

# Designing Networks with Adaptation Rules and Optimal Transport

## Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät  
der Eberhard Karls Universität Tübingen  
zur Erlangung des Grades eines  
Doktors der Naturwissenschaften  
(Dr. rer. nat.)

vorgelegt von

**Alessandro Lonardi**

aus Verona, Italien

Tübingen

2024

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen  
Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation: 26.04.2024

Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Dr. Caterina De Bacco
2. Berichterstatter:	Prof. Dr. Peter Ochs





# Abstract

Efficient transportation of resources is critical for network functionality at all scales. However, while natural systems adapt over time to achieve optimal structures for transportation, man-made networks are not built with a comparable evolutionary mechanism. Consequently, these structures frequently fall short of meeting their intended design criteria.

This thesis presents adaptation rules rooted in biological systems that enable the design of plausible man-made infrastructures. Specifically, we extrapolate mathematical models classically used to study, for instance, the transport of nutrients in plants or the human body and extend them to model different problems with a paradigm shift: Use such equations to get instrumental insight on how to build artificial networks.

We connect adaptation rules and optimality with Optimal Transport (OT) theory. Initially, we formulate adaptation equations tailored to the problem at hand. Then, we aim to find a well-defined Lyapunov functional for these equations, which is interpretable as the cost to transport mass along the edges of a network. This is the cost minimized in OT. This link allows us to leverage optimization insights and methods to enhance performance and validate our adaptation schemes.

While this mechanism is established for greedy routing problems, we extend it to more complex scenarios.

First, we consider a multicommodity problem where different immiscible mass types move in a shared network. By interacting in one infrastructure, the mass types contribute to minimizing a unique cost. We observe that thoughtfully devising the coupling of mass types is pivotal to producing optimal networks. We also explore traffic congestion regimes controlled through a critical exponent entering the adaptation rules and its corresponding optimization formulation. The multicommodity adaptation equations

are used to study the routing of passengers in the Paris Métro and the streets of Bordeaux. These applications showcase which stations are crucial to alleviating traffic under targeted node failures and that trams are a valuable alternative to reduce bus congestion. Furthermore, we employ this method for ameliorating supervised image classification with OT. Here, mass types are RGB color distributions of images, and the OT cost is used as a proxy to assess their similarity.

Second, we study optimal designs of transportation networks with time-dependent input mass loads. Our fundamental assumption is to model the slow evolution of the network infrastructure, which is governed by periodic and fast-fluctuating mass entering its nodes. By postulating the existence of these two different time scales, we derive closed-form adaptation rules that reduce the transport cost upon convergence. Additionally, they enable connecting analytical properties of the mass loads—their Fourier coefficients—with the topology of optimal networks. We use this method to study the robustness of Bordeaux’s bus network.

Third, we frame the competition of a network manager and greedy passengers competing in a bilevel optimization problem. The first aims to minimize traffic by tolling roads, while the second move to reduce their travel costs. To solve the problem, we devise a scheme where adaptation rules for greedy routing are alternated with closed-form Projected Stochastic Gradient Descent for tuning edge weights. Our study on the international E-road network demonstrates that an informed tolling of roads effectively trades off travel time against congestion and can help reduce the carbon footprint of roads.

To make our results reproducible, we complement our methods with open-source codes.

In summary, our models provide a systematic approach to designing optimal transportation networks for different tasks. These tools are valuable for practitioners interested in these problems, for example, policymakers aiming to assess whether a transport infrastructure effectively meets user demand.

# Zusammenfassung

Ein effizienter Ressourcentransport ist für die Funktionalität von Netzen auf allen Ebenen von entscheidender Bedeutung. Während sich natürliche Systeme im Laufe der Zeit anpassen, um optimale Strukturen für den Transport zu erreichen, werden vom Menschen geschaffene Infrastrukturen nicht mit einem vergleichbaren Evolutionsmechanismus gebaut. Infolgedessen erfüllen diese Strukturen häufig nicht die ihnen zugedachten Gestaltungskriterien.

In dieser Arbeit werden Anpassungsregeln vorgestellt, die in biologischen Systemen verwurzelt sind und den Entwurf plausibler vom Menschen geschaffener Infrastrukturen ermöglichen. Konkret werden mathematische Modelle, die klassischerweise zur Untersuchung des Nährstofftransports in Pflanzen oder im menschlichen Körper verwendet werden, extrapoliert und durch einen Paradigmenwechsel zur Modellierung verschiedener Szenarien erweitert: Wir nutzen solche Gleichungen, um instrumentelle Erkenntnisse über den Aufbau künstlicher Netzwerke zu gewinnen.

Wir ziehen eine Verbindung zwischen Anpassungsregeln und Optimalität mit der Optimal Transport (OT) Theorie. Zunächst formulieren wir Anpassungsgleichungen, die auf das jeweilige Problem zugeschnitten sind. Dann versuchen wir, ein wohldefiniertes Lyapunov-Funktional für diese Gleichungen zu finden, das als die Kosten für den Transport von Masse entlang der Kanten eines Netzwerks interpretiert werden kann. Dies sind die Kosten, die in OT minimiert werden. Diese Verbindung ermöglicht es uns, Erkenntnisse und Methoden der Optimierung zu nutzen, um die Leistung zu verbessern und unsere Anpassungsschemata zu validieren.

Während dieser Mechanismus für gierige Routing-Probleme etabliert ist, erweitern wir ihn auf komplexere Szenarien.

Zunächst betrachten wir ein Multicommodity-Problem, bei dem sich verschiedene nicht mischbare Massenarten in einem gemeinsamen Netz bewegen. Indem sie in einer Infrastruktur interagieren, tragen die Massentypen dazu bei, die einmaligen Kosten zu minimieren. Wir stellen fest, dass eine durchdachte Kopplung der Massentypen für die Schaffung optimaler Netze von zentraler Bedeutung ist. Wir untersuchen auch Verkehrsüberlastungsregime, die durch einen kritischen Exponenten in den Anpassungsregeln kontrolliert werden, sowie die entsprechende Optimierungsformulierung. Die Multicommodity-Anpassungsgleichungen werden verwendet, um die Beförderung von Fahrgästen in der Pariser Métro und in den Straßen von Bordeaux zu untersuchen. Diese Anwendungen zeigen, welche Stationen entscheidend sind, um den Verkehr bei gezielten Knotenausfällen zu entlasten, und dass Straßenbahnen eine wertvolle Alternative sind, um die Überlastung der Busse zu verringern. Außerdem setzen wir diese Methode zur Verbesserung der überwachten Bildklassifizierung mit OT ein. In diesem Fall sind die Massentypen RGB-Farbverteilungen von Bildern, und die OT-Kosten werden als Proxy für die Bewertung ihrer Ähnlichkeit verwendet.

Zweitens untersuchen wir die optimale Gestaltung von Transportnetzen, in welche Massen zeitabhängig eingeführt werden. Unsere Hauptannahme ist die Modellierung der langsamen Entwicklung der Netzinfrastruktur, die durch periodische und schnell schwankende Massen, die in die Knoten eintreten, bestimmt wird. Indem wir die Existenz dieser beiden unterschiedlichen Zeitskalen postulieren, leiten wir Anpassungsregeln in geschlossener Form ab, die die Transportkosten bei Konvergenz reduzieren. Außerdem ermöglichen sie es, analytische Eigenschaften der Massen—ihre Fourier-Koeffizienten—mit der Topologie optimaler Netze zu verbinden. Wir verwenden diese Methode, um die Robustheit des Busnetzes von Bordeaux zu untersuchen.

Drittens stellen wir den Wettbewerb zwischen einem Netzmanager und gierigen Fahrgästen dar, die in einem zweistufigen Optimierungsproblem konkurrieren. Der erste versucht, das Verkehrsaufkommen durch die Erhebung von Straßenbenutzungsgebühren zu minimieren, während der zweite versucht, seine Reisekosten zu senken. Zur Lösung des Problems entwickeln



wir ein Schema, bei dem sich Anpassungsregeln für gieriges Routing mit einem in geschlossener Form projizierten stochastischen Gradientenabstieg zur Abstimmung der Kantengewichte abwechseln. Unsere Studie über das internationale E-Road-Netz zeigt, dass eine informierte Mauterhebung für Straßen einen wirksamen Ausgleich zwischen Reisezeit und Staus schafft und dazu beitragen kann, den CO<sub>2</sub>-Fußabdruck von Straßen zu verringern.

Um unsere Ergebnisse reproduzierbar zu machen, ergänzen wir unsere Methoden mit Open-Source-Codes.

Zusammenfassend lässt sich sagen, dass unsere Modelle einen systematischen Ansatz für die Gestaltung von Verkehrsnetzen bieten, die für verschiedene Aufgaben optimal sind. Diese Werkzeuge sind wertvoll für Personen, die sich für diese Probleme interessieren, z.B. für politische Entscheidungsträger, die beurteilen wollen, ob eine Verkehrsinfrastruktur die Nachfrage der Nutzer effektiv erfüllt.



# Acknowledgments

First and foremost, I wish to thank my advisor Caterina De Bacco for her scientific and personal guidance over the last years. I do not know what my future career endeavors are going to be, but I am very happy and feel fortunate that they started with our teamwork. Your tenacity and fervor in doing good science before seeking recognition is an inspiration. Also, thanks to Caterina for fostering the development of my left-hand technique at table soccer. If today my midfielders shoot stronger than ever, it is all thanks to you!

I extend my gratitude to my collaborators (in alphabetical order), Diego Baptista, Enrico Facca, Abdullahi Ibrahim, Mario Putti, Nicolò Ruggeri, and Michael Szell. Your scientific advice and contribution were essential to this thesis and my growth as a researcher. A special thanks goes to Mario, who first presented me with the opportunity to work with science at the end of my Master's degree.

Furthermore, I am grateful to Peter Ochs for reviewing this thesis and for being part of my thesis advisory committee, alongside Anna Levina and Michael Muehlebach, whom I thank in turn. I am grateful again to Peter Ochs, Anna Levina, as well as Georg Martius for forming my examination committee.

I also want to thank all the past and current members of the Physics for Inference and Optimization Group at the Max Planck for Intelligent Systems in Tübingen. Countless hours spent in the office would have not been the same without your encouragement.

My PhD was supported by the International Max Planck Research School for Intelligent Systems (IMPRS-IS) and entirely carried out at the Max Planck for Intelligent Systems. Many dedicated people work to make the school and the institute a wonderful place to make science. I truly appreciate your effort.

On a personal note, I want to thank my friends from back home and every corner of the world. If what you are is the nest where you lay, then thanks for helping me, in one way or another, tearing my nest down and pushing me to be a bit more.

My most heartfelt gratitude is for my family. In particular, I thank our fantastic new additions Gianluca and Pietro, who always spark joy in my life. Then, my polar opposite sister, Valeria. You are and will always be a pillar for my serenity, growing up with you as a model is one of my biggest fortunes. Finally, I wish to thank my Mum and my Dad. I owe you every bit of good that is in me, thank you “for not telling me how to live, but to live and let me watch you do it”<sup>1</sup>.

*Valeria, Mamma e Papà, questo traguardo è dedicato a voi.*

My last thank you goes to Asia. I would like to tell you a million words but thanks for when “I picture myself in a boat on a river, with tangerine trees and marmalade skies, somebody calls me, I answer quite slowly. A girl with kaleidoscope eyes!”<sup>2</sup>.

---

<sup>1</sup>Adapted from Clarence Budington “Bud” Kelland (July 11, 1881 - February 18, 1964).

<sup>2</sup>Adapted from “Lucy in the Sky with Diamonds” by the Beatles (1967).





# Preface

The work presented in this thesis is based on the following peer-reviewed publications and preprints. Papers in each category are listed in chronological order of publication, and asterisks denote equal contributions.

Peer-reviewed publications:

1. Ibrahim *et al.* [1]  
*Optimal Transport in Multilayer Networks for Traffic Flow Optimization*  
Abdullahi Adinoyi Ibrahim, **Alessandro Lonardi**, Caterina De Bacco  
[Algorithms](#), 14(7), 189 (2021) · [arXiv](#) · [GitHub](#)
2. Lonardi *et al.* [2]  
*Designing optimal networks for multicommodity transport problem*  
**Alessandro Lonardi**, Enrico Facca, Mario Putti, Caterina De Bacco  
[Physical Review Research](#) 3, 043010 (2021) · [arXiv](#) · [GitHub](#)
3. Lonardi *et al.* [3]  
*Multicommodity routing optimization for engineering networks*  
**Alessandro Lonardi**, Mario Putti, Caterina De Bacco  
[Scientific Reports](#) 12, 7474 (2022) · [arXiv](#) · [GitHub](#)
4. Lonardi *et al.* [4]  
*Infrastructure adaptation and emergence of loops in network routing with time-dependent loads*  
**Alessandro Lonardi**, Enrico Facca, Mario Putti, Caterina De Bacco  
[Physical Review E](#) 107, 024302 (2023) · [arXiv](#) · [GitHub](#)
5. Lonardi *et al.* [5]  
*Immiscible Color Flows in Optimal Transport Networks for Image*

*Classification*

**Alessandro Lonardi\***, Diego Baptista\*, Caterina De Bacco  
Frontiers in Physics 11:1089114 (2023) · arXiv · GitHub · Poster ·  
CO<sub>2</sub> compensation

6. Lonardi and De Bacco [6]  
*Bilevel Optimization for Traffic Mitigation in Optimal Transport Networks*  
**Alessandro Lonardi**, Caterina De Bacco  
Physical Review Letters 131, 267401 (2023) · arXiv · GitHub

Throughout my doctorate, I also contributed to additional work, which is not included in the thesis.

Additional preprint:

1. Ruggeri *et al.* [7]  
*Message-Passing on Hypergraphs: Detectability, Phase Transitions, and Higher-Order Information*  
Nicolò Ruggeri\*, **Alessandro Lonardi\***, Caterina De Bacco  
Journal of Statistical Mechanics: Theory and Experiment (4), 043403 (2024) · arXiv · GitHub

Additional manuscript in preparation:

1. Lonardi *et al.* [8]  
*Cohesive urban bicycle infrastructure design through optimal transport routing in multilayer networks*  
**Alessandro Lonardi**, Michael Szell, Caterina De Bacco

My scientific contribution to each paper is outlined in Table 1, and it follows the designation:

- Primary Contributor (PC  $> 50\%$ ), for leading research contributions;
- Major Contributor ( $25\% \lesssim MC \leq 50\%$ ), for significant shared contributions that play a substantial role in the research;
- Co-contributor ( $CC \lesssim 25\%$ ), for meaningful but lesser co-authorship contributions.



**Table 1:** Outline of the scientific contribution.

Paper	Model	Exps	Discussion	Writing
Ibrahim <i>et al.</i> [1]	CC	–	CC	CC
Lonardi <i>et al.</i> [2]	MC	PC	PC	MC
Lonardi <i>et al.</i> [3]	MC	PC	PC	MC
Lonardi <i>et al.</i> [4]	PC	PC	PC	PC
Lonardi <i>et al.</i> [5]	MC	MC	MC	MC
Lonardi and De Bacco [6]	PC	PC	PC	PC
Ruggeri <i>et al.</i> [7]	MC	CC	MC	MC
Lonardi <i>et al.</i> [8]	PC	PC	MC	PC



# Table of Contents

<b>Abstract</b>	<b>iii</b>
<b>Zusammenfassung</b>	<b>v</b>
<b>Acknowledgments</b>	<b>viii</b>
<b>Preface</b>	<b>x</b>
<b>1 Introduction</b>	<b>1</b>
1.1 The problem of network modeling . . . . .	2
1.2 Design of networks with adaptation rules . . . . .	6
1.3 Optimal Transport . . . . .	8
<b>2 Goals</b>	<b>11</b>
2.1 Overall aim and contribution . . . . .	11
2.2 Scientific contributions and their organization . . . . .	12
<b>3 Results and discussion</b>	<b>17</b>
3.1 Development of the adaptation rules . . . . .	17
3.1.1 Standard unicommodity routing . . . . .	18
3.1.2 Multicommodity routing . . . . .	22
3.1.3 Routing with time-dependent loads . . . . .	28
3.1.4 Bilevel optimization for traffic mitigation . . . . .	35
3.2 Applications . . . . .	40
3.2.1 Paris Métro . . . . .	40
3.2.2 Bordeaux's trams and buses . . . . .	43
3.2.3 International E-road network . . . . .	46
3.2.4 Image classification . . . . .	49

<b>4 Conclusion and future perspectives</b>	<b>53</b>
<b>Appendix A Published works</b>	<b>57</b>
<b>Bibliography</b>	<b>119</b>





# 1 Introduction

## Content

- Section 1.1 —
1. Connection between adaptation rules and optimization problems;
  2. Central question: Can we leverage optimization to design and study principled (with theoretical guarantees) and flexible (applicable to disparate problems) adaptation rules?
  3. Optimal Transport is our framework of choice, we harness its power to formalize the problem and boost computational performance;
  4. We apply our models to engineering and machine learning tasks.
- Section 1.2 —
1. Functioning principles of adaptation rules;
  2. Literature review of adaptive systems;
  3. Details on the real-world applications of the thesis.
- Section 1.3 —
1. Description of Optimal Transport;
  2. Targeted literature review Optimal Transport;
  3. Optimal Transport and its connection to adaptation rules.

## 1.1 The problem of network modeling

Designing the “Topology of the Fittest Transportation Network” [9] is indispensable for preserving life in nature and advancing our society’s primary infrastructures.

Paradigms of such a task in the natural world include intricate venation patterns on leaves [10–12], which enable the adequate circulation of nutrients or the universal profiles of plants, emerging from the trade-off between hydraulic resistance and carbon cost minimization [13]. Other remarkable examples are given by rivers [14–17], where water tides form many meanders folding into one another and giving rise to complex morphologies, and by the *Physarum polycephalum*, a single-celled multinucleate myxomycete (slime mold) that shaped research on adaptive network formation since the early 2000s [18].

The slime mold offered an ideal test bed for designing seminal *in vitro* experiments and mathematical modeling, which led to the groundbreaking emulation of artificial transportation networks with a biological organism [19, 20]. Remarkably, the organism’s evolutionary mechanism can be explained using minimal biological adaptation rules that yield a natural algorithm [21] with a desirable convergence property: The slime mold transports nutrients by morphing into a network that connects food sources with minimum total length [22, 23].

The connection between adaptation mechanisms and optimization principles provided by the slime mold pinpoints the first overarching question of this thesis:

**Question 1:** Is it possible to formulate principled adaptation rules suitable for a range of tasks, resulting in optimal network designs?

Posing this problem naturally leads to a cascade of follow-up interrogations, which are essential to narrow the domain of investigation and, therefore, to make the formulation of a satisfactory answer feasible.

We need to specify what optimality refers to within different contexts. Its definition can essentially change, and it is intimately dependent on the target



of the system's evolution. For example, while many plants' leaves contain venation with dense loops, some ancient plant species such as the *Ginkgo biloba* organize their vascular pattern in tree topologies [24]. This evidence underscores that optimal transportation networks are potentially very diverse and can emerge from a non-trivial interplay between evolutionary vectors even within the same scientific domain. In the specific case of leaves, optimization may be sought for resilience to external damage [10], among several factors.

The development of many other systems can be explained using different optimization mechanisms. A biological example is the islets of Langerhans in the pancreas, where specialized cells release insulin and glucagon into the bloodstream [25, 26]. Here, optimizing the time to transport these substances allows replicating topologies compatible with those in the pancreatic islets [27].

The mathematical framework we employ to model optimization and to measure optimality is Optimal Transport (OT) theory [28]. Optimal Transport formalizes the problem of finding the best strategy to ship resources from one or multiple origins to one or multiple destinations. More in detail, the goal is to find the transport path that optimally moves every unit of a source mass distribution to any other unit of a target mass distribution. Once again, optimality has been introduced with a broad connotation. However, in OT, the objective function to minimize is typically the product between the so-called ground cost one must pay to move mass units along a path between two locations and the amount of mass traveling onto it. The solution to this problem, which is the minimum of the objective function, is referred to as 1-Wasserstein distance (or Kantorovich-Rubinstein metric) if the ground cost corresponds to the distance between the source and target points.

Optimal Transport is applied across many domains of science, with a large body of literature devoted to establishing a formal bridge between OT and adaptive networks [22, 23, 29–37]. These works develop a robust scientific foundation to answer [Question 1](#). Particularly, they study and formulate the connection between adaptation rules and optimal network design under a dynamical system's perspective. Here, network topologies evolve in time

and adapt their structure to obey Ordinary Differential Equations (ODE) systems. The goal is to design such dynamical systems so that the trajectory drawn by the network's evolution falls down a potential well, reaching a stable and optimal configuration. Crucially, the minimum of the potential well is the 1-Wasserstein distance.

This connection is useful to prescribe a constructive scheme for adaptation rules that follow a minimum energy principle measured in the OT sense. We use this approach in all the works we discuss in this thesis, and as a further driver to develop our models, we pose a second central question:

**Question 2:** How do adaptation rules profit from optimization problems, and vice versa?

In attempting to answer [Question 2](#), we mainly focus on two aspects of the link between OT and dynamical systems for network design.

The first is validating the minimum energy principle followed by adaptation rules via OT. In detail, we use OT theory to prove in what terms and under what conditions the networks extracted at convergence of adaptation are to be considered optimal. We identify that certain aspects of the problems' setup, such as the objective's convexity or the initial configurations of the transported mass, are pivotal in controlling the efficiency of transportation networks. Exploring a large variety of parameter configurations unveils a myriad of transportation networks, each with different features best suited for specific operating functions. By establishing this connection, we also reinterpret OT under a complementary physical perspective where, for example, mass moving along transport paths becomes the current flowing in a resistor network.

The second aspect is computational complexity. With the continually expanding amount of big data we benefit from, devising scalable algorithms for many scientific research tasks has become paramount. Mathematical optimization, particularly OT, is the perfect embodiment of the collective effort aimed at unburdening the computational cost of algorithms. OT garnered popularity in computationally intensive disciplines such as machine learning and deep learning [38] thanks to the development of

high-performing algorithms, for instance, the widely celebrated Sinkhorn’s algorithm [39].

There are plenty of questions we wish to explore in this direction. For example, how do OT-based adaptation rules perform against other OT algorithms? How can they be improved to become faster? Also, what is the difference in efficiency between adaptation algorithms and traditional pathfinding ones, such as Dijkstra’s [40]?

Besides focusing on technical developments of adaptation equations, our studies have a second fundamental target: Addressing real-world applications. We tackle this problem from an orthogonal point of view compared to most of the literature. Instead of studying natural systems that adapt to execute their operating functions, we focus on artificial systems where an external agent designs the networks, and we ask ourselves if the design task can be enhanced by employing insight taken from adaptation rules. In other words,

**Question 3:** Can we leverage adaptation rules for constructing artificial optimal transportation networks?

The *Physarum polycephalum*’s natural algorithm has been employed to study engineering problems such as road and railway design [20, 41–43] and wireless sensor routing [44]. The core idea of this body of work is to understand the structural properties of artificial networks by framing and analyzing biologically plausible adaptation mechanisms that guide the slime mold’s evolution.

In this thesis, we follow a similar strategy to answer [Question 3](#). In particular, we first posit the problem we wish to address, then formulate adaptation rules for solving it and aim to prove their optimality. The adaptation equations we develop do not necessarily need to be biologically plausible but rather give constructive insight into how transportation networks can be built better.

Our primary focus of application lies in urban transportation networks, such as railway systems, roads, and bike lanes. Within this domain, we simulate different real-world scenarios, for example, assuming that

passengers traveling across the network interact with each other and trigger congestion, taking entry and exit rates of metro users that change in time, or making a network manager regulate road tolls that influence optimal routing. These analyses aim to develop a methodology for simulating various critical transportation network scenarios, with the ultimate broad objective of aiding policymakers in designing improved urban services.

We also apply adaptation equations to supervised image classification, showing how accuracy can be boosted with an *ad hoc* adaptation algorithm that uses, at its advantage, samples' color information.

## 1.2 Design of networks with adaptation rules

Adaptation rules are powerful mathematical tools to model the evolution of transportation networks. The idea at their core is that networks evolve in time and organize their structure to allocate mass traveling along their links optimally.

Specifically, mass enters and exits network nodes, moving along the network's edges. The mass displacement between sources (entry nodes) and sinks (exit nodes) generates a flux that obeys local conservation laws, expressing flux conservation at each node, and global conservation laws. The latter formalize that all mass entering the network must also exit. The primary modeling assumption is that edges are endowed with capacities (also referred to as conductivities) that evolve in time to enable efficient mass transfer; namely, they adapt according to a feedback mechanism where edges for which allocating more mass is convenient have large capacities, and conversely, those with less mass have small capacities.

Models using adaptation rules are well-established in the scientific literature. From the first slime mold-inspired equations [18–20, 45, 46], these methods have evolved to tackle several mathematical challenges arising in real-world problems. For example, adaptation rules with stochastic mass loads are used to model leaf vascular structures and blood vessels [10, 47–49]. Spatially distributed loads [50], expanding leaf tissues [51], and external resilience to external damage [10] can also be integrated to achieve

accurate modeling of leaves. To simulate the growth of trees’ conduits, approaches where capacities depend on the distance from plants’ stems have been developed [13].

Complementing these research efforts, OT-based adaptation rules have been extensively studied [22, 23, 29–37, 52–57] with applications ranging from community detection [54], to network extraction from continuous data [35, 57] and noise-induced transportation topologies [55, 56].

Our theoretical focus is to develop adaptation rules for the following case studies. First, we want to investigate a setup where multiple types of mass travel in the same shared infrastructure and collectively contribute to the transportation cost by congesting the edges. To tackle this problem, we develop multicommodity adaptation equations that route fluxes of different mass types (the commodities) by tuning a unique capacity [1–3]. Second, we develop a method for network adaptation with time-varying time loads. By delving into this problem, we discover that leveraging mass load’s periodicity properties permits efficient transportation network design with low computational effort [4]. Third, we frame a bilevel optimization problem where adaptation equations that regulate greedy passengers’ routing are coupled with edge-cost tuning to mitigate traffic. Such a method can effectively alleviate road congestion [6].

We apply these models to different datasets. Using the multicommodity adaptation equations, we design the optimal infrastructure for Bordeaux’s bus and tram multilayer network [1] and for the Paris Métro [3]. Multicommodity adaptation equations enable simulations of different congestion levels for such systems. Additionally, we observe that trams decongest streets by offering greener alternative routes to passengers who would otherwise travel on buses. Applying equations with time-dependent loads to Bordeaux buses shows that optimal routes strongly depend on the evolution of entry-exit inflow over time [4]. The bilevel optimization-adaptation equations model the competition between greedy passengers and a network manager. The first aim is to minimize travel time, and the second is to reduce traffic. Optimal network topologies resulting from such a competition trade-off travel time against congestion; hence, they give an effective

strategy to guarantee transport efficiency while lowering carbon emissions on the International European highways [6].

One further application of the multicommodity adaptation equations deals with supervised image classification [5]. In more detail, we integrate color information of different samples into the adaptation setup by assuming that the types of mass to be transported are RGB color distributions of images. These contribute to a unique cost, with RGB-color differences that better distinguish between samples with different labels than other OT-based methods.

### 1.3 Optimal Transport

Optimal Transport theory endured a long and intricate journey to modern days. Its original development [58] dates back to 1781, when Gaspard Monge formalized the problem of moving most efficiently piles of material from mines to factories. Naturally, its current formulation [28] abstracts from any practical interpretation, primarily thanks to the work of A.N. Tolstoi [59] (1930), Frank L. Hitchcock [60] (1941), and Leonid Kantorovich [61] (1942).

To further lay a sound foundation for what we briefly mentioned, the goal of OT is to find a transport path (also referred to as a transport plan) that connects a source and a target distribution with the least effort. More formally, given two probability distributions lying in a suitable metric space, we want to find the push forward of one source distribution to a target distribution to minimize the transportation cost. In this context, the transportation cost is often defined as the product between a ground cost, which is the effort needed to move a single unit of mass along a trajectory, and the mass moving onto this trajectory. If the ground cost is a distance for the metric space, then the solution of this problem is also a distance, referred to as 1-Wasserstein distance [28, 62]. However, even when this theoretical guarantee is not met, the minimum OT cost may still be used as an interpretable measure of the effort needed to push one distribution into the other.

One of the most significant benefits of OT is that its formulation is well-suited to deal with different mathematical frameworks. Its continuous formulation is connected to variational problems and partial differential equations of physics [28, 63], while its discrete version is growing in popularity thanks to many algorithmic applications [38].

In this thesis, we focus on the latter discrete case. We establish a connection between OT and adaptation systems as follows. Adaptive networks' entry and exit mass inflows are the atomic origin and target distributions to be transported in OT. These are supported on the network nodes and move along the network edges. Each edge is associated with a cost, which is the ground cost entering the transportation cost of OT. Transport paths correspond precisely to the fluxes specifying the displacement of mass along edges, and the goal of OT (respectively for adaptation rules) is to extract transport paths that move source and target distributions (respectively, entry and exit mass inflows) optimally. The capacities introduced in adaptation rules serve as critical variables for reinterpreting the OT problem from a physical perspective and help improve computational performance.

Thanks to its great flexibility to model all kinds of different problems and its excellent computational properties, OT is employed across all fields of science, spanning genomics [64–67], computer vision [68–73], and stochastic thermodynamics [74], among many others. Notable developments of OT stem from machine learning and deep learning [38], with efficient algorithms for the OT problem being the core of such advancements. In particular, substantial effort is put into Sinkhorn's algorithm [39], its improvements [75–79], and its open sourcing [80–82]. Optimal transport is also at the foundation of landmark deep learning architectures such as Wasserstein GANs [83–85].

Finally, OT is bridged to adaptation rules thanks to its dynamical system's interpretation [22, 23, 29–37]. This framework builds upon the idea that adaptive networks evolve their shape to minimize their energy expenditures, hinting that mathematical optimization is the natural setup to describe such physical systems [9, 24, 27, 49, 86–90]. Therefore, adaptation is an instrument to achieve optimality of network topologies asymptotically.

The role of OT is to be the tool with which optimality is measured and to provide a rigorous basis to connect energy minimization with the functional optimality of networks.



## 2 Goals

### Content

Section 2.1 — The recipe used to achieve the goals of the thesis.

Section 2.2 — Organization of the contributions into sections with references to their respective papers.

### 2.1 Overall aim and contribution

The main goal of the research presented in this thesis is to answer [Question 1](#), [Question 2](#), and [Question 3](#) discussed in [Section 1.1](#). These are:

**Question 1:** Is it possible to formulate principled adaptation rules suitable for a range of tasks, resulting in optimal network designs?

**Question 2:** How do adaptation rules profit from optimization problems, and vice versa?

**Question 3:** Can we leverage adaptation rules for constructing man-made optimal transportation networks?

We achieve this goal by developing an OT-based framework for constructing adaptation rules that generate provably optimal networks. The recipe we follow to do so is as follows.

First, we extend traditional adaptation equations to address the task at hand effectively. For example, when modeling multicommodity problems where several mass types share a single infrastructure, we must choose an adequate coupling for commodities. Similarly, when taking time-varying

loads, we must make assumptions on the mass inflow profiles to make the problem analytically tractable.

Then, we study the optimality properties of our adaptation rules by seeking a well-defined Lyapunov functional. This function is the energy well into which adaptive networks fall. Optimal Transport assists us in this step by providing an interpretation of the Lyapunov functional, which we connect to the minimum transportation cost needed for mass to move from entry nodes to exit ones.

Remarkably, we implement the adaptation equations in efficient numerical schemes by exploiting several OT and dynamical systems insights. We release all our algorithms open-source [91–95].

We employ our models to address applications in engineering and machine learning. The primary objective of our empirical analyses is to simulate diverse scenarios that may arise in real-world problems. For example, we address the impact of gridlocks on the functioning of public transportation, or we use bilevel optimization to develop methods to make car traveling environmentally sustainable. We conduct such analyses both at a coarse scale by considering the global efficiency of transport and at a finer level, all the way down to city neighborhoods. Adaptation rules for supervised image classification tasks boost accuracy by meaningfully integrating color information of images.

## **2.2 Scientific contributions and their organization**

Scientific contributions are discussed in detail in Chapter 3. Here, we divide them into theoretical developments in Section 3.1 and applications in Section 3.2. However, it is worth remembering that the approach used to build our models is largely application-driven, and theory and practice are strongly intertwined in all the research work we present.

For this reason, in the following content outline, we reference each theoretical model to its corresponding applications and vice versa. This should facilitate understanding the connection between these two aspects of

the thesis. The outline also serves as a TL;DR summarization to navigate single sections easily.

## Theory (Section 3.1)

### Standard unicommodity routing (Section 3.1.1)

*Main findings* Adaptation rules are an efficient mechanism for network design. They can be carefully developed to connect the minimum energy spent to move mass inflows to the OT cost. Here, we provide a theoretical foundation to understand such systems.

*Research papers* Bonifaci *et al.* [22], Facca *et al.* [29], among others

### Multicommodity routing (Section 3.1.2)

*Main findings* We develop adaptation rules for multicommodity OT, a method to extract optimal networks when multiple mass types move in a shared infrastructure. We prove that asymptotic networks are optimal by finding a suitable Lyapunov functional. We formalize several other connections between multicommodity adaptation and mathematical optimization and find that the interaction of commodities triggers the formation of loops in optimal networks. Our algorithm is tested against different competitors: Dijkstra [40], Sinkhorn [39], and Gradient Descent (GD).

*Applications* Paris Métro (Section 3.2.1), Bordeaux's trams and buses (Section 3.2.2), Image classification (Section 3.2.4)

*Algorithm* McOpt: Multicommodity Optimal Transport [91]

*Research papers* Lonardi *et al.* [2, 3]

### Routing with time-dependent loads (Section 3.1.3)

*Main findings* We study the problem of non-stationary entry and exit mass inflows by assuming the existence of two time scales for adaptation. A fast one is for mass fluxes, and a slow one is for capacities. This leads to closed-form adaptation rules that allow us to connect analytical properties of the mass loads with optimal topologies. Our method offers a computationally efficient alternative to extract adaptation-asymptotic topologies.

*Application* [Bordeaux's trams and buses](#) (Section 3.2.2)

*Algorithm* N-STARK: Non-stationary Loads Routing on Networks [94]

*Research paper* Lonardi *et al.* [4]

### Bilevel optimization for traffic mitigation (Section 3.1.4)

*Main findings* We frame the competition between a network manager and greedy passengers in a bilevel optimization problem. The first minimizes traffic jams, and the second reduce transportation costs. We find closed-form adaptation equations alternating the update of edge capacities and weights to solve the problem. The equations employ ODE integrators and Projected Stochastic GD to return networks that trade-off transport efficiency against edges' saturation.

*Application* [International E-road network](#) (Section 3.2.3)

*Algorithm* BROT: Bilevel Routing on Networks with Optimal Transport [95]

*Research paper* Lonardi and De Bacco [6]

## Applications (Section 3.2)

### Paris Métro (Section 3.2.1)

*Main findings* We model multicommodity routing within the Paris Métro using a well-suited formulation of the multicommodity adaptation equations. Our investigation focuses on congested routing and its impact on the overall network performance. Additionally, we study passengers' rerouting in the event of station failures by selectively eliminating crucial nodes from the network.

*Theory* Multicommodity routing (Section 3.1.2)

*Research paper* Lonardi *et al.* [3]

### Bordeaux's trams and buses (Section 3.2.2)

*Main findings* First, we examine the effect of congested transport on the multilayer network of roads and tram lines in Bordeaux using the multicommodity adaptation equations. With these, we quantify the impact of trams in providing a green and efficient alternative to city trips. Then, we explore the effect of synchronization of passengers' inflows in relation to the presence of loops in the optimal transport network.

*Theory* Multicommodity routing (Section 3.1.2), Routing with time-dependent loads (Section 3.1.3)

*Research papers* Lonardi *et al.* [4], Ibrahim *et al.* [1]

### International E-road network (Section 3.2.3)

*Main findings* We employ our adaptation rules for bilevel optimization to simulate the routing of cars on the International E-road network. We assume a network manager assigns road tolls to mitigate gridlocks, whereas greedy passengers move along the shortest paths. Trading off transport costs against

traffic mitigation effectively alleviates congestion while still outputting convenient origin-destination routes. If the network manager assigns road tolls without periodically accounting for passengers' rerouting, traffic congestion increases, giving an effect opposite to what is intended.

*Theory* [Bilevel optimization for traffic mitigation](#) (Section 3.1.4)

*Research paper* Lonardi and De Bacco [6]

### **Image classification (Section 3.2.4)**

*Main findings* We use multicommodity adaptation equations for supervised image classification. We strategically design the transport network by employing insights from computer vision [68, 69] and unbalanced OT [96, 97] to obtain fast performances and robust results. By taking the RGB color distributions as the mass types to be transported and using the OT cost as a proxy for image similarity, we get higher accuracy than image classification carried out by competitor algorithms.

*Theory* [Multicommodity routing](#) (Section 3.1.2)

*Research paper* Lonardi *et al.* [5]

# 3 Results and discussion

## Content

Section 3.1 — Refer to Section 2.2 (Theory).

Section 3.2 — Refer to Section 2.2 (Applications).

## 3.1 Development of the adaptation rules

In this section, we develop OT-based adaptation rules for different network designing tasks and study their analytical properties. Our methods are built upon complex networks, which we briefly formalize to provide a sufficient mathematical framework.

A network [98]  $G = (V, E)$  is a tuple of nodes  $v \in V$  and edges  $e = (u, v) \in E$ , which join pairs of nodes. Connections of nodes and edges can be stored in the signed incidence matrix of the network, whose elements are  $B_{ve} = \pm 1$  if  $v$  is the head or the tail of  $e$ , and  $B_{ve} = 0$  when it is none of the two. We assume that edges have positive weights  $w_e > 0$ . Such weights potentially hold different domain-dependent interpretations, for example, the Euclidean length between node coordinates or the amount of money to pay to travel on the road. Additionally, we define the network  $x$ -Laplacian as the matrix with elements  $L_{uv}(x) = \sum_e B_{ue}B_{ve}x_e$  for any  $x_e \geq 0$ .

Networks often are stacked into multilayer structures [99]. A multilayer network is made of  $L$  layers where nodes lie. Edges connecting nodes within the same layer  $a = 1, \dots, L$  are called intra-layer edges, whereas those between any layers  $a \neq b$  are called inter-layer edges.

### 3.1.1 Standard unicommodity routing

We present OT-based adaptation rules for unicommodity routing similar to those introduced in the literature [22, 23, 29, 30]. These are equations that, at convergence, output the shortest path fluxes between entry and exit nodes. Initially, we describe their functioning and then formalize their connection with OT. To distinguish this first model from the multicommodity one in Section 3.1.2, we refer to it as unicommodity routing. Additionally, since multicommodity routing is a generalization of the unicommodity one, it is often sufficient to restrict proofs developed for the former setup to the latter. For this reason, we often refer to literature on multicommodity routing [2, 5, 22] when establishing connections between unicommodity adaptation rules and its corresponding optimization framework.

Adaptation rules aim to control a network's time evolution to optimally allocate the entry and exit mass flowing through its nodes. We represent mass using two non-negative arrays  $g_v$  and  $h_v$  with identical total mass  $\sum_v g_v = \sum_v h_v$ . The first array is the mass distribution inflowing from nodes, and the second is the one flowing out.

Our primary modeling assumption is that edges are endowed with non-negative capacities  $\mu_e \geq 0$ . These variables admit a straightforward physical interpretation: They can be thought of as the physical space allocated on edges to accommodate mass fluxes. For example, they can be the width of roads where passengers travel or the size of tubular veins where nutrients are transported in plants and the human body.

Fluxes on edges are denoted with  $F_e$  and obey Kirchoff's law, formalizing the local conservation of mass at every node. Namely, the difference in entering and exiting flux at each node  $v$  must equal its net inflow of mass  $S_v = g_v - h_v$ . In formula,

$$\sum_e B_{ve} F_e = S_v \quad \forall v \in V. \quad (3.1)$$

Typically, the number of edges is larger than the number of nodes; hence, Eq. (3.1) is undetermined. To make the fluxes well-defined, we assume that they are related to the difference of a scalar pressure potential  $p_v$  by the



following relation, often referred to as Hagen-Poiseuille equation in fluid dynamics [100],

$$F_e = \frac{\mu_e}{w_e}(p_u - p_v) \quad \forall e = (u, v) \in E. \quad (3.2)$$

Substituting Eq. (3.2) into Eq. (3.1) allows us to find the potential

$$p_v = \sum_u L(\mu/w)_{vu}^\dagger S_u \quad \forall v \in V, \quad (3.3)$$

where  $\dagger$  denotes the Moore-Penrose inverse of the  $(\mu/w)$ -Laplacian. The system in Eq. (3.3) has a solution since  $\sum_v S_v = 0$  [101], which is guaranteed by construction from global conservation of mass, specifically, the entry and exit arrays  $g_v$  and  $h_v$  sum to the same value. Importantly, through Eq. (3.3), the potential, hence the potential-based fluxes of Eq. (3.1), become solely a function of the capacities, which are the only independent variable of our model.

With this construction, we can build the unicommodity OT adaptation equations. These establish a feedback mechanism between fluxes and capacities: The more the flux to be accommodated on an edge, the larger its capacity becomes; conversely, edges where there is no flux shrink. The unicommodity adaptation rules are

$$\frac{d\mu_e}{dt} = \frac{|F_e|^2}{\mu_e} - \mu_e \quad \forall e \in E. \quad (3.4)$$

By coupling the closed-form potential in Eq. (3.3) with Eq. (3.4), we can solve the adaptation equations and find asymptotic capacities  $\mu_e^* = \lim_{t \rightarrow +\infty} \mu_e(t)$ . These are useful to extract asymptotic fluxes  $F_e^* = F_e(\mu^*)$ , representing mass distribution over the network edges at convergence.

The critical properties of this model are the following. The adaptation rules in Eq. (3.4) asymptotically converge to the minimum of a convex Lyapunov functional. The functional can be interpreted in physical terms. Indeed, it is the sum of the energy dissipated by the mass moving through the edges and the cost of building the network infrastructure. Furthermore,

the functional's minimum is the 1-Wasserstein distance between entry and exit mass inflows, and the asymptotic flux converging to such a minimum is exactly the transport path that solves the OT problem [2, 5, 22, 23, 31, 34].

The Lyapunov functional reads [31]

$$\mathcal{L}(\mu) = \frac{1}{2} \sum_v p_v S_v + \frac{1}{2} \sum_e w_e \mu_e. \quad (3.5)$$

Its first addend is equivalent to  $J = (1/2) \sum_e w_e F_e^2 / \mu_e$ , which is, analogously to Joule's first law in electromagnetism where the fluxes are the currents moving in a resistor network, the dissipated energy. The energy  $J$  is also the objective function in unicommodity optimization setups to design optimal transportation networks [9, 87]. The second addend  $W = \sum_e w_e \mu_e$  corresponds to the cost needed to design the infrastructure. If we were to interpret the weights as physical lengths and the capacities as widths of edges, such a term in Eq. (3.5) would be the total volume of the network.

To precisely map the adaptation rules to an optimization setup, we show that stationary solutions of Eq. (3.4) are minimizers of the constrained energy minimization problem

$$\min_{\mu, F} \left\{ J = \frac{1}{2} \sum_e \frac{w_e}{\mu_e} F_e^2 \right\} \quad (3.6)$$

$$\text{s.t. } \sum_e w_e \mu_e = K > 0 \quad (3.7)$$

$$\sum_e B_{ve} F_e = S_v \quad \forall v \in V \quad (3.8)$$

in that optimal solutions of both obey the relation  $\mu_e \sim |F_e|$ . Here,  $K > 0$  is an arbitrary constant.

By restricting the search space of Eq. (3.6) with this scaling and opportunely adding Lagrange multipliers for the constraints in Eqs. (3.7)-(3.8),

we get that solving the auxiliary constrained energy minimization problem is equivalent to seeking solutions of the convex problem [9]

$$\min_F \left\{ C = \sum_e w_e |F_e| \right\} \quad (3.9)$$

$$\text{s.t. } \sum_e B_{ve} F_e = S_v \quad \forall v \in V \quad (3.10)$$

with fluxes that scale linearly with capacities.

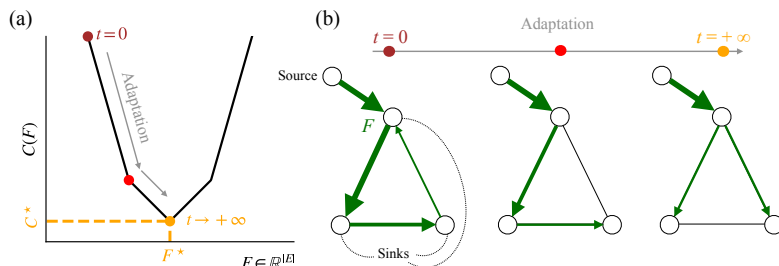
Importantly, the connection with OT can be completed using Eqs. (3.9)-(3.10). In detail, we formulate the OT minimization problem in its so-called primal Kantorovich formulation. This is,

$$\min_{\pi \in \Pi(g, h)} \sum_{uv} w_{uv} \pi_{uv}, \quad (3.11)$$

where  $\Pi(g, h)$  is the set of feasible transport paths  $\pi$ , expressing the probability of moving mass from  $u$  to  $v$  while satisfying conservation of mass  $\sum_v \pi_{uv} = g_u$  and  $\sum_u \pi_{uv} = h_v$ . We show the equivalence between Eq. (3.11) and Eqs. (3.9)-(3.10) by setting  $w_{uv} = w_e$  for each  $e = (u, v)$  and  $w_e = +\infty$  if  $u$  and  $v$  are disconnected. This way, we prove that the objective function  $C$  in Eq. (3.9) and the OT cost in Eq. (3.11) are equal, as well as the search spaces yielded by  $\Pi(g, h)$  and by Kirchhoff's law.

In summary, by solving the adaptation equations in Eq. (3.4) and extracting their asymptotic fluxes, we find minimizers of  $C$  for the optimization problem in Eqs. (3.9)-(3.10). Such minimizers are exactly the OT paths solving the OT minimization problem in Eq. (3.11). Recall that if  $w$  satisfies the properties of a distance, then the minimum OT cost  $C^* = C(F^*)$  is the 1-Wasserstein distance between entry and exit mass inflows [28, 62].

The connection between optimization and adaptation at the core of the thesis is depicted schematically in Fig. 3.1.



**Figure 3.1: Schematic depiction of adaptation equations for unicommodity (greedy) routing.** (a) The convex cost  $C(F)$  is minimized at convergence of Eq. (3.4). The adaptation rules return the stationary capacities  $\mu^*$  with which we can compute  $F^*$ , hence, the optimal cost  $C^* = C(F^*)$  (in orange). (b) Example network where mass enters one source node and exits three sinks. Capacities and their corresponding fluxes (in green) adapt in time. At convergence, the optimal transport network is a tree [9, 87, 89].

### 3.1.2 Multicommodity routing

Multicommodity routing [102] deals with the problem of simulating optimal transport paths for  $M$  distinguishable types of mass moving through a unique network. Such models are necessary for a variety of applied tasks. For example, to obtain realistic paths of passengers using public transport [1, 3, 34, 52], or to simulate data packets' routing in communication networks [103].

To address these problems, practitioners can rely on a variety of methods, from greedy heuristics [102] to belief propagation [103–108]. However, these algorithms often lead to suboptimal results for dense networks and suffer from scalability issues. Here, we present multicommodity adaptation rules developed in Lonardi *et al.* [2, 3], providing an efficient and optimal alternative to address multicommodity routing.

First, we illustrate the model in relation to its unicommodity counterpart of Section 3.1.1, mainly focusing on the critical modeling assumptions

and their central differences. Then, we discuss the connection between multicommodity congested transport (which is introduced through the adaptation equations) and the generation of loops in optimal transportation networks. Lastly, we perform an analytical and computational comparison between OT-based adaptation rules and other algorithms.

We model commodities  $i = 1, \dots, M$  by introducing indistinguishable mass types stored in entry and exit distributions  $g_v^i$  and  $h_v^i$ , respectively. The  $M$  commodities generate as many fluxes,  $F_e^i$ , obeying Kirchhoff's law in Eq. (3.1), but where conservation of mass is enforced between each  $i$ -th flux and each mass matrix entry  $S^i = g_v^i - h_v^i$ . Similarly to the unicommodity case, to make the fluxes well-defined, we introduce a scalar potential on nodes, with the difference that now the potential is commodity-dependent, namely, it is of the form  $p_v^i$ .

We want to model the scenario where  $M$  commodities travel in one unique network infrastructure. Therefore, our first important modeling assumption is that capacities are commodity-independent. This is enforced by setting

$$\mu_e = \mu_e^i \quad \forall e \in E, \forall i = 1, \dots, M. \quad (3.12)$$

With Eq. (3.12), we impose that no mass moving in the network is prioritized; an example of such a scenario is a data routing network where users have access to equal bandwidth. The Hagen-Poiseuille law becomes  $F_e^i = \mu_e(p_u^i - p_v^i)/w_e$  for all edges  $e = (u, v)$  and commodity indexes  $i = 1, \dots, M$ .

Our second crucial modeling choice is establishing the coupling between commodities in the adaptation rules, i.e., the mechanism determining the optimal network's topology. We define multicommodity adaptation rules as

$$\frac{d\mu_e}{dt} = \frac{f_e(F)}{\mu_e^\gamma} - \mu_e \quad \forall e \in E \quad (3.13)$$

$$f_e(F) = \sum_i (F_e^i)^2 \quad \forall e \in E, \quad (3.14)$$

thus taking the fluxes' squared 2-norm over the commodities. Other choices can be made, as taking the 1-norm of fluxes [3], the whole spectrum of  $p$ -norms, or a sigmoid function as done for biological modeling [45]. However, as discussed below, the 2-norm allows us to establish desirable theoretical connections between multicommodity adaptation rules and a suitable optimization framework.

Importantly, in Eq. (3.13), we also introduce a critical exponent  $0 < \gamma < 1$ . This allows us to simulate different levels of traffic congestion by concentrating fluxes over fewer edges when  $\gamma < 1$  and enforcing their distribution over the network when  $\gamma > 1$ . The case  $\gamma = 1$  gives shortest path-like networks. Restricting Eqs. (3.13)-(3.14) to the unicommodity problem where  $M = 1$  and  $\gamma = 1$  returns precisely the shortest path fluxes connecting exit and entry mass inflows.

We find a Lyapunov functional for the multicommodity adaption rules, which can be interpreted as the total transportation cost, the sum of the energy dissipated by the multicommodity fluxes and the cost used to build the network infrastructure. This reads

$$\mathcal{L}_\gamma^M(\mu) = \frac{1}{2} \sum_{iv} p_v^i S_v^i + \frac{1}{2\gamma} \sum_e w_e \mu_e^\gamma. \quad (3.15)$$

Conversely to Eq. (3.5), the functional in Eq. (3.15) is not strictly convex. The presence of multiple commodities combined with the critical exponent  $\gamma$  makes the cost profiles highly non-trivial. In particular, if  $\gamma < 1$ , the  $\mathcal{L}_\gamma$  is strictly convex, and its unique global minimizer is the optimal transport network. If  $\gamma > 1$ , the infrastructure cost  $W_\gamma^M = \sum_e w_e \mu_e^\gamma / 2\gamma$  exhibits multiple local minima, each corresponding to a different locally optimal network topology.

Again, we stress that multicommodity adaptation rules can be modified by setting arbitrary coupling  $f$  between fluxes. However, the specific choice of the 2-norm in Eq. (3.14) allows us to find a well-defined Lyapunov functional and thus to establish a formal connection with OT.

The choice of this coupling lets us formulate a constrained multicommodity energy minimization problem that generalizes Eqs. (3.6)-(3.7). This is

$$\min_{\mu, F} \left\{ J^M = \frac{1}{2} \sum_{ei} \frac{w_e}{\mu_e} (F_e^i)^2 \right\} \quad (3.16)$$

$$\text{s.t. } \sum_e w_e \mu_e^\gamma = K^\gamma > 0 \quad (3.17)$$

$$\sum_e B_{ve} F_e^i = S_v^i \quad \forall v \in V, \forall i = 1, \dots, M. \quad (3.18)$$

Minimizers of Eqs. (3.16)-(3.18) and stationary solutions of Eqs. (3.13)-(3.14) satisfy the scaling  $\mu_e = [\sum_i (F_e^i)^2]^{1/(1+\gamma)}$ , which leads to the definition of a multicommodity OT cost finalizing the connection between adaptation and OT for multicommodity routing. In particular, we derive the minimization problem

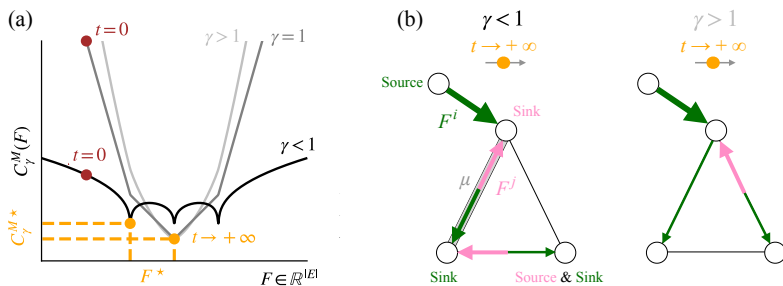
$$\min_F \left\{ C_\gamma^M = \sum_e w_e f_e(F)^{\Gamma(\gamma)} \right\} \quad (3.19)$$

$$\text{s.t. } \sum_e B_{ve} F_e = S_v \quad \forall v \in V, \quad (3.20)$$

where  $f_e(F) = \sum_i (F_e^i)^2$  as in Eq. (3.14) and  $\Gamma(\gamma) = \gamma/(1 + \gamma)$ . Solving the adaptation rules in Eqs. (3.13)-(3.14) and the optimization problem in Eqs. (3.19)-(3.20) is equivalent, meaning that the two setups, given identical inputs, output the same transportation networks.

Noticeably, at optimality, stationary solutions  $F^*$  of the multicommodity dynamics lie in a Pareto front that can be expressed in closed-form as

$$\frac{J^M(F^*)}{W_\gamma^M(F^*)} = \gamma. \quad (3.21)$$



**Figure 3.2: Schematic depiction of adaptation equations for multicommodity routing.** (a) The convexity of the transportation cost  $C_\gamma^M(F)$  depends on the critical exponent  $\gamma$ . The cost is strictly convex for  $\gamma \geq 1$ , whereas if  $\gamma < 1$ , then the minimization landscape becomes rugged. This causes adaptation rules to fall in any of the local minima of the transportation cost (in orange). (b) Example network where there are different types of mass (commodities) being transported. Here, nodes can be sources for one type of mass, e.g., the green one, but sinks for the pink one, and vice versa. The colored fluxes  $F^i$  (green) and  $F^j$  (pink) are coupled by a commodity-independent capacity  $\mu$ . The critical exponent  $\gamma$  concentrates fluxes together if  $\gamma < 1$ . Conversely, fluxes are more distributed if  $\gamma > 1$ .

This relation arises in various applied problems, ranging from urban transportation [3] to biology [13]. Exploring its connection with the topologies of transport networks allows us to interpret the driving evolutionary vectors traded off by the multicommodity equations to design optimal infrastructures. Precisely, we measure the number of edges unused during transport and the Gini coefficient [109] of the multicommodity fluxes, which measures traffic congestion over the edges. This study shows dense transport networks (high  $\gamma$ ) are energy-expensive but robust against edge failure. On the contrary, sparse networks (low  $\gamma$ ) are energy-efficient but not resistant to failures.

We schematically outline the connection between multicommodity adaptation rules and their optimization setup in Fig. 3.2.



One important consequence of having multiple commodities moving in a unique infrastructure is that they trigger the generation of loops in non-trivial congestion regimes. In particular, it is established that for  $\gamma \leq 1$ , the unicommodity adaptation rules in Eq. (3.4) converge to trees [9, 87, 89]. However, if, for example, mass loads change location stochastically, loops can be observed even at  $\gamma \leq 1$  [10, 49, 50].

Similarly, the multicommodity adaptation rules output networks with loops when optimal unicommodity transport networks are trees. However, the mechanism prompting the formation of such loops is different than those classically explored in the literature. In fact, the multicommodity problem is entirely deterministic, and loops emerge only because multiple mass types simultaneously occupy the same edges, rather than being a product of random mass fluctuations. The absence of stochasticity in the problem setup allows us to extend classical arguments [89] to analytically demonstrate that loops may be optimal for  $\gamma \leq 1$ . We also validate our argument with several experiments.

Concerning computational aspects of our model, we numerically integrate the multicommodity adaptation equations using the forward Euler method. This is coupled with a sparse direct solver for the  $M$  linear system in Kirchhoff's law, the Unsymmetric MultiFrontal sparse LU Factorization (UMFPACK) method [110]. The forward Euler method exhibits remarkable convergence properties, this is because Eqs. (3.13)-(3.14) yield a mirror-descent dynamics for  $\mathcal{L}_\gamma$  in Eq. (3.15) [34], i.e., they automatically preserve positiveness of the capacities, enabling the usage of large numerical integration steps.

The solution of adaptation rules has been made faster in the unicommodity case and on sparse networks by using Multigrid solvers of Kirchhoff's law [111], and the inexact Newton-Raphson methods for updating the capacities [33].

We compare our solver against GD with momentum for the fluxes and without the capacities in the problem formulation and Dijkstra's algorithm. The multicommodity adaptation dynamics outperforms competitors. Here, one should note that the multicommodity OT cost in Eq. (3.19) is non-linear with respect to the commodities. Therefore, multicommodity optimal

transport networks are not equivalent to those returned by Dijkstra, where  $M$  independent shortest paths overlap. Our analysis highlights the difference between multicommodity networks and Dijkstra’s network.

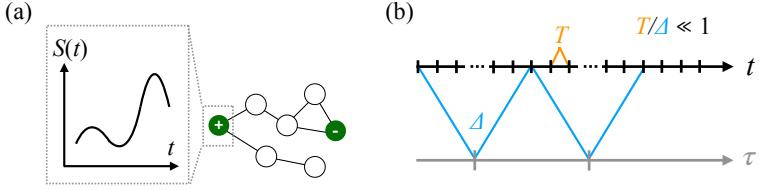
In the literature, Monte Carlo algorithms are also employed [87]. However, these require restricting the search space of optimal networks to trees to become feasible. Such an assumption cannot be made in multicommodity setups, where we show that loops can arise optimally for any congestion regime.

### 3.1.3 Routing with time-dependent loads

All the formulations discussed so far assumed a constant set of mass loads on nodes entering and exiting the system. This is fixed as an input to the models. For instance, in urban transportation one can use average travel demands at stations at any given day.

This section aims to present the model of Lonardi *et al.* [4], which deals with the problem of finding optimal transport networks when entry and exit mass loads on nodes *change in time*. This scenario arises in many case studies. For example, blood vessels adapt their structures to accommodate time-varying metabolic demands [47, 112–114], or passengers in urban transportation travel in periodical irregular intervals [115].

Despite numerous experimental evidence, most of the adaptation models presented in the literature only consider stationary loads [22, 23, 29–32, 50, 51]. However, remarkable changes in optimal network topologies occurs when, for example, studying a network’s evolution with respect to the ensemble average of its position changing loads [10, 47–49]. Remarkably, this yields a classically unobserved (in setups with stationary loads) phase transition for the loops present in the network, which now emerge to provide routes to accommodate mass fluctuations. As discussed in Section 3.1.2, loops also show up in non-trivial regimes in multicommodity routing because of the interaction of multiple mass types over the network edges. To complement these approaches, we develop adaptation rules for non-stationary mass distributions.



**Figure 3.3: Model construction for adaptation equations with time-dependent loads.** (a) Loads change rapidly in time, and their values are stored in  $S(t)$ . (b) We assume the existence of two time scales. The first is a fast scale  $t$  that dictates the evolution in time of mass and fluxes. We also assume that loads are periodic in the short period  $T$ . The second is a slow characteristic time  $\tau$  that regulates adaptation, the evolution of capacities. The integration window  $\Delta$  is large in  $t$  but small in  $\tau$ .

Our method is based on a crucial underlying assumption: Two different time scales, a fast ( $t$ ) and a slow ( $\tau$ ) one. Mass loads, and therefore their fluxes, change fast. On the other hand, capacities adapt slowly to the flux changes. One could interpret this as the case where passengers move in a transportation network with daily or weekly rates, but changes in the infrastructure happen at a much coarser scale, e.g., months. We schematically draw the model setup in Fig. 3.3.

By making this modeling assumption, we derive closed-form adaptation rules and analytical conditions on the mass matrix under which optimal topologies are trees or contrarily, may have loops. We propose a candidate Lyapunov functional for the adaptation rules, prove its optimality for a particular class (formalized later) of input mass loads, and numerically show its monotonically decreasing profile over time under looser assumptions. Although not evidently, our model also connects with long-run solutions of other adaptation rules where capacities change fast. Therefore, it provides a computationally cheap alternative to finding optimal networks that would otherwise be extracted by numerically integrating a time-dependent dynamical system at long times.

We consider the case of unicommodity routing, that is, when only one type of mass moves across the network, as in Section 3.1.1. However, now we take the mass matrix to be time-dependent, with its entries being  $S_v(t) > 0$  for entry inflows and  $S_v(t) < 0$  for exit ones. Global mass conservation is always satisfied, formally  $\sum_v S_v(t) = 0$  for all  $t \geq 0$ . Similarly, flux conservation in Kirchhoff's law becomes  $\sum_e B_{ve} F_e(t) = S_v(t)$  for all  $t \geq 0$ . These relations put into equations the dependence on fast times of fluxes.

We develop closed-form adaptation equations for slow time-varying capacities by starting from the fast time-varying dynamical system,

$$\frac{d\mu_e(t)}{dt} = \frac{F_e(t)^2}{\mu_e(t)^\gamma} - \mu_e(t) \quad \forall e \in E, \quad (3.22)$$

where we write explicitly all time dependencies. As in Eq. (3.13), we introduce a critical exponent  $\gamma$  to control fluxes concentration over the edges. Specifically,  $\gamma < 1$  encourages mass concentration over fewer links,  $\gamma > 1$  its distribution, and  $\gamma = 1$  gives shortest path-like networks.

To model the evolution of slow capacities, we first formalize the relation between the time scales as  $\tau = Kt$ , with  $K \gg 1$ . Secondly, we assume that in a time window  $\Delta$ , small for  $\tau$  but large for  $t$ , slow evolving capacities obey

$$\hat{\mu}_e(\tau + t') \approx \hat{\mu}_e(\tau) \quad \forall t' \in [0, \Delta), \forall e \in E, \forall \tau \geq 0. \quad (3.23)$$

We couple these two hypotheses with a third one, namely, slow capacities evolve with the time integral over the window  $\Delta$  of the product of the mass loads. Formally, we define the slow dynamics

$$\frac{d\hat{\mu}_e(\tau)}{d\tau} = \hat{\Phi}_e(\hat{\mu}(\tau), \tau) \quad \forall e \in E \quad (3.24)$$

$$\hat{\Phi}_e(\hat{\mu}, \tau) = \frac{\hat{\mu}_e^{2-\gamma}}{w_e^2} \sum_{uv} A_{eu}(\hat{\mu}) A_{ev}(\hat{\mu}) \frac{1}{\Delta} \int_{\tau}^{\tau+\Delta} S_u(t) S_v(t) dt - \hat{\mu}_e \quad \forall e \in E, \quad (3.25)$$

where we also introduced the auxiliary matrix  $A_{ev}(\hat{\mu}) = \sum_u B_{eu} L(\hat{\mu})_{vu}^\dagger$  for all  $e \in E, v \in V$ . It can be proved that the forcing term defined in Eq. (3.25) is the natural approximation of the dynamics for the fast time-varying capacities in Eq. (3.22) under the three stated assumptions for the slow capacities. In principle, one could use a numerical integrator to solve the time integral in Eq. (3.25), and therefore solve Eq. (3.24) until convergence. However, we wish to carry out analytical calculations further, and for this task, we restrict the class of functions  $S(t)$ .

In particular, we assume the mass matrix to be changing periodically fast in time, which in formulas reads

$$S_v(t+T) = S_v(t), \text{ for } T/\Delta \ll 1 \quad \forall v \in V, \forall t \geq 0. \quad (3.26)$$

By further decomposing the mass loads with their Fourier decomposition  $S_v(t) = \sum_{n_v \in \mathbb{Z}} c_v^{n_v} \exp(i\omega n_v t)$ , with  $\omega = 2\pi/T$ , and then substituting Eq. (3.26) into Eq. (3.25), we obtain the central result of our derivations. This is

$$\frac{1}{\Delta} \int_{\tau}^{\tau+\Delta} S_u(t) S_v(t) dt = C_{uv} + O(\Delta) \quad \forall u, v \in V \quad (3.27)$$

$$C_{uv} = \sum_{n_v \in \mathbb{Z}} (c_u^{n_u})^* c_v^{n_v} \quad \forall u, v \in V, \quad (3.28)$$

with  $O(\Delta)$  containing small terms  $\varepsilon$  decaying as  $\lim_{\Delta \rightarrow \infty} \varepsilon/\Delta = 0$ . The asterisk indicates complex conjugation.

Using Eqs. (3.27)-(3.28), we define a second slow dynamical system, which dictates the slow time evolution of a second set of capacities as

$$\frac{d\bar{\mu}_e(\tau)}{d\tau} = \bar{\Phi}_e(\bar{\mu}) \quad \forall e \in E \quad (3.29)$$

$$\bar{\Phi}_e(\bar{\mu}) = \frac{\bar{\mu}_e^{2-\gamma}}{w_e^2} \sum_{uv} A_{eu}(\bar{\mu}) A_{ev}(\bar{\mu}) C_{uv} - \bar{\mu}_e \quad \forall e \in E. \quad (3.30)$$

Contrarily to Eqs. (3.24)-(3.25), the slow dynamical system in Eqs. (3.29)-(3.30) does not explicitly depend on  $\tau$ . Thus, its asymptotic solutions can be efficiently extracted by numerically integrating its right-hand side without needing to calculate the integral over  $\Delta$  of the product at each time step between the mass load components. We also note that the adaptation rules in Eqs. (3.29)-(3.30) are not equivalent to the fast ones of Eq. (3.22), with the time loads that are integrated over a  $T$ . This latter case would imply that  $C_{uv} = \tilde{S}_u \tilde{S}_v$ , where  $\tilde{S}_v$  are mass load entries integrated on a period. This is a particular case of the more general expression introduced in Eq. (3.28).

We characterize solutions of the fast dynamics Eq. (3.22) with respect to those of Eqs. (3.29)-(3.30) using semi-analytical arguments.

We experimentally observe that fast capacities  $\mu_e(t)$  undergo two distinct phases when evolving in time. First, there is a transient interval where they largely change their average value; then, they stabilize around a steady value and oscillate. We label with  $t_{\text{STAB}}$  the transition time between these two regimes, and upon experimental inspection of both the fast and slow long-run capacities, we propose for  $t > t_{\text{STAB}}$  the ansatz

$$\mu_e(t) = a_e + b_e(t) \quad \forall e \in E \quad (3.31)$$

$$a_e = \bar{\mu}_e^* := \lim_{\tau \rightarrow +\infty} \bar{\mu}_e(\tau) \quad \forall e \in E \quad (3.32)$$

$$b_e(t) = b_e(t + T) \quad \forall e \in E. \quad (3.33)$$

The fast time-varying fluxes also oscillate in times after stabilizing, i.e., for  $t > t_{\text{STAB}}$ . We use this evidence to deduce that capacities' oscillatory modes are resonant with the squared fluxes precisely

$$b_e(t) = \sum_{n,m \in \mathcal{N}} b_e^m b_e^n \exp[i\omega(n+m)t] \quad \forall e \in E, \quad (3.34)$$

where  $\mathcal{N} = \{n_v\}$  is the set of Fourier modes of the mass inflows. These findings also validate Eq. (3.23) since they suggest that we may neglect

fast oscillatory terms when studying the average properties of asymptotic solutions of Eq. (3.22).

To connect stationary solutions of adaptation equations with optimality conditions, we adopt the same recipe we have followed in the rest of the thesis. This is, we seek a candidate Lyapunov functional for the adaptation rules at hand and, analytically and numerically, show its well-posedness. Here, we propose

$$\bar{\mathcal{L}}_\gamma(\bar{\mu}) = \frac{1}{2} \sum_e \frac{w_e}{\bar{\mu}_e} \bar{F}_e^2(\bar{\mu}) + \frac{1}{2\gamma} \sum_e w_e \bar{\mu}_e^\gamma, \quad (3.35)$$

where the fluxes  $\bar{F}_e(\bar{\mu}) = (\bar{\mu}_e^2/w_e^2) \sum_{uv} A_{eu}(\bar{\mu}) A_{ev}(\bar{\mu}) C_{uv}$  have been introduced.

In our experiments, we experimentally demonstrate that for a wide variety of input mass loads, and therefore matrixes  $C$ , the functional in Eq. (3.35) is monotonically decreasing with respect to the dynamics in Eqs. (3.29)-(3.30). Furthermore, if  $\text{rank}(C) = 1$ , the functional has a provably negative (zero at convergence) Lie derivative along solutions of Eqs. (3.29)-(3.30). As discussed below, such a rank condition on  $C$  is pivotal in determining the emergence of loops in stationary network topologies. A particular case where  $C$  has unitary rank is when the input loads are stationary. Here,  $\bar{\mathcal{L}}_\gamma(\bar{\mu})$  becomes  $\mathcal{L}_\gamma^{M=1}(\mu)$ , i.e., the functional of Eq. (3.15) evaluated for  $M = 1$ .

A key observation is that the functional Eq. (3.35) converges to the long-run running average of

$$\langle \mathcal{L}_\gamma^1(\mu) \rangle_T = \frac{1}{T} \int_t^{t+T} \left( \frac{1}{2} \sum_v p_v(\mu) S_v(t') + \frac{1}{2\gamma} \sum_e w_e \mu_e^\gamma \right) dt'. \quad (3.36)$$

Such a consideration is essential since it allows us to link, under an optimization perspective, networks returned by the dynamics in Eqs. (3.29)-(3.30) to optimal long-run networks that are outputted on average by Eq. (3.22). This connection highlights that our newly derived adaptation

equations are a computationally convenient scheme for finding optimal networks that would otherwise require solving fast dynamic systems for a long time. Numerically integrating Eqs. (3.29)-(3.30) requires the calculation of  $C$  only once, at the beginning of the dynamics, which then has a constant right-hand side.

The last part of the theoretical model development is devoted to answering whether and how the profile of  $S_v(t)$  influences the emergence of loops in optimal transport networks. We analytically prove that the optimal transport network is a tree if  $\text{rank}(C) = 1$  and  $\gamma \leq 1$ . This condition is satisfied, for example, when the mass inflows are stationary, but it also holds in many other less trivial case studies.

We examine the case study where loads are sums of decoupled harmonic oscillators, namely

$$S_v(t) = \sum_{i=1}^{N_v} A_v^i \cos(\omega n_v^i t + \phi_v^i) + d_v \quad \forall v \in V, \quad (3.37)$$

with  $\omega = 2\pi/T$ ;  $n_v^i, N_v \in \mathbb{N}$ ;  $A_v^i, d_v \in \mathbb{R}$ . By construction, Eq. (3.37) returns input loads that are periodic in  $T$ ; hence, we equate them with their Fourier series expansion  $S_v(t) = a_v^0/2 + \sum_{n_v \geq 1} a_v^{n_v} \cos(\omega n_v t + \varphi_v^{n_v})$ . Putting these two expressions together gives the important relation

$$c_v^{n_v} = \frac{A_v^i}{2} \exp(i\phi_v^i) \delta_{n_v n_v^i} \quad \forall v \in V, \forall n_v \in \mathbb{N}, \quad (3.38)$$

where we conventionally set  $\phi_v^0 = 0$  for all  $v \in V$  and where only a finite set of Fourier coefficients is non-zero since the sum in Eq. (3.37) is finite.

Our main result is translating the condition  $\text{rank}(C) = 1$  to a relation written with the harmonic oscillators' amplitudes, modes, and phases. Starting from Eq. (3.37), we find that for  $\gamma \leq 1$ , stationary solutions of Eqs. (3.29)-(3.30) are trees if

1.  $\phi_v^i = \phi_u^i + k\pi$  for  $k \in \mathbb{Z}$  and for all  $u, v \in V$ , i.e., sources and sinks are in phase;



2.  $A_v^i \delta_{n_v n_v^i} = \lambda(-1)^k A_u^i \delta_{n_u n_u^i}$  for  $\lambda \neq 0$  and all  $u, v \in V$  (yielding  $N_v = N$ ), i.e., amplitudes are non-zero for the same modes.

Numerically, we also show that the rank of  $C$  can be employed as a proxy for the total number of loops in the optimal transport network and, therefore, quantify its robustness. In fact, by examining a series of semisynthetic experimental setups, we see that the dimension of the cycle basis monotonically increases with the rank, where a basis for cycles is a minimal collection of cycles that allows expressing any other cycle as an “exclusive or” of the edges.

### 3.1.4 Bilevel optimization for traffic mitigation

The models introduced above assume one main objective in designing the transportation network. This takes into account the cost to operate the network and that to build it, with their relative contribution tuned by a parameter. From a network manager’s point of view this is a meaningful trade-off that allows to select globally optimal networks. In practice though, individual passengers may not follow the ideal scenario represented by optimal flows at convergence. Rather, they may move greedily from multiple entry nodes to multiple exit ones, as in the shortest path-like unicommodity routing model of Section 3.1.1. This routing mechanism potentially triggers traffic congestion since the total flux on some edges may be larger than a prefixed threshold. Therefore, how do we model the competition between greedy mass routing and an agent that tunes edge weights to reroute fluxes to mitigate congestion? Additionally, what effect does this competition have on optimal network topologies?

An answer is provided by bilevel optimization, as proposed in Lonardi and De Bacco [6], the work we discuss. To hone the intuition on the model setup, one could think of it as the competition between greedy passengers wanting to reach their destinations on a highway while lowering expenses (a behavior that is experimentally observed and referred to as Wardrop’s first principle [116–119]), and a road manager that sets road tolls to avoid traffic.

These types of problems have been tackled in the literature with a variety of methods, including message passing [103, 120–122], Markov chain Monte Carlo [123], cellular automata [124, 125], and heuristic routing strategies [126].

Our approach is to develop adaptation rules for extracting fluxes that balance transport efficiency with link congestion. In detail, we alternate the unicommodity adaptation equations with a Projected Stochastic Gradient Descent (PSGD) scheme to update the weights. Remarkably, we can derive closed-form gradients for the PSGD by making physics-inspired assumptions. We study the networks' profiles at convergence and examine how they relate to global network metrics, such as the transportation and congestion costs or the Gini coefficient of fluxes over the edges. We also empirically observe that the uncoordinated action of the agent (network manager) and the greedy routing of mass (passengers) increases the Price of Anarchy [127], measured as the total travel time needed to move from every entry node to every destination.

We frame the optimization problem for greedy routing using its formulation in Eqs. (3.9)-(3.10). However, in the following discussion, we introduce a slight change of notation that does not change the underlying model but can aid in understanding the mathematical framework when considering the case study of passenger routing in a transportation network. In detail, we introduce Origin-Destination (OD) indexes  $r = 1, \dots, R$  to specify different mass inflows that travel each from their origin to their destination greedily. It is essential to distinguish fluxes  $F_e^r$ , which represent the displacement of mass of type  $r$  and are independent for  $r' \neq r$ , and fluxes  $F_e^i$  of Section 3.1.2. The latter are components of multicommodity fluxes that jointly minimize a unique objective, as in Eq. (3.19). This distinction highlights a crucial conceptual difference between commodity indexes and OD ones. Commodity indexes are more general than the second, being, in principle, commodities not only represent the entry and exit nodes of mass but also any relevant information about the problem at hand. For example, commodities can be wavelengths in optical communication networks [103] or RGB color channels of images [5].

Thus, we seek for fluxes minimizing the convex cost  $C = \sum_{e,r} w_e |F_e^r|$ , while satisfying Kirchhoff's law,  $\sum_e B_{ve} F_e^r = S_v^r$  for all  $r = 1, \dots, R$ . The cost  $C$  is convex; its solution is the overlap of the  $R$  shortest path between all OD pairs.

The crucial difference with all models considered in the rest of the thesis is that the greedy routing task is only the lower-level problem in a larger bilevel optimization setup, where the agent's action on the weights is formalized in the upper-level problem. We assume that fluxes trigger congestion if larger than a critical value, i.e.,  $\sum_r |F_e^r| \geq \theta$ . Hence, we conveniently introduce the variable  $\Delta_e = \sum_r |F_e^r| - \theta$  and quantify congestion by defining the minimization problem

$$\min_{w \geq \varepsilon} \left\{ \Omega = \frac{1}{2} \sum_e \Delta_e^2 H(\Delta_e) \right\}, \quad (3.39)$$

where  $H$  is the Heaviside step function, and  $\varepsilon > 0$  is a small threshold that guarantees the network's Laplacian well-posedness. The multiple solutions of Eq. (3.39) are edge weights that minimize the quadratic loss  $\Omega$ , thus lowering congestion. Other congestion costs are studied in the literature, for example, the Hinge loss [120, 128].

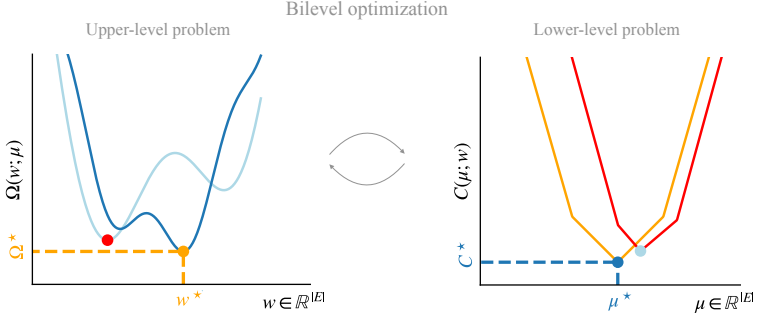
The upper and lower-level problems are put together into a unique bilevel minimization problem, reading

$$\min_{w \geq \varepsilon} \Omega(w; \tilde{\mu}) \quad (3.40)$$

$$\text{s.t. } \tilde{\mu} = \underset{\mu \geq 0}{\operatorname{argmin}} C(\mu; w), \quad (3.41)$$

where we made explicit all the problem's variables (before the semicolon) and parameters (after the semicolon), and with the equality in Eq. (3.41) coming from the convexity of  $C$ .

With Eqs. (3.40)-(3.41), we prescribe the optimization task we introduced in words and wish to model. This is, Eq. (3.41) is optimized by  $\tilde{\mu}_e$ , linked to its respective greedy fluxes. The capacities  $\tilde{\mu}_e$  implicitly constrain the upper-level problem in Eq. (3.40), where edge weights are tuned to minimize



**Figure 3.4: Sketch of the bilevel optimization problem.** The optimal solutions of a bilevel optimization simultaneously minimize the upper- and lower-level optimization problems. For the specific case of traffic mitigation, the optimal capacities  $\mu^*$  (dark blue) are minimizers of the transport cost  $C$  and return the shortest path fluxes. These enter as constraints in  $\Omega$ , which is minimized by  $w^*$  (orange). In turn, the weights parametrize the profile of  $C$ . Given the rugged landscape of  $\Omega$ , we update the weights stochastically to converge to different local minima of  $\Omega$  (red), which are associated with their optimal capacity (light blue).

congestion caused by the fluxes. The edge weights re-enter the lower-level problem of Eq. (3.41) as a constraint. Solutions of this joint minimization process are optimal minimizers of Eqs. (3.40)-(3.41). We schematically represent the model setup in Fig. 3.4.

The method we use to search for minimizers is alternating adaptation rules to update  $\mu_e$  and the PSGD scheme for  $w_e$ . As explained in Section 3.1.1, greedy routing can be solved using a set of equations as in Eq. (3.4), which update the edge capacities, giving the fluxes  $F_e^r = F_e^r(\mu^r)$ . Here, we update  $R$  independent adaptation rules at all time steps, corresponding to the  $R$  OD indexes.

We derive closed-form gradients of  $\Omega$  to update the weights. In particular, we exploit linear algebra insights to treat non-trivial dependencies on the

network Laplacian [129], and make use of Hagen-Poiseuille law in Eq. (3.2) to find

$$\frac{\partial \Omega}{\partial w_e} = \sum_{e' \in E'} \Delta_{e'} \sum_r \frac{F_e^r}{w_e} \left( \frac{\mu_{e'}^r}{w_{e'}} \text{sgn}(F_{e'}^r) G_{e'e}^r - \text{sgn}(F_e^r) \delta_{e'e} \right) \quad \forall e \in E, \quad (3.42)$$

where  $E'$  is the set of congested edges with  $\Delta_e > 0$ , and  $G_{e'e}^r = \sum_{vu} B_{ve'} B_{ue} (L(\mu^r/w^r)^\dagger)_{vu}$ .

The weight update is performed using a dropout mask that sets to zero a random number of entries of the gradients in Eq. (3.42). We add stochasticity to explore the rugged minimization landscape of the bilevel optimization problem. Moreover, we apply a projection step to SGD to constrain the weights to their feasibility set defined by  $w_e \geq \varepsilon$  for all  $e \in E$ . For this, we clip the weights since such a simple expedient is effective in the case of linear constraints as in Eq. (3.40). If one had to deal with complex non-linear constraints, a possibility would be to add a cleverly designed momentum term to the adaptation equations to satisfy the problem's constraints at infinite times [130, 131].

We validate our method for bilevel optimization on different synthetic networks. Our findings show that the algorithm effectively trades off traffic congestion, encoded by  $\Omega$ , against the transport efficiency, encoded by  $C$ . By choosing different dropout masks for the stochastic update of the weights, we can explore this trade-off and showcase several transport network topologies that counterweight less over-trafficking with higher transport costs and vice versa.

We delve deeper into the nature of traffic congestion by computing the Gini coefficient of the flux distribution and estimating the total travel time using an affine latency model, which sets an affine dependence between fluxes and travel time on edges [120, 132]. An essential finding of this analysis is the observation that the total travel time and the Gini coefficient are worse when the network manager tunes the weight but does not cooperate with the passengers than when they do not take any action. This happens because, by excessively lowering the cost of an edge, passengers may

largely reroute onto it, triggering more congestion than they would have had initially. Such a phenomenon is similar to Braess’s paradox [133], which is theoretically explained and empirically verified in many physical systems from power grids [134, 135] to urban transportation [136, 137].

## 3.2 Applications

After developing the mathematical foundations for various approaches to address optimal transport problems on networks through the lens of adaptation equations, we proceed to describe several concrete applications. Here, the problem’s variables assume different interpretations that may broadly differ based on the task. Therefore, to clarify the modeling setup, we discuss such interpretations for each application before going into the details of our experiments. We also stress once more that these applications are the primary motivator for the theoretical developments presented in Sections 3.1.1-3.1.4, with their connection precisely outlined in Section 2.2. For this reason, the thesis has enriched significance when these two parts are taken in unison.

Our experiments mainly deal with engineering problems, specifically designing urban transportation networks. In particular, we apply our methods to the Paris Métro [138, 139], the bus and tram system of Bordeaux [138], and the International E-road network [43]. In the last part of the discussion, we shift our focus to machine learning, and we use multicommodity adaptation equations for supervised image classifications on two image datasets [140, 141].

### 3.2.1 Paris Métro

We use multicommodity routing to study passengers’ trajectories in the Paris Métro [138, 139]. The network used for our experiments has  $|V| = 302$  nodes and  $|E| = 359$  edges. The central modeling assumption of this application is that the commodity index  $i$  distinguishes groups of passengers entering from different stations, therefore,  $i \in V$ . The passengers’ groups

are the commodities that interact along the network edges and trigger traffic congestion. Naturally, other modeling choices can be made. However, separating passengers by their origin station allows us to effectively capture congestion phenomena in real-world setups where unicommodity adaptation equations fall short of achieving that.

Under this assumption, the variables entering the multicommodity adaptation equations can be interpreted as physical quantities directly connected to transportation networks. In detail, the mass matrix  $S_v^i$  contains the rates of passengers entering and exiting stations, which are the network nodes. Specifically, the mass matrix has positive entries in  $i = v$ , negative ones for all  $v \neq i$  where passengers exit, and zero otherwise. Passengers move in between stations and generate a multicommodity flux  $F_e^i$ ; components of the flux represent displacements of passengers entering from  $i$  along  $e$ . Crucially, the commodity-independent capacity  $\mu_e = \mu_e^i$  for all  $i$  couples the passengers together and separates the multicommodity dynamics from the unicommodity dynamics with multiple sources and multiple sinks of Section 3.1.1. One can think of such a capacity as the capability of the infrastructure, e.g., the rails, to allocate passengers traveling in the network. In this application, the weights  $w_e$  are the Euclidean lengths between the nodes' geographical coordinates.

We use the multicommodity dynamics of Eqs. (3.13)-(3.14) by both coupling commodities with their 2-norm as in Eq. (3.14), and by using their 1-norm squared,  $f_e(F) = (\sum_e |F_e^i|)^2$ . We adopt two different couplings because the former (the 2-norm) gives a series of theoretical guarantees, as discussed in Section 3.1.2, which allows us to prove that networks outputted by adaptation rules are provably optimal. Meanwhile, the latter (the 1-norm) is arguably a more natural choice to measure the total occupancy of a link, which is the total number of users traveling onto it rather than the sum of its squares. The 1-norm is squared since, by doing so, we can formalize a mapping to a minimization setup admitting a constrained and an unconstrained formulation as in Eqs. (3.16)-(3.17) and Eqs. (3.19)-(3.20), respectively. The critical difference between the two norms is that for the 2-norm, we find a suitable Lyapunov functional associated with the multicommodity dynamics, while for the 1-norm, we do

not draw this connection; hence, we cannot analytically prove the optimality of its asymptotic networks. Nevertheless, we experimentally observe that the multicommodity cost  $C_\gamma^M$  in Eq. (3.19), with  $f_e$  being the squared 1-norm, decreases along trajectories of the multicommodity dynamics, and converges to an energy plateau.

We study optimal transportation networks returned by different passengers' input configurations. Particularly, we extract real passengers' entry and exit data [139] and use them to create semisynthetic mass matrixes  $S_v^i$  with inflows interpolated between the real ones taken from the data and homogeneous ones. Passengers entering each node  $i = v$  are assigned to outflow  $u$  with the same rate as the inflowing passengers of  $u$ , with a proper renormalization to ensure mass conservation. This way, passengers exit at high rates from stations where they enter at high rates, and vice versa for small ones.

In our experiments, we observe that the 1-norm and the 2-norm output comparable transport networks when the critical exponent  $\gamma$  of Eq. (3.13) distributes fluxes, i.e.,  $\gamma > 1$ . However, their transport networks differ if fluxes are aggregated  $\gamma < 1$ , with the 1-norm congesting more fluxes over a smaller amount of edges than the 2-norm. To better quantify the extent of congestion, we measure the Gini coefficient [109] of the flux distribution over the edges. A higher Gini coefficient corresponds to transport networks where passengers primarily concentrate on a few links, a low one to those where fluxes homogeneously distribute on the whole infrastructure. The 1-norm returns a higher Gini than the 2-norm for the same mass matrix and with the same  $\gamma$ .

We also measure the fraction of idle edges, i.e., the number of tracks unused by passengers, which exhibits a sharp first-order phase transition at  $\gamma = 1$ . This metric stays constant at zero for all  $\gamma > 1$  and starts to grow from  $\gamma \leq 1$ . Such a profile further confirms the effect of the critical exponent  $\gamma$  on the dynamics, i.e., concentrating and distributing passengers over the network edges. While we observe a phase transition for a fraction of idle edges, it is important to remark that for  $\gamma > 1$ , the interaction between commodities still prompts loops' formation. This effect is unobserved for unicommodity adaptation rules.



We conclude our study on the Paris Métro by probing its robustness to structural failures. Precisely, we iteratively remove the four largest stations of the network in descending order, measured by their number of yearly entry passengers. These are: Châtelet, Gare du Nord, Saint-Lazare, and Gare de Lyon. For each network with failures, we posit that passengers who would have entered a now-removed station instead enter neighboring stations, with their entry rates set to be proportional to the size of the respective neighboring station. The main takeaway of our analysis is that when  $\gamma$  is low, and thus when traffic congestion is mainly caused by node removal, stations trigger traffic with varying levels.

In detail, removing the Châtelet station causes the fluxes' Gini coefficient to abruptly jump, i.e., passengers concentrate more than in the original network. Removing Gare du Nord does not affect congestion since passengers entering it now move to its neighboring station, Gare de l'Est, and do not majorly change their trajectories. Removing the station of Saint-Lazare causes the Gini coefficient to jump again. This station appears crucial for linking the city center with the city's North area. Conversely, removing Gare de Lyon does not critically affect congestion.

### 3.2.2 Bordeaux's trams and buses

#### Multicommodity routing

The transportation network of Bordeaux considered in this first application is a multilayer network with  $L = 2$  layers [138]. The first layer,  $a = 1$ , corresponds to the city's roads where buses move, and it consists of  $E_1 = 2347$  intra-layer bus edges. The second layer,  $a = 2$ , represents the city's tram rails and comprises  $E_2 = 112$  intra-layer tram edges.

We study this network using a variation of the multicommodity adaptation equations, which generalizes the model described in Section 3.1.2 to multilayer networks. There are two important modeling assumptions we make. The first is taking a layer-dependent critical exponent  $\gamma_a$ , whose value may differ on layers  $a = 1$  and  $a = 2$ . For instance, we could concentrate fluxes on the tram by taking  $\gamma_2 < 1$  while diluting them on the bus layer with  $\gamma_1 > 1$  (for inter-layer edges, we fix  $\gamma = 1$ ). The second is that we

assign edge weights as the Euclidean length between nodes' geographical coordinates,  $\ell_e$ , multiplied by a layer-dependent coefficient representing vehicles' (inverse) velocity,  $\nu^a$ . In equations, this is  $w_e = \nu^a \ell_e$  for all  $e \in E_a$ , with  $a = 1, 2$ . Concretely, suppose  $\nu^a$  is set to a lower value on layer  $a = 2$  than on  $a = 1$  to signify the higher speed of trams relative to buses. In that case, passengers find it more convenient to cover the same geographical distance on layer  $a = 2$  than on layer  $a = 1$ . The weights of inter-layer edges are set to be negligible compared to the others, so passengers move between layers without additional costs. This choice can be thought of as not adding extra cost for commuting between buses and trams.

These modeling assumptions translate to the minimization setup associated with the adaptation rules. In particular, we map the multilayer-multicommodity adaptation rules to a suitable optimization framework using the scheme of Section 3.1.2. This is, we formulate energy minimization constrained and unconstrained problems that are solved by the adaptation rules, which asymptotically converge to the minimum of a Lyapunov functional, being exactly the unconstrained minimization problem's objective. The mapping leads us to find [1]

$$C_{\gamma_a}^M = \sum_{a=1}^L \sum_{e \in E_a} \nu^a \ell_e f_e(F)^{\Gamma(\gamma_a)} \quad (3.43)$$

with  $f_e(F) = \sum_i (F_e^i)^2$  and  $\Gamma(\gamma_a) = \gamma_a / (1 + \gamma_a)$ , which is the natural generalization of the multicommodity cost in Eq. (3.10) under the assumptions for multilayer networks discussed above. Minimizers of Eq. (3.43) are found by solving the multilayer-multicommodity adaptation equations and by using its asymptotic capacities to extract the fluxes as  $F_e^{i*} = F_e^i(\mu^*)$ .

We run our algorithm by inputting a mass matrix  $S$  that contains monocentric passenger inflows. This means that we simulate the scenario where passengers uniformly enter the network nodes and exit from a single exit at the center of the city [142]. Practically, we assign positive entries  $S_v^i = +1$  to a different node  $v$  per commodity  $i$ , which is the station passengers of

type  $i$  enter, and negative entries  $S_u^i = -1$  to a node  $u$  that is identical for all commodities.

Our experiments are performed with  $\nu^{a=2} = 0.2$ , i.e., supposing that traveling by tram is faster than traveling by bus. Additionally, we fix  $\gamma_2 = 0.5$  and  $\gamma_1 = 1.5$ . We observe that passengers distribute over the whole bus infrastructure while predominantly concentrating on a few central tram lines.

We compare the multilayer network of the city with a single-layer network where only the bus edges are present. The latter experimental setup is obtained by setting  $w_e$  to a large value for all  $e \in E_1$ . Our central finding is that the tram lines are crucial for decongesting the city's roads. In particular, we observe that adding the tram to the city's transportation infrastructure allows for reducing the total number of passengers traveling with buses by 17%. We also measure the Gini coefficient of the flux distribution on the bus network. This increases when trams are added to the system since passengers do not use many previously largely and uniformly populated edges.

### Routing with time-dependent loads

In this second application, we consider Bordeaux's bus network a single-layer network. The goal is to test and explore the loop formation mechanism derived in Section 3.1.3 on real-world data. That is, we want to validate that loops emerge in the transport network when passenger entry and exit time-dependent inflows are not in phase with each other.

To this end, we devise a first experimental setup to randomly extract two sources,  $v_1$  and  $v_2$  and five sinks among the network nodes. Then, we design the loads' profiles  $S_v(t)$  so that their Fourier transforms produce the forcing matrix  $C$  in Eq. (3.28) so that, in one case, (i)  $\text{rank}(C) = 1$ , and in another case (ii)  $\text{rank}(C) = 2$ . For the two case studies, we set the sources to be (i)  $S_{v_1}(t) = S_{v_2}(t) = 100 \cos(\omega t)$ , with  $\omega = 2\pi$ , and (ii)  $S_{v_1}(t) = 100 \cos(\omega_1 t)$  for one source and  $S_{v_2}(t) = 100 \cos(\omega_2 t)$  for the other, with  $\omega_1 = 2\pi$  and  $\omega_2 = 4\pi$ . We set in both cases  $S_v(t) = -[S_{v_1}(t) + S_{v_2}(t)]/5$  to ensure conservation of mass at all times. Notice that the terms "sources"

and “sinks” here are misused since the values attained by the mass matrix on these nodes oscillate between positive and negative. However, at each time  $t$  when the sources’ loads are positive, the sinks’ are negative, and vice versa.

When running the slow dynamics in Eqs. (3.29)-(3.30) with  $\gamma < 1$ , we expect the transport network at convergence to be a tree for (i) and to possibly contain loops for (ii). Our numerical experiments confirm the theory with  $\gamma = 0.9$ .

We perform a second larger validation experiment by setting  $S_v(t) = \sum_{k=1}^n S_v^k(t)$  where  $S_v^k(t) = (100/|Q_n|) \cos(\omega kt)$  on a set of  $Q_n$  of randomly extracted nodes, while  $S_v^k(t) = -[100/(|V| - |Q_n|)] \cos(\omega kt)$  on the remaining ones. The modes range as  $n = 1, \dots, 6$ , while the number of nodes extracted for each mode is  $|Q_n| = \{1, 5, 10, 20\}$ . Again,  $\omega = 2\pi$ . This particular problem’s construction produces Fourier coefficients’ matrixes  $C$  for which  $1 \leq \text{rank}(C) \leq 6$ .

Numerically integrating the adaptation rules in Eqs. (3.29)-(3.30) with  $\gamma = 0.9$  on this setup shows that the fraction of basis loops generated by the passengers’ fluxes is monotonically increasing with  $\text{rank}(C)$ . Furthermore, consistently with our theory, no loops are present when  $\text{rank}(C) = 1$ . This experiment hints that  $C$  can be employed as a proxy to infer the number of loops in the optimal transport network, i.e., its robustness.

### 3.2.3 International E-road network

We apply the adaptation scheme for bilevel optimization, as discussed in Section 3.1.4, to the International E-Road network [43]. In our study, we simulate the action of a network manager tasked with assigning tolls to the network edges by adjusting their weights ( $w_e$ ) to alleviate traffic bottlenecks. This task prompts competition with greedy passengers who aim to travel from their origin nodes to their destinations by minimizing travel costs. The two problems are coupled into a bilevel optimization formulation, as in Eqs. (3.40)-(3.41).

The solution to this problem is found by alternating adaptation equations for capacity updates, which are controlled by passengers, with edge weights

tuning. The capacities enter as parameters into the PSGD scheme devised to minimize traffic congestion. In turn, the weights get updated and prompt passengers' rerouting.

All variables entering the problem have similar physical interpretation to what is discussed in Section 3.2.1 or Section 3.2.2. However, here, it is helpful to think of the edge weights as tolls the users have to pay to travel on a road; additionally, each capacity  $\mu_e^r$  (the capacities now are  $R$ , one per greedy group of passengers) may be interpreted as the space one should allocate on a road to let passengers of group  $r$  flow freely. We stress that the theory developed in Section 3.1.4 is independent of such choices, and it is valid for different modeling tasks, significantly far apart from engineering.

The E-Road network connects cities around Europe with highways; in total, it comprises  $|V| = 541$  nodes and  $|E| = 712$  links. We assume that passengers are divided into  $R = 15$  greedy groups, each entering the network from a different large city, with rates proportional to the city's population. Similarly to what is done in Section 3.2.1, for a particular entry node  $r$ , we assign outflows to be proportional to the exit cities' populations. This way, if a node has a high volume of passengers entering it, it also has a large volume of passengers exiting. The total number of passengers to be routed is approximately 30 million.

We compare our bilevel optimization scheme algorithm against two baseline methods. The first consists of routing the greedy passengers using  $R$  independent systems of adaptation rules as in Eq. (3.4), with fixed edge weights that are  $w_e = \ell_e$ , where  $\ell_e$  are the Euclidean distance between cities' geographical coordinates. This method simulates the scenario where the network manager does not act on the system. The second scheme involves tuning the edge tolls after assigning fixed shortest path capacities to the edges. Practically, we find the shortest path fluxes connecting passengers' entry and exit nodes with Dijkstra's algorithm [40]. Then, we use the optimal capacities associated with such fluxes as parameters and run the PSGD scheme for updating weights until convergence. We extract the final passengers' fluxes with adaptation equations using the weights at convergence as parameters. The scope of this procedure is to simulate the action of an uninformed network manager that tunes the edge weights while

disregarding passengers' rerouting. Indeed, information on passengers' paths enters the manager's problem only as an initial condition.

Our findings highlight that adequately considering traffic rerouting in the edge tuning task of the network manager is crucial to trade off traffic congestion against the efficiency of transport efficiently. In particular, solving Eqs. (3.40)-(3.41) by alternating a PSGD scheme and adaptation equations returns distributed networks where passengers can move efficiently while traffic is mitigated. In contrast, both setups with no action from the network manager and where the network manager is uninformed give transport networks with high congestion levels.

We quantify congestion by computing the average travel time of users using an affine latency model [120, 132]. In equations, the average travel time is defined as  $\langle T_\theta(s) \rangle = \sum_{er} t_{\theta,e}(s) |F_e^r| / \sum_{er} |F_e^r|$ , where the over-trafficked edges for which  $\sum_r |F_e^r| > \theta$  take  $t_{\theta,e} = \ell_e [1 + s(\sum_r |F_e^r| - \theta)/\theta] / v^\infty$  to be traversed, and the non-congested ones  $t_{\theta,e}(s) = \ell_e$ . Here,  $s$  is a sensitivity parameter, while  $v^\infty = 100$  (km/h) is the free flow speed of passengers. To dimensionalize the model properly, all quantities are to be multiplied by unitary constants.

The time traveled by passengers whose routes are extracted by alternating adaptation rules and the PSGD scheme is substantially lower than that found with the other two baseline methods. In particular, for low sensitivity ( $s = 1$ ) the bilevel optimization scheme gives  $\langle T_\theta(s) \rangle \simeq 1.7$  (hours). The greedy routing method with no intervention from the network manager returns  $\langle T_\theta(s) \rangle \simeq 2.3$  (hours). In contrast, the scheme with the uninformed network manager outputs  $\langle T_\theta(s) \rangle \simeq 3.1$  (hours). The differences in travel time become even starker for a high sensitivity value ( $s = 5$ ).

We also observe that travel time is higher when the network manager tunes the edge weights while being uninformed about the passengers' routes, as opposed to when there is no intervention. As discussed in Section 3.1.4, this evidence can be interpreted similarly to Braess's paradox. Specifically, it shows that the Price of Anarchy, which measures how much a system degrades due to the greedy behavior of its users, is higher when the network manager intervenes and gets an effect opposite to what was initially intended.

### 3.2.4 Image classification

In this last part of our discussion, we shift the focus from engineering problems to machine learning. More in detail, we employ adaptation equations to perform supervised image classification. Our main goal is to accurately categorize images into their classes, assigning them a label. To do so, we compare them with other pictures of which we know the label, hence, the category they belong to. Using the multicommodity adaptation equations of Section 3.1.2 together with a thoughtfully devised experimental setup that allows us to leverage color information of images to boost classification performance, we carry out such a task using the optimal multicommodity transport cost  $C_\gamma^M(F^*)$  in Eq. (3.19) as a measure of similarity between images.

Optimal Transport-based schemes are widely used for image classification [39, 68, 69, 79, 143–145]. The underlying idea of these methods is to extract the 1-Wasserstein distance between two distributions efficiently—representing the two images that are compared—and use it as a proxy to measure their similarity: If the distance is lower, then the images are similar since intuitively “it does not take much effort to move one distribution into the other”. This approach proved itself to be accurate and holds several technical advantages. For instance, the 1-Wasserstein distance is more robust over domain shift for train and test data than other metrics, such as the Kullback-Leibler divergence [68]. Additionally, it provides well-defined and meaningful gradients to learn data on non-overlapping domains, which is the crucial feature at the core of Wasserstein GANs [83]. Because of these, and many other desiderata, a lot of effort has been put into reducing the computational cost for computing the 1-Wasserstein distance on large-scale problems [75–79].

Our experiments complement these approaches by integrating image colors directly in the OT formulation. In particular, we assess whether leveraging the physical intuition of using colors as immiscible fluxes can boost classification performance.

Our experimental setup is as follows. We consider  $M = 3$  commodities, denoting the RGB color distributions composing each colored image. In

detail,  $g_v^1$  are the vectorized entries of the red pixels' intensity of an image, and similarly,  $g_v^2$  and  $g_v^3$  are the green and blue ones. These color distributions must be transported through a network that connects all pixels of one image to all the pixels of another. The crucial modeling assumption is constructing the network connecting the two pictures and, more in detail, the weights  $w_e$  assigned to its edges. These have to be chosen so that comparing images of the same class gives low transport cost, i.e., high image similarity, and vice versa for images of different classes.

To build the transport network's topology between two images,  $g_v^i$  and  $h_v^i$ , that are transported one into the other, we start from a complete bipartite network between their pixels. In principle, this network already allows us to solve the adaptation rules of Eqs. (3.13)-(3.14) to find the multicommodity transport cost. However, this would come at a high computational price since the number of edges of the network scales poorly as  $|E| = O(|V|^2)$ .

For this reason, we use two expedients; the first is to remove transport edges whose cost is above a tunable threshold  $\tau > 0$  [68, 69]. There are various advantages to doing so; from an intuitive standpoint, humans perceive distances as saturated distances [146]; therefore, using a thresholded ground cost could produce results that arguably match human intuition. Most importantly, many color distributions of images are noisy and heavy-tailed. Thus, thresholding leads to discarding better outliers. Moreover, thresholding boosts the accuracy and speed of OT [69]. To integrate this expedient into the model construction, we assign weights

$$w_{e=\{u,v\}} = \min\{(1 - \theta)\|v_u - v_v\|_2 + \theta\|g_u - h_v\|_1; \tau\} \quad (3.44)$$

to those edges that are not removed from the bipartite network. In Eq. (3.44), we take the 2-norm of the pixels' coordinates and the 1-norm of the images' color distributions. We weigh these two quantities in a convex combination controlled by a parameter  $\theta$ , which we choose with cross-validation. Importantly, this construction gives  $|E| = O(|V|)$ .

The second expedient is to relax the optimization problem in Eq. (3.19) with a penalty for unbalanced OT as in [69]. The idea is to allow the total masses  $\sum_{vi} g_v^i$  and  $\sum_{vi} h_v^i$  to be different and to penalize



images’ pairs proportionally to their net difference in color, namely,  $P = (\max_e w_e/2) \sum_i |\sum_v g_v^i - \sum_v h_v^i|$ . This way, images with largely different color distributions, e.g., dark and light ones, are unlikely labeled with the same class. This penalty is embedded into the network construction by adding an auxiliary “transshipment” edge with cost  $w_e = \max'_e w'_e/2$  (the maximum is taken over the edges of Eq. (3.44)) that is connected to all images’ pixels. A mass equal to the net difference between the images’ color intensities exits from such a node so that the final network is isolated.

We test our methods on two datasets with images of flowers in an unsupervised environment and fruits in a supervised one (white background) [140, 141]. We compare our multicommodity algorithm against its unicommodity counterpart, where the distributions to be transported are the gray-scale intensities of the images. Also, we benchmark performance against a stabilized implementation of Sinkhorn’s algorithm [147] executed on gray-scale and RGB images. For colored images, we average the  $M = 3$  OT costs outputted by Sinkhorn’s algorithm, one per color channel, to measure the images’ similarity.

Our findings highlight that the multicommodity adaptation rules allow us to classify images better than their competitors. In particular, they give a substantial improvement in terms of classification accuracy. As one could expect, we also observe that methods that classify colored images have higher accuracy than those that classify gray-scale ones. We assess the stability of our methods’ prediction by measuring its sensitivity on flower images. The sensitivity is the ratio of true positive samples over the sum of true positive and false negatives. Our results exhibit how the multicommodity adaptation rules have higher sensitivity than Sinkhorn’s algorithm for most classes.

We perform additional experiments where we focus on a subset of all the classes of the fruit images. Here, our results are tailored to show how the multicommodity adaptation equations can better cluster images with similar colors and shapes. This task is more challenging when colors are aggregated and not considered by the OT setup, i.e., in the unicommodity formulation. Additionally, we observe that when the shapes of fruits play

### 3 Results and discussion

a central role in distinguishing samples, both the unicommodity and the multicommodity equations give good classification results.

# 4 Conclusion and future perspectives

## Content

- Chapter 4 —
1. Broad summary of the contributions;
  2. Detailed scientific advancements;
  3. Limitations and possible future research directions.

The efficiency of transport networks is pivotal in numerous real-world applications, emphasizing the importance of devising systematic approaches to address this challenge.

This thesis tackles the problem by formulating a series of methods rooted in biologically-inspired adaptation rules. Specifically, we frame ordinary differential equations that, together with conservation laws, control the temporal evolution of transport networks to return optimal networks upon convergence. We use OT theory to measure optimality. Specifically, we seek a well-defined Lyapunov functional for our dynamical systems—representing the energy well into which adaptive networks fall, and link it to the cost associated with transporting mass along the network edges. This cost is the objective minimized in OT. Such a connection is not only of theoretical significance, allowing us to prove that adaptive networks are provably optimal at convergence, but it also facilitates the connection and validation of adaptation equations with and against classical OT algorithms.

In detail, we extend classical adaptation equations for unicommodity (greedy) mass routing to various modeling setups. First, we consider

the multicommodity scenario, where different types of mass move along the edges of a network. These contribute to minimizing one unique cost being coupled with a shared capacity. To guarantee that multicommodity adaptation equations are well posed, i.e., they reduce the cost of transporting mass so that a formal connection with OT can be drawn, choosing an adequate coupling for the mass types is crucial. We apply this model to study traffic on the Paris Métro and the streets of Bordeaux. Among other findings, our studies exhibit what stations are essential to alleviating congestion in Paris when targeted node failures occur and showcase how the trams of Bordeaux decongest the city’s bus network. We also use multicommodity adaptation rules to perform supervised image classification. Our methods outperform other OT-based algorithms in terms of classification accuracy.

We then formulate adaptation rules to model networks’ evolution with time-varying mass loads. Here, we consider two time scales for the problem’s variables. Namely, we assume mass loads to evolve rapidly in time, whereas capacities, i.e., the network infrastructure, change slowly. This assumption is crucial for deriving closed-form adaptation equations to govern the evolution of the optimal networks. By further assuming that mass loads are periodic, we formally connect the temporal profiles of mass loads with the properties of optimal networks. In particular, we show how the resonance of loads prompts loop formation. We apply this method to Bordeaux’s bus network.

We conclude by modeling the competition between greedy passengers and a network manager controlling traffic mitigation in a bilevel optimization problem. We couple adaptation rules with a Projected Stochastic Gradient Descent scheme that tunes edge weights to enforce decongesting edges. By applying our method to passengers’ routing on the International E-road network, we demonstrate how an informed tuning of tolls can shorten travel times, consequently reducing carbon emissions from car travel.

Our methods are all complemented by open-source codes, which make the results presented in this thesis fully reproducible [91–95].

Naturally, our analyses spark many future research questions in disparate directions, of which we present only a few. First, it is essential to remark that while our models offer a principled mechanism to design transport

networks, in reality, passengers move unpredictably, and the theoretical assumptions made in our works may not hold when validating our methods against traffic data. Extending our models to integrate other real-world assumptions and understanding the difference between networks predicted by our methods and traffic flows extracted, for instance, with digital traces, is an avenue for future research.

Additionally, the broader scope of this thesis is to build a set of tools for optimal network design. These can be relevant to policymakers interested in understanding whether an infrastructure meets the needs of its users. To make such studies robust, it is crucial to complement the insights of our methods with detailed knowledge of the urban development of a geographical area. We aim to bridge this gap in Lonardi *et al.* [8], focusing on the city of Copenhagen.

Going more in detail into the technical aspects of the models presented in the thesis, we propose the following ideas for future research questions. Multicommodity routing has been addressed with several methods, among which belief propagation [103–108]. In particular, in Tai and Yeung [104] the authors showed that coordinated passengers’ routing reduces travel costs by 66% on the England highway network with respect to the case where passengers travel greedily. We could ask ourselves if a similar result can be recovered with adaptation equations. Particularly, by comparing against each other multicommodity adaption equations, that allow us to optimize the joint cost of all mass types moving along the network, and standard unicommodity (greedy) ones.

A possible extension for the routing model with time-dependent loads would be to drop the assumption that the input loads and the edge capacities change with two different time scales, and instead assume the same time scale for both quantities. Renouncing to such a modeling assumption would mutate the traditional idea of optimization on a fixed fitness landscape to one of a “seascape”—a concept largely diffused, for example, in genetics—where the landscape’s parameter change as the optimization procedure gets carried out. This would add several technical difficulties since the analytical derivations performed in our model could no longer be performed, and one should resort to different mathematical and numerical tools.

## 4 Conclusion and future perspectives

As for the bilevel optimization routing scheme, it would be interesting to tackle particular engineering challenges through adjustments to the established method. For instance, one might consider a particular relationship between edge costs and road tolls, and therefore derive a scheme where optimization is performed by directly adjusting tolls. Alternatively, one could explore different upper-level costs that involve not only edge over-trafficking but also incorporate additional metrics for the global functioning of the network, such as an estimation for the carbon emissions of transportation.

Concerning supervised image classification, our methods naturally do not attain the highest accuracy compared to more complex competitors. However, our study serves a different purpose. Similar ideas to multicommodity adaptation could be incorporated into deep architectures to enhance their state-of-the-art performance. For example, the multicommodity equations could include other image features, such as edges, shapes, and contours, to facilitate classification.

# A Published works




The following publications are accompanied by Supplementary Information, which is made available through the publishers.

Supplementary Information:

1. Lonardi *et al.* [3]  
*Multicommodity routing optimization for engineering networks: Supplementary Information*
2. Lonardi *et al.* [5]  
*Immiscible Color Flows in Optimal Transport Networks for Image Classification: Supplementary Information*
3. Lonardi and De Bacco [6]  
*Bilevel Optimization for Traffic Mitigation in Optimal Transport Networks: Supplemental Material*

Article

# Optimal Transport in Multilayer Networks for Traffic Flow Optimization

Abdullahi Adinoyi Ibrahim , Alessandro Lonardi  and Caterina De Bacco 

Max Planck Institute for Intelligent Systems, Cyber Valley, 72076 Tuebingen, Germany

\* Correspondence: [abdullahi.ibrahim@tuebingen.mpg.de](mailto:abdullahi.ibrahim@tuebingen.mpg.de) (A.A.I.); [alessandro.lonardi@tuebingen.mpg.de](mailto:alessandro.lonardi@tuebingen.mpg.de) (A.L.); [caterina.debacco@tuebingen.mpg.de](mailto:caterina.debacco@tuebingen.mpg.de) (C.D.B.)

**Abstract:** Modeling traffic distribution and extracting optimal flows in multilayer networks is of the utmost importance to design efficient, multi-modal network infrastructures. Recent results based on optimal transport theory provide powerful and computationally efficient methods to address this problem, but they are mainly focused on modeling single-layer networks. Here, we adapt these results to study how optimal flows distribute on multilayer networks. We propose a model where optimal flows on different layers contribute differently to the total cost to be minimized. This is done by means of a parameter that varies with layers, which allows to flexibly tune the sensitivity to the traffic congestion of the various layers. As an application, we consider transportation networks, where each layer is associated to a different transportation system, and show how the traffic distribution varies as we tune this parameter across layers. We show an example of this result on the real, 2-layer network of the city of Bordeaux with a bus and tram, where we find that in certain regimes, the presence of the tram network significantly unburdens the traffic on the road network. Our model paves the way for further analysis of optimal flows and navigability strategies in real, multilayer networks.

**Keywords:** optimal transport; networks; multilayer networks; routing optimization



**Citation:** Ibrahim, A.A.; Lonardi, A.; Bacco, C.D. Optimal Transport in Multilayer Networks for Traffic Flow Optimization. *Algorithms* **2021**, *14*, 189. <https://doi.org/10.3390/a14070189>

Academic Editor: Lijun Chang

Received: 29 May 2021  
Accepted: 21 June 2021  
Published: 23 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Investigating how a network operates and assessing an optimal network design in interconnected networks is a critical problem in several areas [1]. Examples of these include economics [2], climate systems [3], epidemic spreading [4–6] and transportation networks [7]. The main challenge of these problems is to account for the various types of connections that nodes can use to travel through the network efficiently. For example, in transportation networks, the main application considered here, passengers can travel using various means of transport within the same journey. The different transportation modes can operate in significantly different ways [8,9]. For instance, traveling along a rail network (e.g., by tram or subway) is usually faster than along a road network (e.g., by car or bus). The rail network is less sensitive to traffic congestion but the road network has wider coverage and thus allows to reach more destinations. The question is how to combine all these different features to design optimal networks and predict the optimal trajectories of passengers.

Multilayer networks [1,10–12] are a powerful tool to study multi-modal transportation networks [13–15]. Transport in a multilayer network, where layers correspond to transport modes, is often studied using diffusion or spreading processes [1,16–18]. Many of these works use shortest-path minimization [14,19–21] as the main method to extract the passengers' trajectories. However, this can be a restrictive choice: on one side, this assumes that different layers share the same cost function to be minimized; on the other side, shortest-path minimization is not sensitive to traffic congestion and thus, may not be realistic in certain scenarios. Empirical studies [22] have also indicated that passengers may not necessarily choose the shortest paths.



Here, instead, we propose a model that considers more general transport cost minimization, based on a regularized version of the Monge–Kantorovich optimal transport problem [23]. The regularization is obtained via a parameter  $\beta$  that allows to flexibly tune the cost between settings where traffic is penalized or consolidated. Optimal transport is a proven, powerful tool to model traffic in networks and optimal network design [24–39]. Recent works [30,40] extended this formalism to a multi-commodity case that properly accounts for passengers with different origins and destinations. All these studies consider the case of a single-layer network, i.e., one transportation mode. The existence of multiple connections on different layers invites a generalization of these recent results of optimal transport to cope with multilayer networks.

Here, we make this effort and propose a model that uses optimal transport theory to design optimal multilayer networks and finds optimal path trajectories on them. We show how such networks operate under various transport costs tuned by  $\beta$  on both synthetic and real data. We see how the traffic evolves from being more homogeneous to a more unbalanced traffic distribution when a second layer is present and the cost to travel through it changes.

In summary, the goal of this work is to propose an efficient optimal transport-based method for modeling optimal network flows in multilayer networks. Our model finds optimal flows by naturally incorporating the different nature of transportation modes and is computationally efficient. While here, we focus on transportation networks, our method is applicable to a broader set of practical applications involving flows on multilayer networks.

#### *What Makes Multilayer Networks Different Than Single-Layer in Transportation*

Having given the broader context for our work, we now highlight the main features of transport on multilayer networks. The presence of edges between layers (inter-layer edges) makes a multilayer network fundamentally distinct from a standard single-layer one, as these edges allow passengers to switch between transportation modes. However, this is not the only difference. In fact, in a multilayer network, the various layers have different characteristics. The main one is that the type of transportation cost varies across layers. For example, the cost to build and maintain the infrastructure differs depending on the transportation mode, with subway or rail tracks costing more than a road network. Moreover, the cost assigned to traffic congestion is also different, as road networks are more sensitive to traffic bottlenecks than rail ones. In addition, the power dissipated differs depending on the means of transportation, as running a tram generally produces fewer CO<sub>2</sub> emissions than running a bus. All these different features impact the results of an optimal transport problem, as the network features contributing to the cost function to be optimized vary with layers, and thus also the optimal solution.

Finally, the network topologies themselves vary with layers [41], as a bus network has many edges with short lengths, while a rail network tends to have fewer but longer edges. In addition, the weights assigned to each edge differ based on the layer, which can induce coupling between layers [42].

## 2. Materials and Methods

### 2.1. Multilayer Transportation Networks

In general, a multilayer network is represented as a graph  $G(\{\mathcal{V}_\alpha\}_\alpha, \{\mathcal{E}_\alpha\}_\alpha, \{\mathcal{E}_{\alpha\gamma}\}_{\alpha,\gamma})$ , where  $\mathcal{V}_\alpha$  and  $\mathcal{E}_\alpha$  are the set of nodes and edges in layer  $\alpha$ , respectively, and  $\mathcal{E}_{\alpha\gamma}$  is the set of edges between nodes in layer  $\alpha$  and nodes in layer  $\gamma$ . Here,  $\alpha = 1, \dots, L$ , where  $L$  is the number of layers. We denote with  $N_\alpha = |\mathcal{V}_\alpha|$  the number of nodes in layer  $\alpha$ , and with  $E_\alpha = |\mathcal{E}_\alpha|$  the number of edges in layer  $\alpha$ ,  $E_{\alpha\gamma} = |\mathcal{E}_{\alpha\gamma}|$  is the number of edges between nodes in layer  $\alpha$  and  $\gamma$ . Finally, we denote with  $\mathcal{V}_0 = \cup_\alpha \mathcal{V}_\alpha$  the total set of nodes, with  $\mathcal{E}_0 = (\cup_\alpha \mathcal{E}_\alpha) \cup (\cup_{\alpha\gamma} \mathcal{E}_{\alpha\gamma})$  the total set of edges, and with  $N_0 = |\mathcal{V}_0|$  and  $E_0 = |\mathcal{E}_0|$ , their cardinalities. We assume that edges have lengths  $l_e > 0$ , which determine the cost to travel through them.

Transportation networks are relevant examples of this type of structure, where nodes are stations, edges are connections between stations and layers are transportation modes, for instance, rails or bus routes. A convenient way to represent a multilayer network is with two tensors [43]: (i) an intra-layer adjacency tensor  $A$  with entries  $A_{uv}^\alpha = 1$  if there is an edge between nodes  $u$  and  $v$  in layer  $\alpha$ , and 0 otherwise. We refer to this type of edge as an *intra-layer* edge; (ii) an inter-layer adjacency tensor  $\hat{A}$  with entries  $\hat{A}_{uv}^{\alpha\gamma} = 1$  if there is an edge between node  $u$  in layer  $\alpha$  and node  $v$  in layer  $\gamma$ , and 0 otherwise. Without loss of generality, in our applications, we have  $\hat{A}_{uv}^{\alpha\gamma} = 0$  if  $u \neq v$ , meaning that different layers are connected solely by shared nodes. We refer to edges connecting nodes in different layers as *inter-layer* edges. In the case of transportation networks, the main application studied here, a station could have a bus stop, a train platform and a subway entrance, which allows passengers to switch between communication modes within the same station. For example, one can think of an inter-layer edge as the stairs connecting the subway entrance with the entrance to the train station. Typically, inter-layer edges are, thus, much shorter than intra-layer edges.

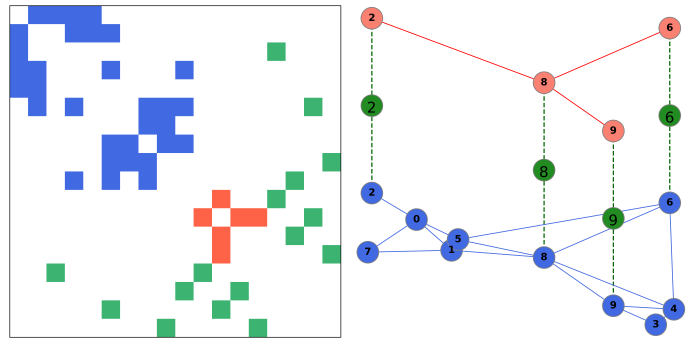
In the case of multilayer networks, we need to be careful with how stations connecting multiple transportation modes are represented. In fact, if an entry station connects more than one layer, we may not be able to distinguish in what layer a passenger enters. In other words, if a node  $u$  belongs to more than one layer, i.e., a node  $u_\alpha$  exists for more than one value of  $\alpha$ , we may not be able to tell whether the passengers entering  $u$  entered from  $u_\alpha, u_\gamma$  or from any of the other instances of node  $u$  in the various layers. To alleviate this problem, we build auxiliary *super* nodes  $u$ , which do not belong to any layer in particular but instead connect the various instances of the same node in the various layers together. Specifically, we remove all the inter-layer edges  $(u_\alpha, u_\gamma)$  and replace them with auxiliary *inter-super* edges  $(u_\alpha, u)$ , connecting all the instances  $u_\alpha$  of node  $u$  with the super node  $u$ , as in a star graph, so that the original edge  $(u_\alpha, u_\gamma)$  is replaced by a two-edge path  $\{(u_\alpha, u), (u_\gamma, u)\}$ .

This auxiliary structure allows the model to allocate in an optimal way the passengers along the inter-super edges when they enter from a station with connections to more than one layer, thereby avoiding the selection of arbitrary entrances a priori. This becomes relevant in applications where the cost to travel along inter-layer edges is non trivial, for instance, in situations where changing connection impacts the comfort of the passengers.

Moreover, the introduction of super nodes and edges facilitates how we represent the multilayer network. In fact, by adding these auxiliary super nodes and inter-super edges, we only need to consider an individual network adjacency matrix  $A$ , instead of two separate tensors. This matrix has entries  $A_{uv} = 1$  if an edge exists between nodes  $u$  and  $v$  and 0 otherwise, where a node  $u$  can be a node  $u_\alpha$  in layer  $\alpha$  or a super node  $u$ . The set of nodes is then  $\mathcal{V} = \mathcal{V}_0 \cup \mathcal{V}_{super}$ , where  $\mathcal{V}_{super}$  is the set of super nodes, and  $|\mathcal{V}_{super}| = N_{super}$  is their number, which corresponds to the number of nodes that belong to more than one layer. Similarly, the new set of edges is  $\mathcal{E} = (\cup_\alpha \mathcal{E}_\alpha) \cup \mathcal{E}_{super}$ , where  $\mathcal{E}_{super}$  is the set of inter-super edges. The final numbers of nodes and edges are  $N = |\mathcal{V}| = N_0 + N_{super}$  and  $E = |\mathcal{E}| \geq E_0$ . Notice that this construction is equivalent to assume that the network has  $L + 1$  layers, where the extra layer is made of inter-super edges  $\mathcal{E}_{super}$  and all nodes incident to them (without loss of generality, we assume that all the inter-super edges are treated equally). We denote it as the *super* layer and this corresponds to  $\alpha = L + 1$ , so that  $\mathcal{E}_{L+1} \equiv \mathcal{E}_{super}$ . We show an example of this structure in Figure 1.

Finally, we consider a coupling between layers as in [42] that controls how the layers are linked. Specifically, we multiply the lengths of each edge by a factor  $w_\alpha \in [0, 1]$  that depends on what layer the edge belongs to. For convenience, we introduce  $q_e \equiv q_e(\alpha)$  taking values  $q_e = \alpha$  for each  $e \in \mathcal{E}_\alpha$  and with  $\alpha = 1, \dots, L + 1$ . Using this, we define the resulting length as  $\ell_e := w_{q_e} l_e$ . This ensures that edges in different layers can be navigated differently. If we interpret  $w_\alpha$  as the inverse of a velocity, then  $\ell_e$  is proportional to the time needed to travel along edge  $e$ , which can be seen as an “effective” length. When  $w_\alpha < 1$  and  $w_\gamma = 1$ , a passenger takes less time to travel along an edge of length  $l_e$  in  $\alpha$  than one in

$\gamma$ . Typically,  $\ell_e$  are small for inter-super edges. Nevertheless, one can tune the cost to travel along them by tuning  $w_{L+1}$ .



**Figure 1.** Example of multilayer structure. We show an example of a 2-layer network with  $N = 18$  ( $N_1 = 10, N_2 = 4$  and  $N_{super} = 4$ ). (Left) adjacency matrix  $A$ , colors denote the layer type: blue is layer 1, red is layer 2 and green is the super layer. (Right) the 2-layer network with layer 1 on the bottom, layer 2 on top, and the super nodes in between.

### 2.2. The Model

We consider the formalism of optimal transport theory, and in particular, recent works that map the setting of solving a standard optimization problem into that of solving a dynamical system of equations [24–30,40]. Specifically, we model two main quantities defined on network edges: (i) fluxes  $F_e$  of passengers traveling through an edge  $e$ ; and (ii) conductivities  $\mu_e$ , which are quantities determining the flux passing through an edge  $e$ . Intuitively, the conductivity  $\mu_e$  of an edge can be seen as proportional to the size of the edge  $e$ . To keep track of the different routes that passengers have, we consider multi-commodity formalism as in [40], i.e., we distinguish passengers based on their entry station  $a \in S$ , where  $S \subseteq \mathcal{V}$  is the set of stations where passengers enter, and we denote with  $M = |S|$  the number of passenger types. With this formalism, we have that the fluxes  $F_e$  are  $M$ -dimensional vectors, where the entries  $F_e^a$  denote a number of passengers of type  $a$  traveling on edge  $e$ . The important modeling choice is that the conductivities  $\mu_e$  are shared between passengers, thus they are scalar numbers contributing to the cost for all passenger types traveling through  $e$ . This formalism can be equally applied to both edge types: intra-layer and inter-super edges.

We assume that fluxes are determined by pressure potentials  $p_u^a$  defined on nodes as follows:

$$F_e^a := \frac{\mu_e}{\ell_e} (p_u^a - p_v^a), \quad e = (u, v) \quad (1)$$

We model the number of passengers entering a station  $a$  with a positive real number  $g^a$ . For notational convenience, we define a  $N \times M$  dimensional matrix of entries  $g_u^a$  such that  $g_u^a := 0$  if  $u \neq a$ , and  $g_u^a := g^a$  if  $u = a$ . Similarly, we define with  $h_u^a$  the number of passengers of type  $a$  exiting at node  $u$ . Here, the only constraint is that  $h_u^a = 0$  if  $u = a$  to avoid unrealistic situations where passengers entering in one station exit from the same station. Finally, we define the  $N \times M$ -dimensional source matrix with entries  $S_u^a = g_u^a - h_u^a$ , which indicates the number of passengers of type  $a$  entering or exiting a station. Notice that for each  $a \in S$  we have  $\sum_u S_u^a = 0$ , meaning the system is isolated, i.e., all the passengers of a certain type who enter the network also exit.

With this in mind, we enforce mass conservation by imposing Kirchhoff’s law on nodes. To properly enforce this constraint, we need to consider all the edges, both intra-layer and inter-layer edges. This can be compactly written by considering the multilayer

network signed incidence matrix  $B$  with entries  $B_{ve} = 1, -1, 0$  if node  $v \in \mathcal{V}$  is the start, end of edge  $e \in \mathcal{E}$ , or none of them, respectively. With this in mind, Kirchhoff’s law can be written as follows:

$$\sum_e B_{ve} F_e^a = S_v^a, \quad \forall a \in \mathcal{S}, \forall v \in \mathcal{V} \quad (2)$$

Finally, we assume that the conductivities follow the following dynamics:

$$\dot{\mu}_e = \mu_e^{\beta_{q_e}} \frac{\sum_{a \in \mathcal{S}} (p_u^a - p_v^a)^2}{\ell_e^2} - \mu_e, \quad \forall e \in \mathcal{E} \quad (3)$$

where  $q_e$  encodes the type of edge, as defined in Section 2.1. The parameter  $0 < \beta_{q_e} < 2$  is important, as it determines the type of optimal transport problem that we aim to solve, which we describe in more detail later. Interpreting the conductivities as quantities proportional to the size of an edge, this dynamics enforces a feedback mechanism such that the edge size increases if the flux through that edge increases, it decreases otherwise. This feedback mechanism was observed in biological networks, such as the one made by slime mold *Physarum polycephalum* [24,44], which adapts its body shape to optimally navigate the space, searching for food.

The important property of this dynamics is that its stationary solutions minimize a multilayer transport cost function:

$$J_\beta = \sum_{\alpha=1}^{L+1} \sum_{e \in \mathcal{E}_\alpha} \ell_e \|F_e\|_2^{\Gamma(\beta_\alpha)} \quad (4)$$

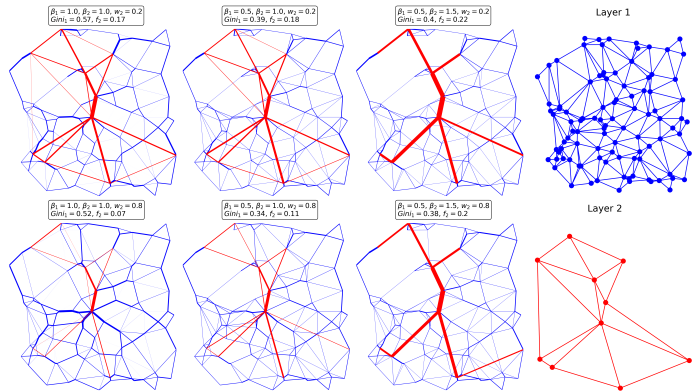
where  $\Gamma(\beta_\alpha) = 2(2 - \beta_\alpha) / (3 - \beta_\alpha)$  for all  $\alpha$  and the 2-norm is calculated over the  $M$  entries of each  $F_e$ . This means that solving the systems of Equations (1)–(3) is equivalent to finding the optimal trajectories of passengers in a multilayer network, where optimality is given with respect to the cost in Equation (4). An extended discussion and a formal derivation of this property can be found in [32,40].

The parameter  $\beta_{q_e}$  (taking value  $\beta_\alpha$  on layer  $\alpha$ ) regulates how the fluxes should distribute in each of the layers. In fact, according to Equation (4), when  $\beta_\alpha > 1$ , the fluxes are encouraged to consolidate into few edges of a layer  $\alpha$ , being  $\Gamma(\beta_\alpha) < 1$ , and thus the cost in Equation (4) is sub-linear. In the opposite scenario, when  $0 < \beta_\alpha < 1$ , we have that the fluxes are encouraged to distribute over more edges and with lower values in order to keep traffic congestion low. Finally, when  $\beta_\alpha = 1$ , we obtain the shortest path-like minimization. The consequence of having different  $\beta_\alpha$  in different layers is that the optimal trajectories have different topologies in each of the layers. At the same time, layers are coupled together, thus the final trajectories are a complex combination of the weights  $w_\alpha$  and the  $\beta_\alpha$ . We give an example of optimal flows for various combinations of these parameters in Figure 2.

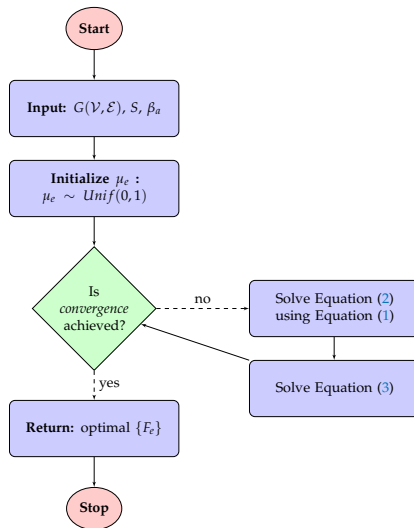
### 2.3. The Algorithmic Implementation

The numerical implementation consists of initializing the  $\mu_e > 0$  at random. Then, one iterates between (i) extracting the pressure potentials (or the fluxes) using Equations (1) and (2), and (ii) using these to recompute the  $\mu_e$  by means of Equation (3), which can be solved numerically with finite difference discretization. The iteration is repeated until convergence. In our experiments, we terminate a run of the algorithm when the difference  $J_\beta^{(t+1)} - J_\beta^{(t)}$  between two successive updates is lower than a threshold (the superscript  $(t)$  is the iteration step). The cost  $J_\beta$  in Equation (4) is not strictly convex in general, hence the solution of Algorithm 1 may converge to a local optima. One should then run the algorithm several times, each time initializing to a different random initial realization of  $\mu_e > 0$ . A possible choice for a final optimal solution is the one that has lower  $J_\beta$ . We give the pseudocode for this in Algorithm 1; this is complemented with the block diagram in Figure 3. Most of the computational effort required by Algorithm 1 is in the solution of

$M$  linear systems as in Equation (2). In our implementation, this is performed by a sparse direct solver (UMFPACK), performing a LU decomposition of each column of the right hand side of Equation (2), and having complexity scaling as  $\mathcal{O}(MN^2)$ .



**Figure 2.** Example of optimal paths. We show an example of optimal paths obtained with:  $p = 0.2$  and (top)  $w_1 = 0.2$ , (bottom)  $w_1 = 0.8$ . Values of  $\beta_1, \beta_2$  are those reported on top of each network. The statistics  $Gim_1$  and  $f_2$  are those defined in Section 3.1. The width of edges is proportional to the optimal  $\|F_e\|_2$ . Blue and red edges are for layers 1 and 2, respectively. The two layers are plotted individually on the rightmost column.



**Figure 3.** Block diagram of Algorithm 1. We give a pictorial representation of the pseudocode in Algorithm 1. Here, rectangular blocks are *action* blocks, corresponding to the update of a variable, to an input initialization, or to the output of the fluxes at convergence. *Conditional* blocks are diamond-shaped; elliptical blocks denote the *start* and *stop* points.

---

**Algorithm 1** Multilayer optimal transport.

---

- 1: **Input:** multilayer network  $G(\mathcal{V}, \mathcal{E})$ , source matrix  $S, \beta_\alpha$
  - 2: **Initialize:**  $\{\mu_e\}$  (e.g., sampling as i.i.d.  $\mu_e \sim Unif(0, 1)$ )
  - 3: **while** convergence not achieved **do**
  - 4:     use Equation (1) to solve Kirchhoff’s law as in Equation (2)  $\rightarrow \{p_u^q\}$
  - 5:     solve the dynamics in Equation (3):  $\{\mu_e^t\} \rightarrow \{\mu_e^{t+1}\}$
  - 6: **end while**
  - 7: **Return:** fluxes  $\{F_e\}$  at convergence, computed using Equation (1)
- 

The resulting  $\{F_e\}$  capture how passengers travel along the network via optimal trajectories. The norms  $\|F_e\|_2$  measure the total number of passengers along an edge  $e$ .

**3. Results**

*3.1. Results on Synthetic Data*

We show how the model works on synthetic data where each layer is planar, to mimic realistic scenarios of transportation networks in space. We generate 2-layer networks and the source matrix  $S$  as done in [42]. Specifically, we generate one layer by randomly placing  $N$  nodes in the square  $[0, 1] \times [0, 1]$  and then extract their Delaunay triangulation [45]. We then select a subset of nodes and use this to build the second layer with an analogous procedure. An example of this is given in Figure 2. After having constructed the network topology, we assign entry and exit stations to each node in the network, starting from a monocentric scenario where all passengers exit from a central station, regardless of their origin. We then randomly re-assign with a probability  $p \in [0, 1]$  the exit station of each set of passengers. When  $p = 0$ , all the passengers travel to the city center, while when  $p = 1$ , the destinations are assigned completely at random.

We generate 20 networks with  $N_1 = 100$  and  $N_2 = 10$ , so that layer 1 has, on average, shorter edges than layer 2. For each sampled network, we take 50 random samples of  $S$ . We consider  $p \in \{0.2, 0.8\}$  to study two opposite situations of having a majority or a minority of the passengers directed to a common central node. Then, we fix  $w_1 = 1$  and vary  $w_2 \in \{0.2, 0.8\}$  to mimic a scenario where traveling on the second layer is faster.

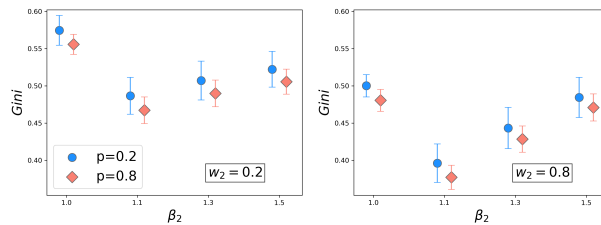
Overall, with these combinations of parameters, we obtain 2-layer networks that resemble a road–rail network. With this in mind, we run our model with the following combination of parameters for the dynamics:  $(\beta_1, \beta_2) \in \{(0.5, 1.1), (0.5, 1.3), (0.5, 1.5), (1, 1)\}$ . This is because we expect to penalize traffic congestion in a road network, hence  $\beta_1 = 0.5$ . Instead, a rail network is less sensitive to traffic but it may cost more to build connections, thus once should consolidate traffic along fewer edges, hence  $\beta_2 > 1$ . The case  $(\beta_1, \beta_2) = (1, 1)$  is used as a baseline for comparison with the shortest path-like optimization.

We measure how passengers distribute along the optimal trajectories to assess how the network operates under various regimes of  $w$  and  $\beta$ . For this, we consider  $\|F_e\|_2$  and measure the distribution of this quantity along the edges to see how this varies across parameters’ values and in each of the two layers. In addition, we calculate the current flow edge betweenness centrality (FBC) [46], which captures how important an edge is based on how many passengers travel through it. This is different than the standard edge betweenness centrality [47] in that it considers random paths connecting two points, instead of only the shortest paths. We argue that FBC is more appropriate in our case, as the shortest paths may not be the optimal trajectories where passengers travel. We calculate the weighted version of FBC, where the edge weight is  $\|F_e\|_2$ , so that the random paths are more likely to follow edges with higher flux. We use the Gini coefficient  $Gini \in [0, 1]$  to characterize the disparity in the flow assignment along edges. We consider the following definition [48]:

$$Gini := \frac{1}{2E^2\bar{x}} \sum_{r,q} |x_r - x_q| \quad , \quad (5)$$

where  $r, q$  denote edges,  $x$  is the quantity that we want to measure this coefficient with, and  $\bar{x} = \sum_e x_e / E$  is its average value. Here, we use  $x_e = \|F_e\|_2$  and  $x_e = FBC_e$ . When *Gini* is close to one, most of the flow passes through few edges, whereas when *Gini* is small, the flows are distributed evenly across the edges.

Looking at Figure 4, we see that *Gini* increases with  $\beta_2$  and thus the network usage becomes more hierarchical, as expected in this case (we report here results for *Gini* w.r.t. the flux, but similar results are observed for FBC, see Figure A1). The exact value of *Gini* depends on the travel demand, as for  $p = 0.2$ , i.e., when the central node is a destination in 80% of the journeys, *Gini* is higher than when  $p = 0.8$ . This is because with fewer destinations, there are also fewer possible path trajectories, and thus more passengers use the same part of the network. We can also see how *Gini* decreases for higher  $w_2$ , i.e., when traveling by tram is not much faster than traveling on the road network. Finally, we can notice the drop in *Gini* compared to the shortest path-like scenario  $\beta_1 = \beta_2 = 1$ . In this case, the traffic distribution is the most hierarchical, suggesting that possible traffic congestion can be avoided by setting lower values of  $\beta_1$ .



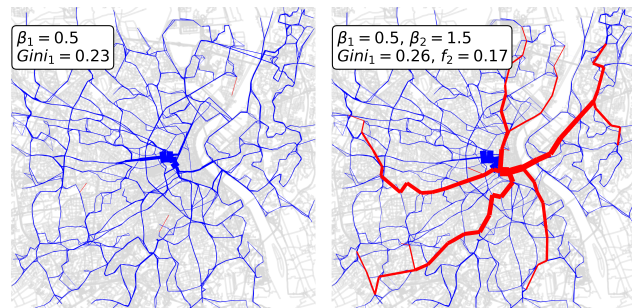
**Figure 4.** Results on synthetic data. We show the *Gini* w.r.t. the optimal  $\|F_e\|_2$  (y axis) vs.  $\beta_2$  (x axis) for synthetic 2-layer networks generated as in Section 3.1. Blue and red markers denote  $p = 0.2, 0.8$ , respectively,  $w_1 = 1$  in all cases, while  $w_2 = 0.2$  (left) and  $w_2 = 0.8$  (right);  $\beta_1 = 0.5$  in all cases, except for the case where  $\beta_2 = 1$  for which  $\beta_1 = 1$ . This case is the shortest path-like baseline. Markers are averages over 20 network samples and 50 source matrix samples (for a total of 1000 individual samples).

Our model can be used to simulate traffic distributions under various conditions. In fact, tuning  $p$ ,  $\{w_\alpha\}$  and  $\{\beta_\alpha\}$ , one can simulate disparate scenarios. For instance, in Figure 2 we show results for different parameters' choices on a particular realization of a 2-layer synthetic network. Several conclusions can be drawn from this simple experiment. For instance, the second layer, which ideally can represent a tram network, is only partially used when  $\beta_2 = 1.5$ . This value encourages traffic to consolidate on fewer main connections, simulating the scenario where building the rail infrastructure is expensive. Our model can guide a network manager to decide what edges should be prioritized when designing the network. In this example, we can distinguish which set of edges are the most utilized. These are mainly central edges, but the exact set can change depending on the other parameters. For example, if the travel demand, tuned by  $p$ , switches from a monocentric to a more heterogeneous set of entry-exit stations, one of the main central edges changes from connecting a periphery to the center, to connecting two locations in the periphery.

### 3.2. Results on Real Data

We illustrate our model on a real 2-layer network of the city of Bordeaux, where the two layers are the bus and tram, respectively. Data are taken from [49]. We simulate a monocentric source matrix  $S$ , i.e.,  $p = 0.0$ , to assess the scenario where all the passengers travel to the city center; however, the results are similar for other values of  $p$  (not reported here). Optimal paths are extracted using our model for  $\beta_1 = 0.5$ ,  $\beta_2 = 1.5$ ,  $w_2 = 0.2$  and compared against the case where the tram network is absent. This can be simulated by setting a high value of  $w_2$ , so that the cost on the tram edges makes it extremely unlikely to

use any tram connection (here, we use  $w_2 = 100$ ). We measure the total percentage flux  $f_2 = \sum_{e \in E_2} \|F_e\|_2 / (\sum_{e \in E_1} \|F_e\|_1 + \sum_{e \in E_2} \|F_e\|_2)$  passing through layer 2. Remarkably, in this scenario, the tram network absorbs  $f_2 = 17\%$  of the total flow of passengers, even though the tram network contains only  $E_2 = 112$  edges, compared to  $E_1 = 2347$  bus edges. This allows to reduce significantly the traffic along the road network, as can be seen in Figure 5, and the road edges, and, in particular, those parallel to the tram line and close to the city center get thinner as more passengers use the tram. This also results in a higher  $Gini_1 = 0.26$  (calculated on edges in layer 1 w.r.t.  $\|F_e\|_2$ ), compared to the  $Gini_1 = 0.23$  when the tram is absent: as the passengers use the tram, they decrease traffic on many road edges. While the traffic distribution on layer 1 gets more hierarchical (higher  $Gini_1$ ), this does not necessarily lead to more traffic congestion. In fact, the total percentage flow  $f_1$  decreases, as we saw above. Additional plots can be seen in Figure A2.



**Figure 5.** Example of optimal paths in the city of Bordeaux for a bus and tram network. The paths are obtained with (left) and without (right) the tram layer. Here,  $\beta_1 = 0.5$  in both cases, while  $\beta_2 = 1.5$  in the second case. The width of the edges is proportional to the optimal  $\|F_e\|_2$ . The reported  $Gini_1$  coefficient for the bus network (layer 1) is calculated using  $\|F_e\|_2$ . The total percentage flux  $f_2 = \sum_{e \in E_2} \|F_e\|_2 / (\sum_{e \in E_1} \|F_e\|_1 + \sum_{e \in E_2} \|F_e\|_2) = 0.17$ , distributed over  $E_2 = 112$  tram edges, compared to  $E_1 = 2347$  bus edges.

#### 4. Discussion

We have presented a model that extracts optimal flows on multilayer networks based on optimal transport theory. Our models accounts for different contributions from different layers to the total transport cost by means of a parameter  $\beta_\alpha$ . Our modeling choice is relevant in scenarios where passengers can travel using different transport modalities on an interconnected transportation network. We have shown how the optimal distribution of passenger flows on network edges is influenced by different factors. In fact, a complex combination of the parameter  $\beta_\alpha$  on each layer, the coupling between layers and the distribution of the origin and destination pairs determine how heterogeneous the flow distributions are inside the various layers. In particular, when  $\beta_\alpha < 1$  in one layer and  $\beta_\alpha > 1$  in another layer, the network topologies are significantly different in the two layers, as in one, the traffic is more balanced and distributed along many edges, while in the other, the traffic is consolidated along a few main arteries. To show the potential of our model, we considered an application to the 2-layer bus and tram network of Bordeaux, showing how the presence of the tram changes the traffic distribution on the road network.

#### 5. Conclusions

In this work, we proposed a model that uses optimal transport theory to find optimal path trajectories on multilayer networks. By means of the regularization parameter  $\beta_\alpha$ , we were able to take into account different contributions from the different layers for the total transportation cost. We illustrated the model on both synthetic and real data and



showed how the optimal distribution of passenger flows on network edges is influenced by different parameters used for the construction of the model (i.e.,  $w$ ,  $p$ ,  $\beta_\alpha$ ).

In the absence of real data, we simulated the entry and exit destination of passengers. However, if travel demands are known, for instance, using mobile data [50], it would be interesting to investigate the distribution of traffic obtained with our model and compare it with real usage data as done in [51]. We considered a cost assigned on edges where  $\beta_\alpha$  tunes the impact of traffic on them, but one can generalize this to include penalties on nodes based on their degrees, as considered in [52]. Our model can be used to extract the main features of multilayer transportation networks [53] or to study the existence of several congestion regimes in both synthetic and real data [21] and investigate how this changes, varying  $\beta_\alpha$ . Finally, in our experiments, we fixed the weight of inter-super nodes to be small. Potentially, one could suitably increase this to account for the cost of changing transportation modes within a journey and use our model to see how optimal trajectories change. This would be relevant in scenarios where the passengers' comfort contributes to the total transport cost. To facilitate future analysis, we provide an open source implementation of our code at <https://github.com/cdebacco/MultiOT> (accessed on 28 May 2021).

**Author Contributions:** All authors contributed to developing the models, conceived the experiments, analyzing the results and reviewing the manuscript. A.A.I. conducted the experiments. All authors have read and agreed to the published version of the manuscript.

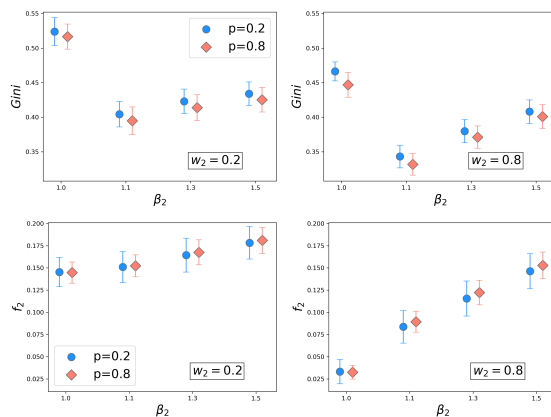
**Funding:** This research received no external funding.

**Data Availability Statement:** Publicly available data sets were analyzed in this study. These data can be found here: <http://transportnetworks.cs.aalto.fi> (accessed on 28 May 2021).

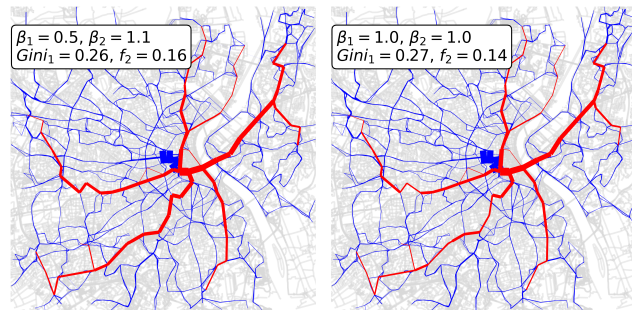
**Acknowledgments:** The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Abdullahi Adinoyi Ibrahim and Alessandro Lonardi.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A



**Figure A1.** Additional results on synthetic data. We show the *Gini* w.r.t. the optimal *FBC* (top) and the total percentage flux  $f_2$  on layer 2 (bottom) vs.  $\beta_2$  (x axis), for synthetic 2-layer networks generated as in Section 3.1;  $w_2 = 0.2, 0.8$  (left, right),  $\beta_1 = 0.5$  in all cases, except for the case where  $\beta_2 = 1$ , for which  $\beta_1 = 1$ . This case is a shortest path-like baseline. Markers are averages over 20 network samples and 50 source matrix samples.



**Figure A2.** Additional example of optimal paths in the city of Bordeaux for a bus and tram network. Here  $p = 0.0$ ,  $w_2 = 0.2$ ,  $\beta_1, \beta_2 = (0.5, 1.1), (1.0, 1.0)$  (left, right). The width of the edges is proportional to the optimal  $\|F_e\|_2$ .  $Gini_1$  is calculated w.r.t. to the flux on layer 1;  $f_2 = \frac{\sum_{e \in \mathcal{E}_2} \|F_e\|_2}{(\sum_{e \in \mathcal{E}_1} \|F_e\|_1 + \sum_{e \in \mathcal{E}_2} \|F_e\|_2)}$ .

## References

- Boccaletti, S.; Bianconi, G.; Criado, R.; Del Genio, C.I.; Gómez-Gardenes, J.; Romance, M.; Sendina-Nadal, L.; Wang, Z.; Zanin, M. The structure and dynamics of multilayer networks. *Phys. Rep.* **2014**, *544*, 1–122. [\[CrossRef\]](#)
- Gómez-Gardenes, J.; Reinares, I.; Arenas, A.; Floría, L.M. Evolution of cooperation in multiplex networks. *Sci. Rep.* **2012**, *2*, 620. [\[CrossRef\]](#) [\[PubMed\]](#)
- Donges, J.F.; Schultz, H.C.; Marwan, N.; Zou, Y.; Kurths, J. Investigating the topology of interacting networks. *Eur. Phys. J. B* **2011**, *84*, 635–651. [\[CrossRef\]](#)
- Saumell-Mendiola, A.; Serrano, M.Á.; Boguná, M. Epidemic spreading on interconnected networks. *Phys. Rev. E* **2012**, *86*, 026106. [\[CrossRef\]](#)
- Dickison, M.; Havlin, S.; Stanley, H.E. Epidemics on interconnected networks. *Phys. Rev. E* **2012**, *85*, 066109. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chen, J.; Hu, M.B.; Li, M. Traffic-driven epidemic spreading in multiplex networks. *Phys. Rev. E* **2020**, *101*, 012301. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kurant, M.; Thiran, P. Layered complex networks. *Phys. Rev. Lett.* **2006**, *96*, 138701. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wu, J.; Pu, C.; Li, L.; Cao, G. Traffic dynamics on multilayer networks. *Digit. Commun. Netw.* **2020**, *6*, 58–63. [\[CrossRef\]](#)
- Zhuo, Y.; Peng, Y.; Liu, C.; Liu, Y.; Long, K. Traffic dynamics on layered complex networks. *Phys. A Stat. Mech. Appl.* **2011**, *390*, 2401–2407. [\[CrossRef\]](#)
- Aleta, A.; Moreno, Y. Multilayer networks in a nutshell. *Annu. Rev. Condens. Matter Phys.* **2019**, *10*, 45–62. [\[CrossRef\]](#)
- Bianconi, G. *Multilayer Networks: Structure and Function*; Oxford University Press: Oxford, UK, 2018.
- Kivelä, M.; Arenas, A.; Barthélemy, M.; Gleeson, J.P.; Moreno, Y.; Porter, M.A. Multilayer networks. *J. Complex Netw.* **2014**, *2*, 203–271. [\[CrossRef\]](#)
- Strano, E.; Shai, S.; Dobson, S.; Barthélemy, M. Multiplex networks in metropolitan areas: generic features and local effects. *J. R. Soc. Interface* **2015**, *12*, 20150651. [\[CrossRef\]](#)
- Solé-Ribalta, A.; Gómez, S.; Arenas, A. Congestion induced by the structure of multiplex networks. *Phys. Rev. Lett.* **2016**, *116*, 108701. [\[CrossRef\]](#)
- Aleta, A.; Meloni, S.; Moreno, Y. A multilayer perspective for the analysis of urban transportation systems. *Sci. Rep.* **2017**, *7*, 44359. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gomez, S.; Diaz-Guilera, A.; Gomez-Gardenes, J.; Perez-Vicente, C.J.; Moreno, Y.; Arenas, A. Diffusion dynamics on multiplex networks. *Phys. Rev. Lett.* **2013**, *110*, 028701. [\[CrossRef\]](#) [\[PubMed\]](#)
- De Domenico, M.; Granell, C.; Porter, M.A.; Arenas, A. The physics of spreading processes in multilayer networks. *Nat. Phys.* **2016**, *12*, 901–906. [\[CrossRef\]](#)
- De Domenico, M.; Solé-Ribalta, A.; Gómez, S.; Arenas, A. Navigability of interconnected networks under random failures. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 8351–8356. [\[CrossRef\]](#)
- Barthélemy, M. Spatial networks. *Phys. Rep.* **2011**, *499*, 1–101. [\[CrossRef\]](#)
- Solé-Ribalta, A.; Arenas, A.; Gómez, S. Effect of shortest path multiplicity on congestion of multiplex networks. *New J. Phys.* **2019**, *21*, 035003. [\[CrossRef\]](#)
- Lampo, A.; Borge-Holthoefer, J.; Gómez, S.; Solé-Ribalta, A. Multiple abrupt phase transitions in urban transport congestion. *Phys. Rev. Res.* **2021**, *3*, 013267. [\[CrossRef\]](#)

22. Quercia, D.; Schifanella, R.; Aiello, L.M. The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In Proceedings of the 25th ACM Conference on Hypertext and Social Media, Santiago, Chile, 1–4 September 2014; pp. 116–125.
23. Kantorovich, L. On the Transfer of Masses. *J. Math. Math. Sci.* **1942**, *133*, 2006.
24. Bonifaci, V.; Mehlhorn, K.; Varma, G. Physarum can compute shortest paths. *J. Theor. Biol.* **2012**, *309*, 121–133. [[CrossRef](#)] [[PubMed](#)]
25. Santambrogio, F. Optimal channel networks, landscape function and branched transport. *Interfaces Free. Boundaries* **2007**, *9*, 149–169. [[CrossRef](#)]
26. Facca, E.; Cardin, F.; Putti, M. Towards a Stationary Monge-Kantorovich Dynamics: The Physarum Polycephalum Experience. *SIAM J. Appl. Math.* **2016**, *78*, 651–676. [[CrossRef](#)]
27. Facca, E.; Daneri, S.; Cardin, F.; Putti, M. Numerical Solution of Monge-Kantorovich Equations via a Dynamic Formulation. *J. Sci. Comput.* **2020**, *82*, 68. [[CrossRef](#)]
28. Facca, E.; Cardin, F.; Putti, M. Branching Structures Emerging from a Continuous Optimal Transport Model. Available online: <http://xxx.lanl.gov/abs/1811.12691> (accessed on 28 May 2021).
29. Baptista, D.; Leite, D.; Facca, E.; Putti, M.; De Bacco, C. Network extraction by routing optimization. *Sci. Rep.* **2020**, *10*, 088702. [[CrossRef](#)]
30. Bonifaci, V.; Facca, E.; Folz, F.; Karrenbauer, A.; Kolev, P.; Mehlhorn, K.; Morigi, G.; Shahkarami, G.; Vermande, Q. Physarum Multi-Commodity Flow Dynamics. Available online: <http://xxx.lanl.gov/abs/2009.01498> (accessed on 28 May 2021).
31. Kirkegaard, J.B.; Sneppen, K. Optimal Transport Flows for Distributed Production Networks. *Phys. Rev. Lett.* **2020**, *124*, 208101. [[CrossRef](#)]
32. Bohn, S.; Magnasco, M.O. Structure, Scaling, and Phase Transition in the Optimal Transport Network. *Phys. Rev. Lett.* **2007**, *98*, 088702. [[CrossRef](#)]
33. Banavar, J.R.; Maritan, A.; Rinaldo, A. Size and form in efficient transportation networks. *Nature* **1999**, *399*, 130–132. [[CrossRef](#)]
34. Hu, D.; Cai, D. Adaptation and optimization of biological transport networks. *Phys. Rev. Lett.* **2013**, *111*, 138701. [[CrossRef](#)]
35. Ronellenfitch, H.; Katifori, E. Global optimization, local adaptation, and the role of growth in distribution networks. *Phys. Rev. Lett.* **2016**, *117*, 138301. [[CrossRef](#)]
36. Katifori, E.; Szöllösi, G.J.; Magnasco, M.O. Damage and Fluctuations Induce Loops in Optimal Transport Networks. *Phys. Rev. Lett.* **2010**, *104*, 048704. [[CrossRef](#)]
37. Ronellenfitch, H.; Katifori, E. Phenotypes of Vascular Flow Networks. *Phys. Rev. Lett.* **2019**, *123*, 248101. [[CrossRef](#)] [[PubMed](#)]
38. Baptista, D.; De Bacco, C. Principled network extraction from images. *arXiv* **2020**, arXiv:2012.12758. Available online: <http://xxx.lanl.gov/abs/2012.12758> (accessed on 28 May 2021).
39. Kaiser, F.; Ronellenfitch, H.; Witthaut, D. Discontinuous transition to loop formation in optimal supply networks. *Nat. Commun.* **2020**, *11*, 1–11. [[CrossRef](#)]
40. Lonardi, A.; Facca, E.; Putti, M.; De Bacco, C. Optimal transport for multi-commodity routing on networks. *arXiv* **2020**, arXiv:2010.14377. Available online: <http://xxx.lanl.gov/abs/2010.14377> (accessed on 28 May 2021).
41. Halu, A.; Mukherjee, S.; Bianconi, G. Emergence of overlap in ensembles of spatial multiplexes and statistical mechanics of spatial interacting network ensembles. *Phys. Rev. E* **2014**, *89*, 012806. [[CrossRef](#)] [[PubMed](#)]
42. Morris, R.G.; Barthelemy, M. Transport on coupled spatial networks. *Phys. Rev. Lett.* **2012**, *109*, 128703. [[CrossRef](#)]
43. De Domenico, M.; Solé-Ribalta, A.; Cozzo, E.; Kivela, M.; Moreno, Y.; Porter, M.A.; Gómez, S.; Arenas, A. Mathematical formulation of multilayer networks. *Phys. Rev. X* **2013**, *3*, 041022. [[CrossRef](#)]
44. Tero, A.; Takagi, S.; Saigusa, T.; Ito, K.; Bebbler, D.P.; Fricker, M.D.; Yumiki, K.; Kobayashi, R.; Nakagaki, T. Rules for Biologically Inspired Adaptive Network Design. *Science* **2010**, *327*, 439–442. [[CrossRef](#)] [[PubMed](#)]
45. Guibas, L.; Stolfi, J. Primitives for the manipulation of general subdivisions and the computation of Voronoi. *ACM Trans. Graph. (TOG)* **1985**, *4*, 74–123. [[CrossRef](#)]
46. Newman, M.E. A measure of betweenness centrality based on random walks. *Soc. Netw.* **2005**, *27*, 39–54. [[CrossRef](#)]
47. Brandes, U.; Fleischer, D. Centrality measures based on current flow. In *Annual Symposium on Theoretical Aspects of Computer Science*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 533–544.
48. Dixon, P.M.; Weiner, J.; Mitchell-Olds, T.; Woodley, R. Bootstrapping the Gini coefficient of inequality. *Ecology* **1987**, *68*, 1548–1551.
49. Kujala, R.; Weckström, C.; Darst, R.K.; Mladenović, M.N.; Saramäki, J. A collection of public transport network data sets for 25 cities. *Sci. Data* **2018**, *5*, 1–14. [[CrossRef](#)] [[PubMed](#)]
50. Alexander, L.; Jiang, S.; Murga, M.; González, M.C. Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transp. Res. Part C Emerg. Technol.* **2015**, *58*, 240–250. [[CrossRef](#)]
51. Wang, P.; Hunter, T.; Bayen, A.M.; Schechtner, K.; González, M.C. Understanding road usage patterns in urban areas. *Sci. Rep.* **2012**, *2*, 1001. [[CrossRef](#)]
52. Gao, L.; Shu, P.; Tang, M.; Wang, W.; Gao, H. Effective traffic-flow assignment strategy on multilayer networks. *Phys. Rev. E* **2019**, *100*, 012310. [[CrossRef](#)]
53. Orozco, L.G.N.; Battiston, F.; Iniguez, G.; Szell, M. Extracting the multimodal fingerprint of urban transportation networks. *Transp. Find.* **2020**, *13171*. [[CrossRef](#)]

## Designing optimal networks for multicommodity transport problem

Alessandro Ionardi<sup>1,\*</sup>, Enrico Facca,<sup>2</sup> Mario Putti,<sup>3</sup> and Caterina De Bacco<sup>1,†</sup><sup>1</sup>Max Planck Institute for Intelligent Systems, Cyber Valley, Tübingen 72076, Germany<sup>2</sup>Centro di Ricerca Matematica Ennio De Giorgi, Scuola Normale Superiore, Piazza dei Cavalieri 3, Pisa, Italy<sup>3</sup>Department of Mathematics “Tullio Levi-Civita”, University of Padua, Via Trieste 63, Padua, Italy

(Received 27 October 2020; revised 26 March 2021; accepted 14 September 2021; published 4 October 2021)

Designing and optimizing different flows in networks is a relevant problem in many contexts. While a number of methods have been proposed in the physics and optimal transport literature for the one-commodity case, we lack similar results for the multicommodity scenario. In this paper we present a model based on optimal transport theory for finding optimal multicommodity flow configurations on networks. This model introduces a dynamics that regulates the edge conductivities to achieve, at infinite times, a minimum of a Lyapunov functional given by the sum of a convex transport cost and a concave infrastructure cost. We show that the long-time asymptotics of this dynamics are the solutions of a standard constrained optimization problem that generalizes the one-commodity framework. Our results provide insights into the nature and properties of optimal network topologies. In particular, they show that loops can arise as a consequence of distinguishing different flow types, complementing previous results where loops, in the one-commodity case, were obtained as a consequence of imposing dynamical rules on the sources and sinks or when enforcing robustness to damage. Finally, we provide an efficient implementation of our model which converges faster than standard optimization methods based on gradient descent.

DOI: [10.1103/PhysRevResearch.3.043010](https://doi.org/10.1103/PhysRevResearch.3.043010)

### I. INTRODUCTION

Optimizing networks for the distribution of quantities such as passengers in a transportation network or data packets in a communication network is a relevant matter for network planners. Similar problems arise in natural systems such as river basins and vascular networks. A variety of models have been proposed to study these systems within an optimization framework [1–4]. The standard goal is to find the values of flow and the network topology that minimize a transportation cost. A common choice for this cost is the total power dissipation [1,2,5–9], but alternatives can be adopted depending on the application; see, for instance, Ref. [10]. More recently, different approaches based on a dynamical adaptation of network properties coupled with conservation laws have been proposed [5,6]. These models can be reformulated within the framework of optimal transport theory, following the work in Refs. [11–17]. Very efficient computational techniques have been developed for solving such optimal transport-based models [13–15].

In all these systems there is a unique indistinguishable flow traveling through the network. However, it may occur that

flows of different types compete in the network infrastructure; yet all the physical models mentioned above have been developed for one type of flow only. One could use these methods to analyze multicommodity problems either by aggregating together all flow types or by treating them independently. In either case, one loses the important information of how interacting commodities affect the flow, which constitutes the multicommodity character of these settings. Multicommodity-specific methods that rely on standard optimization suffer from high computational costs caused by the simultaneous assignment of multiple interacting paths to minimize a global cost function. As a consequence, existing multicommodity flow algorithms rely on ignoring these interactions, or use greedy heuristics and approximations that lead to suboptimal solutions [18]. Approaches based on statistical physics and message-passing algorithms have improved results [19,20] but remain computationally costly.

In this paper, we propose a model to design the topology of optimal networks where multiple resources are moved together. This is based on principles of optimal transport theory similar to those studied in Refs. [16,17]. Assuming potential-driven flows, this optimal design problem is posed as that of finding the distribution of multicommodity fluxes that minimize a global cost functional, or equivalently, as that of finding the optimal edge conductivities. The cost functional is the multicommodity extension of the optimal transport Lyapunov functional proposed in Refs. [14,15]. It is given by the sum of the convex cost incurred in transporting all the commodities across the network, summed to a concave cost proportional to the total flux on the network. This second term can be interpreted as the cost of building and maintaining the transport infrastructure and controls traffic congestion on the

\*alessandro.ionardi@tuebingen.mpg.de

†caterina.debacco@tuebingen.mpg.de

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by the Max Planck Society.

network edges either by distributing fluxes on many edges or by concentrating them on fewer edges following a principle of economy of scale.

Additionally, we show that the problem of minimizing the proposed cost functional is equivalent to a constrained optimization problem that generalizes the one-commodity case. The optimal distribution of fluxes is used to identify the optimal network topology by discarding edges where conductivities are small. Within this optimization framework, numerical experiments supported by analytical evidence lead to the important result that optimal network topologies may have loops as a consequence of distinguishing flow types. Generally, loops are pervasive in both natural and anthropic networks [7,21–24]. However, in one-commodity settings, several studies have shown that trees are often optimal [1,2], while few results show that loops can be obtained by fluctuating flows or by aiming at increased robustness to damage [3,5,7]. This implies either changing the type of cost function or introducing stochasticity in the sources and sinks. Instead, in our multicommodity model, loops emerge naturally as a consequence of the presence of different flow types.

In order to minimize the highly nonlinear and nonconvex cost functional mentioned before, we propose a particular set of dynamical equations for the edge conductivities, generalizing to a multicommodity scenario those proposed in Refs. [16,17], and find their stationary solution. We demonstrate that the cost functional is indeed a Lyapunov functional (i.e., it is strictly decreasing along the solution trajectories) for the proposed dynamics. Altogether, our results extend the theoretical insights of two separate lines of literature, optimal transport and network dynamics. Two principled algorithms for solving the multicommodity problem are proposed. They have similar computational complexity that largely improves on that of techniques based on gradient descent or Monte Carlo methods, thus making the model scalable to large data sets and the only computationally viable optimization alternative for large problems.

## II. MODEL

Consider a graph  $\mathcal{G}$  made of a set of  $N$  nodes  $\mathcal{V}$  interconnected by a set  $\mathcal{E}$  of  $E$  edges. We want to model transport through the network of  $M \geq 1$  commodities, each identified by a color. The inflow-outflow rate of each commodity is given by a vector  $S^i \in \mathbb{R}^N$  such that  $\sum_v S_v^i = 0$  for all  $i = 1, \dots, M$  to ensure global mass preservation. Let the “colored flux”  $F_e = (F_e^1, \dots, F_e^M)$  be a vector with entries  $F_e^i$ , which represent the commodities flux passing through edge  $e$ . In standard one-commodity cases, the flux per unit time could represent a water or an electrical current and typically is “colorless”, i.e.,  $F_e$  is a scalar quantity. In turn, the components  $F_e^i$  can be thought of as fluxes of immiscible substances traveling through the same edge. Denote with  $B$  the signed network incidence matrix, with entries  $B_{ve} = +1, -1$  if node  $v \in \mathcal{V}$  is the starting or ending point of edge  $e \in \mathcal{E}$ , respectively, and zero otherwise. We require the flux to obey the “colored” local Kirchhoff’s law:

$$\sum_{e \in \mathcal{E}} B_{ve} F_e^i = S_v^i, \quad \forall v \in \mathcal{V}, \quad \forall i = 1, \dots, M, \quad (1)$$

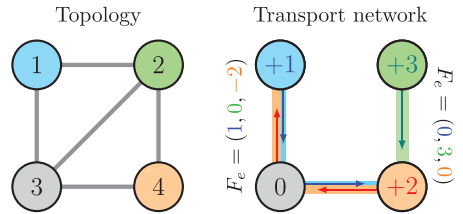


FIG. 1. Multicommodity problem illustration for  $M = 3$ . Left: topology of the graph; numbers inside nodes correspond to their indices  $v$ . Right: an admissible configuration of fluxes. Here, numbers inside nodes correspond to their mass inflow, and we assume each commodity to have its mass concentrated in a single vertex; in the gray node, no mass is entering or exiting, i.e.,  $S_v^i = 0$  for every  $i$ . Widths of edges are drawn proportional to  $F_e^i$ , and in the case where all  $F_e^i = 0$ , links are not drawn; arrows denote the direction of the colored fluxes  $F_e^i$ . Notice how blue mass and orange mass share the same edges, thus creating possible traffic congestion. The inflow-outflow rates of each color are  $S^1 = (1, 0, 0, -1)$  (blue),  $S^2 = (0, 3, 0, -3)$  (green), and  $S^3 = (-2, 0, 0, 2)$  (orange).

where each edge  $e = (u, v)$  has length  $\ell_e > 0$ . We could assume that the components (colors) of the flux derive from differences in a colored potential (pressure) defined on nodes  $p_v^i$ , and a colored conductivity  $\mu_e^i$ :

$$F_e^i = \frac{\mu_e^i}{\ell_e} (p_u^i - p_v^i). \quad (2)$$

The commodity index  $i$  can be any arbitrary attribute of the mass traveling through the network without impacting the validity of our model. In fact, the important idea behind multicommodity optimization is that different types of mass interact while being transported in a shared infrastructure, and a suitable cost needs to be minimized. In Fig. 1 we show a simple example of the model construction.

Up to this point, we have a set of independent one-commodity flows, one per color  $i$ . Taking them separately and then superimposing each individual flux or conductivity *a posteriori* would be a naive strategy, neglecting possible complex interactions. For example, it may be more convenient to gather multiple flows through one channel with high capacity (the conductivity  $\mu_e^i$ ). More generally, the optimal network design mechanism must take into account all commodities at once. Deciding how this should be done is an open problem in the context of optimal transport theory, the approach we take here.

### A. Introducing a shared conductivity

Our first model assumption is that all the conductivities must be equal, namely,

$$\mu_e^i = \hat{\mu}_e, \quad \forall e \in \mathcal{E}, \quad \forall i = 1, \dots, M. \quad (3)$$

The quantity  $\hat{\mu}_e$  plays the role of a colorless conductivity. Given that the conductivity can be seen as proportional to the size of an edge, Eq. (3) can be interpreted as allocating the same edge capacity for the different colors. This is a reasonable assumption in systems for which there is no

priority between commodity types or users. In communication networks such as the Internet, this captures the situation where all users share the same bandwidth, and no privileged user exists who has access to more bandwidth, which is often the case. Notice, however, that the flux  $F_e^i$  still depends on the color, because the difference in potential does. This implies that users can transfer different amounts of data packages, with potential for traffic congestion when they overload the network, e.g., when streaming videos. This is one of the many possible alternatives of coupling between colors. Other more complex choices could be made, for instance, by introducing an explicit coupling involving the fluxes or opting for controlling some global functions of the conductivities, e.g., their sum or the magnitude of their fluctuations across colors. However, we find that our choice, while being analytically convenient, allows for a rigorous generalization of the one-commodity case with fixed and fluctuating loads [1–3,5,7] and leads to rich topological behaviors, as we show below.

### B. The dynamics

Having defined how colors move through the network, we now turn our attention to describing the mechanism to design the network. Formally, we propose the following dynamics for the colorless conductivity:

$$\dot{\hat{\mu}}_e(t) = \hat{\mu}_e^\beta(t) \frac{\|\Delta P_{uv}\|_2^2}{\ell_e^2} - \hat{\mu}_e(t), \quad \forall e = (u, v) \in \mathcal{E}, \quad (4)$$

where we define  $\Delta P_{uv}$  as a vector of pressure differences with entries  $\Delta P_{uv}^i = p_u^i - p_v^i$  and  $\|\Delta P_{uv}\|_2^2 = \sum_{i=1}^M (p_u^i - p_v^i)^2$ . Note that  $p^i$ , and thus  $\Delta P_{uv}$ , are implicit functions of  $\hat{\mu}(t)$  because of Eqs. (1) and (2).

The parameter  $\beta$  determines the type of optimization associated with this dynamics. In the standard one-commodity case, for  $\beta < 1$  one aims at minimizing traffic congestion and obtains loopy topologies; for  $\beta > 1$  the aim is to consolidate paths, and optimal networks are trees. The case  $\beta = 1$  is shortest-path-like. This dynamics describes a feedback mechanism. If the total flux through an edge is large, its conductivity increases. If the flux decreases, the conductivity decreases over time and becomes negligible when no flux occurs. The system of Eqs. (1)–(4) represents our model for multicommodity flow optimization.

In the presence of only one commodity, our model is similar to the dynamics used to solve the basis pursuit problem on networks [15] and as a principled mechanism for filtering networks from redundancies [16]. However, both cases are limited to one-commodity scenarios. A similar dynamics is also proposed in Ref. [5], where the authors focus on the average time evolution of a stochastic model with fluctuating loads. Analogously to these one-commodity cases, one can efficiently solve the system in Eqs. (1)–(4) using optimized numerical methods; however, in our case the complexity increases with the number of colors (see Appendix E2 for more details). Our model also bears a close mathematical relationship to a recent work, where similar ideas have been studied in a multicommodity setup [17]. Beyond the fact that this work focuses on the case  $\beta = 1$  (i.e., shortest-path-like) and thus on a convex optimization scenario, there is one other main conceptual difference compared with our model. Notice

that Eq. (4) couples together the various colors by means of  $f(\Delta P_{uv}) = \|\Delta P_{uv}\|_2^2$ , i.e., the 2-norm squared of the pressure difference. Instead, they consider the 1-norm and 2-norm (not squared). Analyzing the solutions of the dynamics under different  $f(\Delta P_{uv})$  is an interesting avenue for future work.

The key insight of optimal transport theory is that Eq. (4) admits a Lyapunov functional (a functional decreasing in time along solution trajectories) having the nice interpretation of being the transportation cost:

$$\mathcal{L}_\beta(\{\hat{\mu}_e\}) = \frac{1}{2} \sum_{i,v} p_v^i(\{\hat{\mu}_e\}) S_v^i + \frac{\sum_e \ell_e \hat{\mu}_e^{2-\beta}}{2(2-\beta)}, \quad (5)$$

where  $p_v^i(\{\hat{\mu}_e\})$  is a function implicitly defined as the solution of Eqs. (1) and (2) when imposing Eq. (3). The first term corresponds to the energy dissipated during transport, and it can be interpreted as the operating costs, whereas the second term is the cost of designing the infrastructure. The equilibrium point of  $\hat{\mu}_e$  is stationary at the previous Lyapunov functional, and for  $\beta \leq 1$  it acts also as the global minimizer due to its convexity. For  $\beta > 1$ , while the first term (operating cost) is convex, the second (infrastructural cost) is not. As a consequence, the transportation cost is not convex; thus in general the functional will present a rich landscape with several local minima towards which the dynamics will be attracted.

We formally show that Eq. (5) defines a well-defined Lyapunov functional for the dynamics of Eq. (4) in Appendix A1, following similar arguments as in Ref. [17]. This extends the work of Bonifaci *et al.* [11], where a similar functional has been proposed to complete the characterization of the dynamics regulating slime molds' evolutionary feedback mechanism.

### C. Mapping to standard optimization setups

Although not evident, our dynamics is connected with an optimization problem analogous to previous models for the one-commodity case [1,2]. Specifically, the stationary solutions of our system minimize the network total transportation cost  $J = \frac{1}{2} \sum_{e \in \mathcal{E}} \frac{\ell_e}{\hat{\mu}_e} \|F_e\|_2^2$  subject to the global constraint of constant material cost  $\sum_{e \in \mathcal{E}} \ell_e \hat{\mu}_e^{2-\beta} = K^{2-\beta}$  and local Kirchhoff's law on nodes as in Eq. (1) [using Kirchhoff's law, one can show that  $J$  is equivalent to the first term in Eq. (5); for more details, see Appendix A2]. Formally, the optimization problem is

$$\{\hat{\mu}_e^*, \{F_e^*\}\} = \arg \min_{\{\hat{\mu}_e, \{F_e\}\}} \left\{ \frac{1}{2} \sum_{e \in \mathcal{E}} \frac{\ell_e}{\hat{\mu}_e} \|F_e\|_2^2 \right\}, \quad (6)$$

such that

$$\sum_{e \in \mathcal{E}} \ell_e \hat{\mu}_e^{2-\beta} = K^{2-\beta}, \quad (7)$$

$$\sum_{e \in \mathcal{E}} B_{ve} F_e^i = S_v^i, \quad \forall v \in \mathcal{V}, \quad \forall i = 1, \dots, M. \quad (8)$$

This optimization problem is analogous to that in Ref. [2], except here the flux appears in terms of its 2-norm. As in the one-commodity case, this leads to an optimal configuration where the conductivities similarly scale with the fluxes,

$$\hat{\mu}_e \sim \|F_e\|_2^{2/(3-\beta)}, \quad (9)$$

and the proportionality constant can be fully determined analytically (see Appendix B for detailed derivations). Using Eq. (9), we can rewrite the total transportation cost in terms of the flux as

$$J_\Gamma = \sum_{e \in \mathcal{E}} \ell_e \|F_e\|_2^\Gamma, \quad (10)$$

where  $\Gamma = 2(2 - \beta)/(3 - \beta)$ , which is analogous to the optimization problem of Banavar *et al.* [1], where there was no conductivity in the setup. Notice that all these results generalize the one-commodity case [1,2] by means of the 2-norm  $\|F_e\|_2$  of the colored flux. If there were only one color, and thus  $\|F_e\|_2^2 = F_e^2$ , our model would reduce exactly to them. Similar relations can be obtained with a stochastic approach such as the one proposed in Refs. [3,7,8], but by considering ensemble averages instead of the 2-norm of the fluxes. In these works, the authors study a setup where sources' and sinks' positions are extracted randomly from a distribution on the network nodes. They also find loops in nontrivial regimes. While the mathematical formulations show some similarities, there are main conceptual differences between these models and ours. These approaches are stochastic; thus the main quantities are calculated with ensemble averages, and loops arise as a consequence of stochastic fluctuations or when randomly cutting edges in the network. Instead, our problem is deterministic, and loops arise as a result of an optimization process while assuming a shared conductivity.

Solving this optimization problem directly by means of gradient descent is computationally expensive (see Appendix E 1). Methods relying on Monte Carlo schemes [2] can also be computationally demanding, and they are valid only when the optimal topology is known to be a tree. Instead, we derive update rules which have similar complexity to that of finding the steady states of our dynamics, and can be implemented with efficient numerical solvers. They consist in iterating between updating conductivities and fluxes as

$$\hat{\mu}_e = \frac{\|F_e\|_2^{2/(3-\beta)}}{(\sum_e \ell_e \|F_e\|_2^{2(2-\beta)/(3-\beta)})^{1/(2-\beta)}} K, \quad (11)$$

$$F_e^i = \frac{\hat{\mu}_e}{\ell_e} (p_a^i - p_b^i), \quad (12)$$

complemented with Kirchhoff's law in Eq. (1), and can be put within the framework of fixed-point iterations. This generalizes results obtained adopting a similar approach for the one-commodity case [2,3]. We make available an open-source implementation of the two approaches which we summarize here: finding the steady state of the dynamics by solving the system of Eqs. (1)–(4) (Dynamics) and extracting the solution of the optimization problem with the iterative updates of Eqs. (11) and (12) (Optimization). We provide a pseudocode for each of these in Algorithms 1 and 2 in Appendix E. They have similar computational complexity that scales as  $O(MN^2)$  and are much faster than techniques based on gradient descent; see Appendix E 1.

### III. ANALYSIS OF THE OPTIMAL TOPOLOGIES

#### A. Optimal topologies may have loops

Now, we address the important question of which network topologies are optimal for the cost in Eq. (10). For the analogous models in the one-commodity case, there is a phase transition at  $\beta = \Gamma = 1$  where optimal networks pass from being trees ( $1 < \beta < 2, 0 < \Gamma < 1$ ) to containing loops ( $0 < \beta < 1, 1 < \Gamma < 4/3$ ) [1,2]; see Ref. [9] for a thorough investigation of this transition. Remarkably, we obtain that in the multicommodity case, loopy structures can be optimal also in the regime where trees were optimal in the previous models, depending on the values and locations of sources  $S_i^j$  and on the edge lengths  $\ell_e$ .

The loopy structures in what was previously a treelike regime arise from the colored Kirchhoff's law (1), distinguishing different commodities entering and exiting a node. Had we imposed a similar but colorless constraint  $\sum_i \sum_e B_{ve} F_e^i = \sum_i S_i^v$ , trees would have been optimal.

#### B. Phase diagram tree-loops ( $S$ fixed)

To illustrate this, we consider the simple triangular loop  $\mathcal{G}(\mathcal{V} = \{1, 2, 3\}, \mathcal{E} = \{a, b, c\})$  represented in Fig. 2, with  $M = 2$  commodities moving in the network and lengths  $\ell = (\ell_a, \ell_b, \ell_c)$ . For simplicity we focus on the phase diagram in  $\{\ell_e\}$  by fixing  $S$ , but similar reasoning applies when doing the opposite. We set  $S^1 = (+1, -1, 0), S^2 = (-1, +2, -1)$ . For this simple case, Kirchhoff's law allows only for three possible tree topologies  $\mathcal{T}_i, i = 1, 2, 3$ ; these are shown at the bottom right of Fig. 2. By solving Kirchhoff's law,

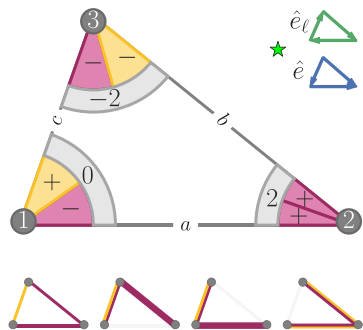


FIG. 2. Toy model where loops are optimal. Here,  $M = 2$ ; hence only two colors move through the network. The triangle network has source vectors  $S^1 = (+1, -1, 0), S^2 = (-1, +2, -1), \ell_a = \ell_b = 1.5, \ell_c = 1$ . The gray patches denote the net loads of each node when ignoring the colors. At the bottom we show one loopy solution on the left and three trees on the right. The green and blue arrows denote the orientation of the loop defined in Appendix D and of the edges, respectively. In detail, the loop has fluxes  $F_a = (0, -1), F_b = (0, -1), F_c = (-1, 0)$ , the leftmost tree has  $F_a = (0, 0), F_b = (0, -2), F_c = (-1, +1)$  (similarly for the other two). The green star refers to the topology of the toy model used in Fig. 3.

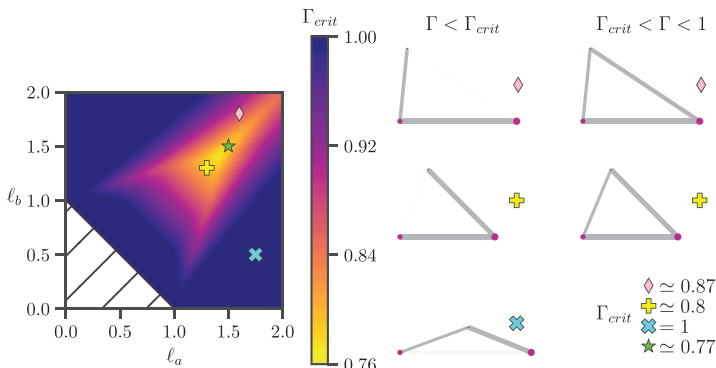


FIG. 3. Phase diagram in  $\ell$ .  $\Gamma_{crit}$  denotes the minimum value of  $\Gamma$  above which loops are optimal. The setup is the same as for the toy model in Fig. 2. Values of  $\Gamma_{crit}$  are found by solving Eq. (13);  $\ell_c = 1$ . The area under the triangular surface with white background is not allowed as the triangular geometry is not defined there. One can notice that there is an entire region where  $\Gamma_{crit} \leq 1$ ; inside it, loops are optimal. On the right-hand side of the figure we show three different topologies, i.e., choices of  $\ell_a, \ell_b$  for which optimal solutions can be loopy; these are associated with the markers drawn on the heat map, and the green star is the configuration of the toy model in Fig. 2. In particular, running our dynamics on the pink-diamond graph (respectively, the yellow-plus graph) leads to a loopy configuration for  $\Gamma > \Gamma_{crit}$  or to a tree if  $\Gamma < \Gamma_{crit}$ . Running our dynamics on the blue-cross graph returns always since its  $\Gamma_{crit}$  is equal to 1. Widths of the edges are proportional to the final  $\|F_i\|_2$ , and edges that are not visible have negligible fluxes; nodes' dimensions are proportional to their inflowing mass. In the bottom right portion of the panel we report the values of  $\Gamma_{crit}$  obtained solving Eq. (13) fixing  $\ell_a, \ell_b$  as given by the markers.

we can write all the fluxes as a function of  $F_a = (F_a^1, F_a^2)$ . Then, by choosing two arbitrary values of  $F_a^1, F_a^2$  we propose a loopy solution  $\mathcal{G}_L$  to compare against the trees; this is the leftmost bottom triangle in Fig. 2. We show that there are values of  $0 < \Gamma < 1$  for which this loopy solution has lower transportation cost than any of the trees. One can compute all the costs using Eq. (10) (see Appendix C for details) and then find values of  $\{\ell^* = (\ell_a^*, \ell_b^*, \ell_c^*), F_a^*, \Gamma^*\}$  for which

$$J_\Gamma(\mathcal{G}_L; \ell^*, F_a^*, \Gamma^*) \leq \min\{J_\Gamma(\mathcal{T}_i; \ell^*, F_a^*, \Gamma^*) : i = 1, 2, 3\} \tag{13}$$

holds. To find an example solution, one could fix certain values for these parameters and then numerically solve Eq. (13); for a few simple cases this can also be done analytically. We show an example phase diagram obtained by varying  $\ell$  in Fig. 3, where we plot the values of  $\Gamma_{crit}$  such that for  $\Gamma \geq \Gamma_{crit}$ , the cost  $J_\Gamma(\mathcal{G}_L)$  is optimal, i.e., we have a phase transition between trees to loopy optimal topologies. Notice that such values of  $\Gamma_{crit}$  depend on the selected values of  $(\ell, F_a)$  and that optimal loopy solutions are not guaranteed to exist for any arbitrary configuration of these values. This can be numerically investigated using similar reasoning to that used for the case above. The important point here is that we could find *at least* one setting of  $(\ell, F_a)$  for which we have loopy solutions in the nontrivial regime  $0 < \Gamma < 1$ . Similar arguments can be used to find phase diagrams in  $S$  when fixing  $\ell$  (see Appendix C, Fig. 5).

**C. Phase diagram tree-loops (lengths fixed)**

To make more clear the consequences of the implicit interaction between different fluxes when imposing a shared conductivity and the optimization process is run, we show results on a simple synthetic toy model where we vary the load of one color while keeping the others fixed.

Specifically, we study the triangle topology of Fig. 2 and consider two different configurations of  $S$ . The first has  $S_1^1 = -S_3^1 = 1$  for the yellow commodity, i.e., one unit of mass of type  $i = 1$  is moving from node 1 to node 3, while the purple commodity has  $S_2^2 = +2$  and  $S_3^2 = S_1^2 = -1$ , i.e., two units of mass of type  $i = 2$  are injected in node 2 and are equally split between destination nodes 1 and 3. This corresponds to the green triangle in the phase diagram of Fig. 4(a). The second configuration has the same sources and sinks for the yellow commodity, while the purple mass is doubled, i.e.,  $S_2^2 = +4$  and  $S_3^2 = S_1^2 = -2$ . This corresponds to the red triangle in Fig. 4(a). As we can see from Fig. 4(b), both the optimal network topologies and the fluxes of individual colors differ in the two configurations. The important point here is that the fluxes of the yellow commodity change, even though its forcing  $S^1$  does not change between the two configurations. This is a consequence of having distinct commodities sharing a common infrastructure: Acting solely on the purple mass impacts also the path taken by the yellow mass and, consequently, the overall optimal network topology.

Additionally, the way the topology changes between these two configurations depends on the exponent  $\Gamma$ . In this simple scenario, we can have either a tree or a loop at  $\Gamma = 0.6$  ( $\Gamma < \Gamma_{crit} \approx 0.77$ ) or  $\Gamma = 0.8$  ( $\Gamma > \Gamma_{crit}$ ), respectively; see



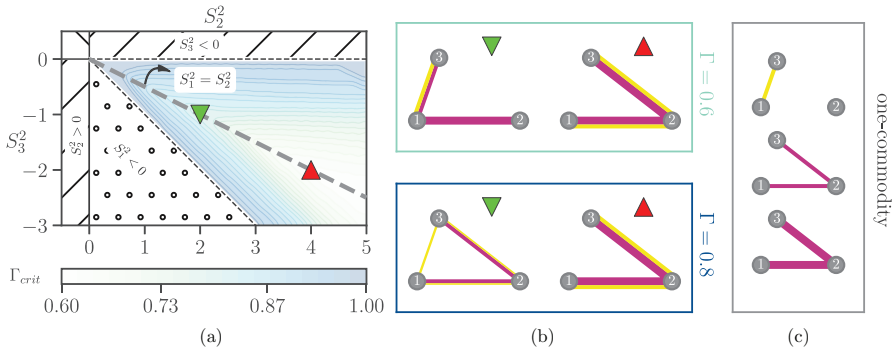


FIG. 4. Interaction between commodities. The figure shows how changing the load of one commodity influences the path taken by others. (a) We plot the heat map of  $\Gamma_{crit}$  obtained using Eq. (13), with  $S_1^i = -S_3^i = +1$  and for different configurations of the purple commodity ( $i = 2$ ): (i)  $S_2^2 = +2, S_3^2 = S_1^2 = -1$ ; (ii)  $S_2^2 = +4, S_3^2 = S_1^2 = -2$ . The areas under the white surface correspond to regions where node 1 is a sink, respectively, nodes 2 and 3 are sources, for the purple mass. The green and the red triangular markers denote the configurations we discuss in Sec. III C and in the rest of the panel. (b) Optimal graphs for the green and the red triangle, fixing  $\Gamma = 0.6$  and  $\Gamma = 0.8$ . The width of each edge  $e$  is proportional to  $|F_e^i|$  for each color. (c) One-commodity solutions obtained injecting only the yellow or the purple mass in the triangle network.

Fig. 4(b). In particular, the case of  $\Gamma = 0.8$  is a simple example of how the routing mechanism is responsible for the generation of loops in a multicommodity setting. Finally, if we were to consider two separate unicommodity scenarios and solve the optimization for the two colors independently, we would have obtained a different result, as shown in Fig. 4(c). In this case, the yellow remains the same in the two configurations, while the purple would simply double the amount of fluxes along edges, but the set of edges being used would stay the same.

In addition to the numerical analysis presented to study the generation of loops, in Appendix D we adapt to our colored case the proof of Proposition 2.1 given by Xia [25], where it was demonstrated that one-commodity (i.e., colorless Kirchhoff’s law) optimal transport paths are trees. Here, we show that for our model, optimal networks may contain loops.

#### IV. CONCLUSIONS

Although we have a rigorous theoretical understanding of the behavior of one-commodity flows in networks, comparable theoretical insights for flows of different types have been lacking. Here, we propose a model for multicommodity flows that extends and generalizes various results obtained for the one-commodity case. It assumes that all the commodities have the same priority by imposing their conductivities to be equal and that their dynamics is regulated by the 2-norm squared of the fluxes. By drawing from theoretical results of optimal transport theory, the equilibrium solutions of our dynamics are also stationary points of a cost function that can be interpreted as the sum of operating and infrastructural costs. As we tune a parameter  $\beta$ , our dynamics can solve various types of routing optimization problems. Its numerical implementation is efficient and scalable to large systems.

Remarkably, our model shows how optimal loopy topologies can arise from simple dynamical rules. We explain how this emerges as a consequence of the colored Kirchhoff’s law and how the theoretical proof valid in the one-commodity case fails when fluxes are vectors. We provide example phase diagrams on a simple toy model that illustrates how optimal topologies evolve from being trees to containing loops.

Our model is applicable to all situations where it is relevant to distinguish flow types and to consider how these interact. One important example of such an instance is in communication networks where packets of information need to be delivered at different destinations.

In our formulation the underlying network topology is given in terms of sets of nodes and edges. While our model allows for edge removal (and node removal as a consequence), it does not provide a mechanism for adding new connections. In order to allow for this, the natural modification of our approach would be to consider a continuous formulation as in Refs. [15,16]. In this case, we would have no underlying topology to start with, except the presence of source and sink nodes at given locations in space. This is an interesting direction for future work.

In addition to solving multicommodity problems, our model allows us to draw a rigorous mapping between two different formalisms. In fact, while both the physics and optimal transport communities are actively investigating these systems, we still miss a clear connection between them, even for the one-commodity case. We make a first attempt to fill this gap by showing how our dynamics maps to a standard optimization setup while also generalizing to the multicommodity case. Furthermore, we deploy two numerical methods that have lower computational complexity compared with others based on gradient descent.

We expect that our formalism can be further extended in the future to accommodate more sophisticated interaction

between commodities or in multilayer networks [26] thus better representing specific application scenarios. Similarly, modifying the dependence of the fluxes in driving the dynamics and investigating possible mappings to suitable optimization setups are natural next steps.

We foresee that the insights gained into the structure of optimal topologies and in combining principles of optimal transport and physics will open the way to further studies targeting these systems. To facilitate this, we provide an open-source implementation of our code [27].

### ACKNOWLEDGMENTS

We thank Kurt Mehlhorn for useful discussions and for the helpful advice about the proof of the Lyapunov functional. We acknowledge the help of Daniela Leite for data preprocessing. The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Alessandro Lonardi.

### APPENDIX A: THE LYAPUNOV FUNCTIONAL

#### 1. The Lyapunov functional is well defined

Here, we prove that the functional proposed in Eq. (5) is a Lyapunov for the dynamics in Eq. (4) for  $0 < \beta < 2$ . To do that, we follow the derivations proposed in Ref. [17] for a similar problem. We need to show that (i)  $\mathcal{L}_\beta \geq 0$ , (ii)  $\dot{\mathcal{L}}_\beta \leq 0$  and  $\mathcal{L}_\beta = 0$  if and only if  $\{\hat{\mu}_e\}$  is a stationary point for the dynamics. The first condition is trivial. In order to prove the second one, we first define the quantity

$$L_{vu} = \sum_e B_{ve}(\hat{\mu}_e/\ell_e)B_{ue}, \quad (\text{A1})$$

which is the entry  $(u, v)$  of the Laplacian of a graph with adjacency matrix with entries  $A_{uv} = \hat{\mu}_{uv}/\ell_{uv}$ . We can thus rewrite Eqs. (1)–(3) as

$$\sum_{e,u} B_{ve} \frac{\hat{\mu}_e}{\ell_e} B_{ue} p_u^i = S_v^i, \quad \forall v \in \mathcal{V}, \quad \forall i = 1, \dots, M, \quad (\text{A2})$$

$$\sum_u L_{vu} p_u^i = S_v^i, \quad \forall v \in \mathcal{V}, \quad \forall i = 1, \dots, M. \quad (\text{A3})$$

Now, we claim that for each edge

$$\partial_{\hat{\mu}_e} \mathcal{L}_\beta = \frac{1}{2} \left[ \ell_e \hat{\mu}_e^{1-\beta} + \partial_{\hat{\mu}_e} \left( \sum_{i,u} S_u^i p_u^i \right) \right] \quad (\text{A4})$$

$$\stackrel{\text{claim}}{=} \frac{\ell_e}{2} \left( \hat{\mu}_e^{1-\beta} - \frac{\|\Delta P_e\|_2^2}{\ell_e^2} \right). \quad (\text{A5})$$

This identity can be obtained differentiating for  $\hat{\mu}_e$  both sides of Eq. (A3). This yields for all  $e, v, i$  the following:

$$\sum_u (\partial_{\hat{\mu}_e} L_{vu}) p_u^i + \sum_u L_{vu} (\partial_{\hat{\mu}_e} p_u^i) = 0, \quad (\text{A6})$$

$$\sum_u L_{vu} (\partial_{\hat{\mu}_e} p_u^i) = - \sum_u B_{ve} (1/\ell_e) B_{ue} p_u^i. \quad (\text{A7})$$

Multiplying both sides of Eq. (A7) by  $p_v^i$ , summing over  $v$ , and exploiting again Eq. (A3) on the left-hand side, we

obtain

$$\sum_{v,u} p_v^i L_{vu} (\partial_{\hat{\mu}_e} p_u^i) = - \sum_{v,u} p_v^i B_{ve} (1/\ell_e) B_{ue} p_u^i, \quad (\text{A8})$$

$$\partial_{\hat{\mu}_e} \left( \sum_u S_u^i p_u^i \right) = - \ell_e \frac{(\Delta P_e^i)^2}{\ell_e^2}, \quad (\text{A9})$$

with  $\Delta P_e$  being an  $M$ -dimensional vector of entries  $\Delta P_e^i = \Delta P_{uv}^i = p_u^i - p_v^i$ , with  $e = (u, v)$ . Summing over  $i$  gives

$$\partial_{\hat{\mu}_e} \left( \sum_{i,u} S_u^i p_u^i \right) = - \ell_e \frac{\|\Delta P_e\|_2^2}{\ell_e^2}; \quad (\text{A10})$$

notice that the term in parentheses on the left-hand side is exactly the “operating cost” of Eq. (5). Using Eq. (A10), the claim in Eq. (A5) immediately follows. It is in force of Eq. (A5) that we see that the Lie derivative of  $\mathcal{L}_\beta$  is not positive. Namely,

$$\dot{\mathcal{L}}_\beta = \sum_e (\partial_{\hat{\mu}_e} \mathcal{L}_\beta) \hat{\mu}_e \quad (\text{A11})$$

$$= - \sum_e \frac{\ell_e}{2} \hat{\mu}_e^\beta \left( \hat{\mu}_e^{1-\beta} - \frac{\|\Delta P_e\|_2^2}{\ell_e^2} \right)^2 \leq 0, \quad (\text{A12})$$

where  $\hat{\mu}_e$  has been substituted with the right-hand side of Eq. (4). Moreover,  $\dot{\mathcal{L}}_\beta = 0$  if and only if  $\hat{\mu}_e = 0$  or  $\hat{\mu}_e^{3-\beta} = \hat{\mu}_e^2 \|\Delta P_e\|_2^2 / \ell_e^2 = \|F_e\|_2^2$ . This exact condition can be recovered setting  $\hat{\mu}_e = 0$  in Eq. (4) and exploiting Eq. (2). In particular, for each  $e$  edge we get

$$\hat{\mu}_e^{1-\beta} = \frac{\sum_i (p_u^i - p_v^i)^2}{\ell_e^2} = \sum_i \left( \frac{\ell_e F_e^i}{\hat{\mu}_e} \right)^2 \frac{1}{\ell_e^2}, \quad (\text{A13})$$

$$\hat{\mu}_e^{3-\beta} = \|F_e\|_2^2. \quad (\text{A14})$$

#### 2. Equivalence between the Lyapunov transportation cost and the dissipated energy

We prove that the transportation cost  $J = \frac{1}{2} \sum_e \frac{\ell_e}{\hat{\mu}_e} \|F_e\|_2^2$  is indeed identical to the first term of the Lyapunov functional of Eq. (5). In fact, combining Eqs. (1)–(3), we can rewrite Kirchhoff's law as

$$\sum_{e,u} B_{ve} \frac{\hat{\mu}_e}{\ell_e} B_{ue} p_u^i = S_v^i, \quad \forall v \in \mathcal{V}, \quad \forall i = 1, \dots, M. \quad (\text{A15})$$

Multiplying both sides of the equation for  $p_v^j$  and summing over  $i$  and  $v$  yields

$$\sum_e \frac{\ell_e}{\hat{\mu}_e} \|F_e\|_2^2 = \sum_{i,v} p_v^i S_v^i, \quad (\text{A16})$$

which is the equality we wanted to show.

### APPENDIX B: MAPPING THE DYNAMICS TO AN OPTIMIZATION PROBLEM

We show that a constrained optimization problem with a cost function representing the total dissipated energy over the whole network has a solution with the same scaling as in Eq. (9).

Formally, consider the constrained optimization problem of Eqs. (6)–(8). This can be turned into an unconstrained optimization problem by introducing Lagrange multipliers:

$$J_\beta(\{\hat{\mu}_e\}, \{F_e\}) = \frac{1}{2} \sum_e \frac{\ell_e}{\hat{\mu}_e} \|F_e\|_2^2 + \frac{\lambda}{2(2-\beta)} \left( \sum_e \ell_e \hat{\mu}_e^{2-\beta} - K^{2-\beta} \right) + \sum_{v,i} \chi_v^i \left( \sum_e B_{ve} F_e^i - S_v^i \right). \quad (\text{B1})$$

Here, we introduced a multiplicative factor  $1/2(2-\beta)$  for the Lagrange multiplier  $\lambda$  to ease calculations. Taking the partial derivatives with respect to  $\hat{\mu}_e$  and setting them to zero (the optimality condition on the derivative of  $J_\beta$  with respect to  $F_e$  will be treated later on) yields, for each edge,

$$\lambda \hat{\mu}_e^{3-\beta} = \|F_e\|_2^2 \rightarrow \hat{\mu}_e = \frac{1}{\lambda^{1/(3-\beta)}} \|F_e\|_2^{2/(3-\beta)}. \quad (\text{B2})$$

This is the same scaling relationship obtained from the stationary state of the dynamics in Eq. (9), up to a multiplicative constant. It is also the natural colored generalization of the one-commodity case presented in Refs. [1,2,10], where instead of having  $\|F_e\|_2$  one has the absolute value  $|F_e|$ , as  $F_e$  is a scalar quantity there. Imposing the global constraint in Eq. (7) allows us to determine the value of the multiplier  $\lambda$ :

$$\sum_e \ell_e \hat{\mu}_e^{2-\beta} = \sum_e \ell_e \frac{\|F_e\|_2^{2(2-\beta)/(3-\beta)}}{\lambda^{2(2-\beta)/(3-\beta)}} = K^{2-\beta}, \quad (\text{B3})$$

yielding

$$\lambda = \frac{1}{K^{3-\beta}} \left( \sum_e \ell_e \|F_e\|_2^{2(2-\beta)/(3-\beta)} \right)^{(3-\beta)/(2-\beta)}. \quad (\text{B4})$$

Substituting back into Eq. (B2), we obtain

$$\hat{\mu}_e = \frac{\|F_e\|_2^{2/(3-\beta)}}{(\sum_e \ell_e \|F_e\|_2^{2(2-\beta)/(3-\beta)})^{1/(2-\beta)}} K. \quad (\text{B5})$$

Setting  $\gamma = 2 - \beta$ , we get the scaling

$$\hat{\mu}_e \sim (\|F_e\|_2^2)^{1/(1+\gamma)}, \quad (\text{B6})$$

which is analogous to that of the one-commodity case in Eq. (5) of Ref. [2]. The same exact scaling can be recovered from our dynamics by setting  $\hat{\mu}_e = 0$  as shown in Eqs. (A13) and (A14).

The total dissipation is obtained by substituting Eq. (B5) inside Eq. (6), leading to

$$J_\beta = \frac{1}{2K} \left( \sum_e \ell_e \|F_e\|_2^{2(2-\beta)/(3-\beta)} \right)^{(3-\beta)/(2-\beta)} \quad (\text{B7})$$

$$= \frac{1}{2K} \left( \sum_e \ell_e \|F_e\|_2^{2\gamma/(1+\gamma)} \right)^{(1+\gamma)/\gamma}. \quad (\text{B8})$$

This cost is again analogous to that of the one-commodity case: Eq. (6) of Ref. [2] for  $\gamma = 2 - \beta$ . Using similar arguments, i.e., noticing that the function  $x^{(3-\beta)/(2-\beta)} = x^{(1+\gamma)/\gamma}$

is monotonically increasing for  $0 < \gamma = 2 - \beta < 2$ , the original minimization problem reduces to that of minimizing with respect to  $\{F_e\}$  the cost of Eq. (10):

$$J_\Gamma(\{F_e\}) = \sum_e \ell_e \|F_e\|_2^{2(2-\beta)/(3-\beta)} = \sum_e \ell_e \|F_e\|_2^\Gamma, \quad (\text{B9})$$

where  $\Gamma = 2(2-\beta)/(3-\beta) = 2\gamma/(1+\gamma)$ , which is analogous to the model of Banavar *et al.* [1]. Lastly, we can set to zero also the derivative with respect to  $F_e^i$  in Eq. (B1):

$$\frac{\partial J_\beta}{\partial F_e^i} = \frac{\ell_e}{\hat{\mu}_e} F_e^i + B_{ue} \chi_u^i + B_{ve} \chi_v^i \quad (\text{B10})$$

$$= \frac{\ell_e}{\hat{\mu}_e} F_e^i + B_{ue} (\chi_u^i - \chi_v^i) \stackrel{!}{=} 0, \quad (\text{B11})$$

$$\rightarrow F_e^i = -\frac{\hat{\mu}_e}{\ell_e} B_{ue} (\chi_u^i - \chi_v^i), \quad (\text{B12})$$

recovering the classical result stating that the pressure  $p$  is (minus) the Lagrange multiplier obtained when we minimize the dissipated energy  $J_\beta(\{\hat{\mu}_e\}, \{F_e\})$  under the Kirchhoff's law constraints.

### APPENDIX C: PHASE DIAGRAM FOR A TOY MODEL

Here, we discuss in more detail the computations described in Sec. III B to enforce the claim that networks with loops can be optimal for  $0 < \Gamma < 1$ . The simple triangular loop of Fig. 2 (top) admits three possible tree topologies  $\mathcal{T}_i$ ,  $i = 1, 2, 3$ , drawn at the bottom right of Fig. 2. Exploiting Kirchhoff's law, we write the fluxes as a function of  $F_a = (F_a^1, F_a^2)$ . Then, computing all the costs using Eq. (10), we get

$$J_\Gamma(\mathcal{T}_1) = 2^\Gamma \ell_b + 2^{\Gamma/2} \ell_c, \quad (\text{C1})$$

$$J_\Gamma(\mathcal{T}_2) = 2^\Gamma \ell_a + 2^{\Gamma/2} \ell_c, \quad (\text{C2})$$

$$J_\Gamma(\mathcal{T}_3) = (\ell_a + \ell_b) 2^{\Gamma/2}, \quad (\text{C3})$$

$$J_\Gamma(\mathcal{G}_L) = [ (F_a^1)^2 + (F_a^2)^2 ]^{\Gamma/2} \ell_a + [ (F_a^1)^2 + (F_a^2 + 2)^2 ]^{\Gamma/2} \ell_b + [ (F_a^1 - 1)^2 + (F_a^2 + 1)^2 ]^{\Gamma/2} \ell_c. \quad (\text{C4})$$

Thus we need to find values of  $\{\ell^* = (\ell_a^*, \ell_b^*, \ell_c^*), F_a^*, \Gamma^*\}$  for which Eq. (13) is satisfied. In practice, the lengths are usually given in input, and thus we set  $\ell_a = \ell_b = \frac{3}{2} \ell_c$ ,  $\ell_c = 1$ ; we then propose  $F_a^* = (0, -1)$ . Thus

$$J_\Gamma(\mathcal{T}_1) = J_\Gamma(\mathcal{T}_2) = \frac{3}{2} \times 2^\Gamma + 2^{\Gamma/2}, \quad (\text{C5})$$

$$J_\Gamma(\mathcal{T}_3) = 3 \times 2^{\Gamma/2}, \quad (\text{C6})$$

$$J_\Gamma(\mathcal{G}_L) = 4. \quad (\text{C7})$$

Analytically from Eqs. (C5)–(C7) or numerically solving Eq. (13) (fixing  $F_a^*$  as proposed for  $\mathcal{G}_L$ ) one can show that  $\Gamma^* \simeq 0.83$ . Notice that such  $\Gamma^*$  is not optimal, in the sense that for other choices of  $F_a$  we may find lower values of the exponent  $\Gamma$  enabling loopy networks to be optimal; we denote the minimum of these values as  $\Gamma_{\text{crit}}$ . Numerically solving Eq. (13) for the toy model just discussed returns  $\Gamma_{\text{crit}} \simeq 0.77$ ,

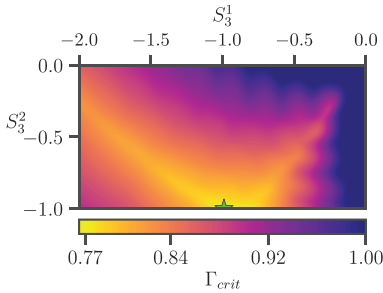


FIG. 5. Phase diagram in  $S$ . This can be numerically found by fixing  $\{\ell_e\}$  as in Fig. 2,  $S_3^1 = -1 - S_3^2$ ,  $S_1^2 = -2 - S_3^2$ , while varying  $S_3^1$  and  $S_3^2$ . The color bar denotes the value of  $\Gamma_{\text{crit}}$  above which optimal solutions can be loopy. The green star denotes the configuration of  $S$  used in Fig. 2.

as shown in Fig. 3 (green star). However, notice that the key point of this derivation is that we could find at least one choice of  $(\ell, F_a)$  for which we have loopy solutions in the nontrivial regime  $0 < \Gamma < 1$ . Indeed, at  $\Gamma = \Gamma^*$  we have  $J_\Gamma(\mathcal{G}_L; \Gamma) = J_\Gamma(\mathcal{T}_1) = J_\Gamma(\mathcal{T}_2) = J_\Gamma(\mathcal{T}_3) = 4$ .

The same procedure can be used to find the phase diagram of Fig. 5; here, the costs  $\{J(\mathcal{G}), J(\mathcal{T}_i)\}$  have been computed fixing the lengths as  $\ell_a = \ell_b = \frac{2}{3}\ell_c$ ,  $\ell_c = 1$ , and using  $(F_a, S_3^1, S_3^2)$  as independent variables.

#### APPENDIX D: TRIMMING LOOPS TO OBTAIN TREES

Any configuration of the edge fluxes  $\{F_e\}$  satisfying the colored Kirchhoff's law (1) can be associated with a weighted graph  $\mathcal{G}(\mathcal{V}, \mathcal{E}, W)$  with weights  $w_e = \|F_e\|_2$ , where  $\mathcal{V}$  and  $\mathcal{E}$  are the set of nodes and edges of the original input network. Denote with  $\mathcal{T}$  the optimal tree topologies among these weighted graphs, i.e., loopless topologies with weights minimizing  $J_\Gamma$  as defined in Eq. (10). These trees (not necessarily unique) can be obtained by taking a weighted graph  $\mathcal{G}_L$  with a single loop denoted as  $L$ , cutting the loop by trimming one of its edges, and then redistributing the fluxes passing through the trimmed edge over the remaining links of  $L$ . We assign an arbitrary orientation  $\hat{e}_i$  to the edges of  $L$  so that  $\langle \hat{e}_i, \hat{e} \rangle = \pm 1$ , where the direction  $\hat{e}$  of each link of a graph is uniquely determined by its incidence matrix. The edge to be cut is the one with smallest weight over the edges in the loop with a negative direction with respect to the graph's orientation. Its flux is redistributed over the remaining edges, which now make a tree. Formally, we assign to the edges of  $\mathcal{T}$  fluxes  $F_e^*$  such that their entries are

$$(F_e^i)^* = F_e^i + \langle \hat{e}_i, \hat{e} \rangle F_{\min}^i, \quad \forall e \in \mathcal{E}, \quad \forall i = 1, \dots, M, \quad (\text{D1})$$

where  $F_{\min} = (F_{\min}^1, \dots, F_{\min}^M) = \arg \min_e \{\ell_e \|F_e\|_2^2 : \langle \hat{e}_e, \hat{e} \rangle = -1\}$  and  $F_e$  are the fluxes of  $\mathcal{G}_L$ . The orientation of  $L$  can be switched in case the set of edges with negative orientation is empty. In the one-commodity case, as given by Xia [25], there is a similar trimming, but with scalar weights on the tree being  $F_e^* = F_e + \langle \hat{e}_i, \hat{e} \rangle F_{\min}$ , where now all the fluxes are

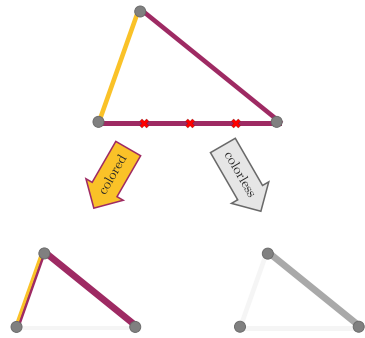


FIG. 6. Sketch of the trimming procedure described in Appendix D, on the toy model of Fig. 2. The tree obtained in the colored case is not optimal, while the one obtained in the colorless case has lower cost but violates Kirchhoff's law.

numbers. The key effect of having a scalar trimming is that  $F_e^*$  can become zero as a result of having a negative orientation  $\langle \hat{e}_i, \hat{e} \rangle$ ; in other words, the flux  $F_{\min}$  adds negatively to the fluxes originally present in  $\mathcal{G}_L$  along the edges in the loop with negative orientation: If  $F_{\min} = F_e$ , then  $F_e^* = 0$ . Here, instead, in Eq. (D1) we add a vector. While we might have that for certain components the flux cancels out, the norm of the whole vector  $F_e$  might not be zero, because not all the components (colors) cancel. This results from imposing a colored Kirchhoff's law.

We illustrate this procedure on the triangle network in Fig. 2. In particular, Fig. 6 shows how this trimming applies in the colored case (our case) against a colorless case. The tree obtained in the colored case is not optimal, while the one obtained in the colorless case has lower cost but is not valid, as it violates the constraints enforced by Kirchhoff's law. The consequence is that now loops can be the optimal solutions while in the one-commodity case optimal networks were trees. Specifically, there exists a  $\Gamma_{\text{crit}} \in (0, 1)$  [or  $\beta_{\text{crit}} \in (1, 2)$ ] such that we have a phase transition between trees and loopy structures. The value of  $\Gamma_{\text{crit}}$  depends on  $(S, \{\ell_e\})$ .

#### APPENDIX E: NUMERICAL IMPLEMENTATION

##### 1. Implementation details and gradient descent

We propose two approaches to solve our problem that are the natural multicommodity generalization of the approaches used in Refs. [2,3,5]. One is based on finding the steady state of the conductivities using Eq. (4) (Dynamics), and one is based on implementing the iterative update of Eqs. (11) and (12) (Optimization). The implementations of these methods are summarized in Algorithms 1 and 2.

These pseudocodes outline our methods; however, practitioners can make further arbitrary choices about what numerical routines to use in the various steps. In our implementation, we solved  $M$  ordinary differential equations as in Eq. (4) by means of an explicit Euler method; thus at each step the local truncation error is approximately proportional

## Algorithm 1. Dynamics.

---



---

```

1: Input:  $\mathcal{G}(\mathcal{V}, \mathcal{E})$  = adjacency list, nodes coordinates;  $M$ ;
   inflows;  $0 < \beta < 2$ 
2: Initialize: (i)  $S$  and (ii)  $\{\hat{\mu}_e\}$  [e.g., sampling as i.i.d.  $\hat{\mu}_e \sim U(0, 1)$ ]
3: while convergence not achieved do
4:   solve Kirchhoff's law as in Eq. (1)  $\rightarrow \{P_e^i\}$ 
5:   update conductivities with a finite difference
   discretization of Eq. (4):  $\{\hat{\mu}_e^i\} \rightarrow \{\hat{\mu}_e^{i+1}\}$ 
6: end while
7: Return: fluxes  $\{F_e^i\}$  at convergence, computed using
    $F_e^i = \hat{\mu}_e(p'_u - p'_v)/\ell_{e^i}$ ,  $e = (u, v)$ 

```

---



---

to  $\Delta t^2$ , with  $\Delta t$  being the difference between two consecutive time steps, which can be arbitrarily set in the input. Solutions of Kirchhoff's law have been computed using a sparse direct solver.

Lastly, we impose the following convergence criteria (*convergence* is achieved when these conditions are satisfied):

$$\text{Dynamics: } \max_e |\hat{\mu}_e^{t+1} - \hat{\mu}_e^t| / \Delta t < \tau_{\text{dyn}}, \quad (\text{E1})$$

$$\text{Optimization: } \max_e \left( \|F_e\|_2^{t+1} - \|F_e\|_2^t \right) < \tau_{\text{opt}}, \quad (\text{E2})$$

where  $\tau_{\text{dyn}}, \tau_{\text{opt}} > 0$  are parameters arbitrarily set in the input. In our experiments we use  $\tau_{\text{dyn}} = 10^{-3}$ ,  $\tau_{\text{opt}} = 10^{-5}$ . To test our methods, we developed a momentum-based gradient descent as a baseline algorithm. This consists in the component-wise iterative update of the fluxes using

$$(V_e^i)^t = \eta (\partial J_\beta / \partial F_e^i)^t + \delta (V_e^i)^{t-1}, \quad (\text{E3})$$

$$(F_e^i)^{t+1} = (F_e^i)^t + (V_e^i)^t, \quad (\text{E4})$$

with  $\eta, \delta > 0$  fixed increment rates and  $(V_e^i)^0 = F_e^i$ . We fixed the convergence criteria analogously to what done for the other two methods:  $\max_e \left( \|F_e\|_2^{t+1} - \|F_e\|_2^t \right) / \eta < \tau_{\text{gd}}$ , with  $\tau_{\text{gd}} > 0$  being a parameter that needs to be set in the input. In our experiments we set it to  $\tau_{\text{gd}} = 10^{-2}$ . From a theoretical point of view, the comparison with a standard gradient descent method was proposed in light of the equivalence of our dynamics and a mirror-descent approach for the Lyapunov functional, as proved for  $\beta = 1$  in Ref. [28]. The dynamics automatically preserves positiveness of the conductivities  $\{\hat{\mu}_e\}$ ,

## Algorithm 2. Optimization.

---



---

```

1: Input:  $\mathcal{G}(\mathcal{V}, \mathcal{E})$  = adjacency list, nodes coordinates;  $M$ ;
   inflows;  $0 < \beta < 2$ 
2: Initialize: (i)  $S$  and (ii)  $\{\hat{\mu}_e\}$  [e.g., sampling as i.i.d.  $\hat{\mu}_e \sim U(0, 1)$ ]
3: while convergence not achieved do
4:   solve Kirchhoff's law as in Eq. (1)  $\rightarrow \{P_e^i\}$ 
5:   update fluxes using Eq. (12)
6:   compute  $\hat{\mu}_e(F_e)$  using Eq. (11)
7: end while
8: Return: fluxes  $\{F_e^i\}$  at convergence

```

---



---

and thus a large time step can be used. In contrast, using purely gradient descent approaches, the time step size must be reduced when some entries of the vector  $\{\hat{\mu}_e\}$  go to zero. After running our algorithms until convergence, the original network is trimmed by removing edges with negligible fluxes. Formally, we remove links for which  $\|F_e\|_2 < \tau$ , with  $\tau > 0$  arbitrarily fixed. Typically, as we empirically found, the distribution of  $\|F_e\|_2$  over the edges is divided into two sets having values differing by several orders of magnitude. It is thus straightforward to distinguish which edges are to be trimmed, i.e., those that have negligible values compared with the rest of the distribution.

## 2. Computational complexity

Each temporal step executed in our algorithms requires the approximate solution of  $M$  linear systems of dimension  $N$ . This operation has been carried out by means of a sparse direct solver (UMFPACK) that performs a LU decomposition for each column of the right-hand side of Eq. (1). The total computational complexity of this process scales as  $O(MN^2)$ . To have a better understanding of this, we tested our models with several synthetic Waxman networks obtained by placing  $N$  nodes uniformly at random in a square domain of size 1. Nodes are connected with probability  $p = A \exp(-d/\alpha L)$ , where  $A, \alpha, L$  are parameters that we fix arbitrarily to  $A = 1/4$ ,  $\alpha = 1/4$ , and  $L = 1$ ;  $d$  is the Euclidean distance between a pair of nodes. The matrix  $S$  is constructed assigning a total inflowing mass of  $10^4$  at random to  $M$  nodes and redistributing on the nodes of the network proportionally to their inflows.

We test the efficiency of our schemes by measuring the total running time (in seconds) to reach convergence for different values of  $\beta, M, N$ . Results are shown in Fig. 7. We notice that the two algorithms, Dynamics and Optimization, have similar computational complexity. Their small running-time differences are negligible and only due to how convergence is precisely defined, i.e., how the corresponding parameters  $\tau_{\text{dyn}}$  and  $\tau_{\text{opt}}$  are set. The running time is shorter for  $\beta < 1$  (traffic optimization, loopy) than in the opposite scenario of  $\beta > 1$  (minimization of infrastructural cost, treelike). The case  $\beta = 1$  is more nuanced, as the cost transitions between two opposite situations. In this case, Optimization fails to converge for  $M/N < 1$ , if convergence is defined in terms of variations of  $\|F_e\|_2$  between iteration steps. This is because the algorithm gets lost in degenerate local minima, configurations with the same cost but different sets of fluxes. This lack of convergence suggests that, for  $\beta = 1$ , the energy landscape around these minima is flat, i.e., there are many configurations with the same cost but non-negligible differences in their fluxes. The Optimization routine keeps switching between these different states. In this case, one can simply pick one of these many possible solutions as an example local optimum. The dynamics does instead converge. This suggests that Dynamics is biased towards one of these degenerate solutions. For  $M/N = 1$ , Optimization converges with the same running time as Dynamics, suggesting that as we enlarge  $M$ , the landscape becomes less flat. A possible cause is that by increasing  $M$  the system has more constraints to be satisfied via Kirchhoff's law, which reduces the number of possible degenerate solutions. This claim is also supported by the

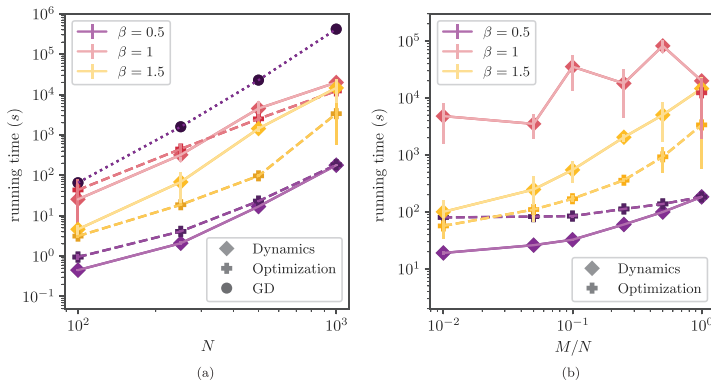


FIG. 7. Computational complexity. (a) Running time (in seconds) as a function of system size  $N$ . (b) Running time as a function of the ratio  $M/N$  between the number of commodities and system size. GD denotes gradient descent, implemented with Eqs. (E3) and (E4); we only show it for  $\beta = 0.5$  as for the other values it fails to converge within a reasonable time. Similarly, for  $\beta = 1$  and  $M/N < 1$ , Optimization fails to converge, and hence we only report Dynamics.

behavior of Dynamics' running time, which does not monotonically increase with  $M/N$  in this case, as shown in Fig. 7(b). These behaviors highlight relevant differences between the

two implementations. Finally, we note that the computational complexity could in principle be further reduced to  $O(MN)$  using multigrid methods [29]; we do not explore this here.

- [1] J. R. Banavar, F. Colaiori, A. Flammini, A. Maritan, and A. Rinaldo, Topology of the Fittest Transportation Network, *Phys. Rev. Lett.* **84**, 4745 (2000).
- [2] S. Bohn and M. O. Magnasco, Structure, Scaling, and Phase Transition in the Optimal Transport Network, *Phys. Rev. Lett.* **98**, 088702 (2007).
- [3] F. Corson, Fluctuations and Redundancy in Optimal Transport Networks, *Phys. Rev. Lett.* **104**, 048703 (2010).
- [4] K. Sinclair and R. C. Ball, Mechanism for Global Optimization of River Networks from Local Erosion Rules, *Phys. Rev. Lett.* **76**, 3360 (1996).
- [5] D. Hu and D. Cai, Adaptation and Optimization of Biological Transport Networks, *Phys. Rev. Lett.* **111**, 138701 (2013).
- [6] H. Ronellenfitch and E. Katifori, Global Optimization, Local Adaptation, and the Role of Growth in Distribution Networks, *Phys. Rev. Lett.* **117**, 138301 (2016).
- [7] E. Katifori, G. J. Szöllösi, and M. O. Magnasco, Damage and Fluctuations Induce Loops in Optimal Transport Networks, *Phys. Rev. Lett.* **104**, 048704 (2010).
- [8] H. Ronellenfitch and E. Katifori, Phenotypes of Vascular Flow Networks, *Phys. Rev. Lett.* **123**, 248101 (2019).
- [9] F. Kaiser, H. Ronellenfitch, and D. Witthaut, Discontinuous transition to loop formation in optimal supply networks, *Nat. Commun.* **11**, 5796 (2020).
- [10] J. B. Kirkegaard and K. Sneppen, Optimal Transport Flows for Distributed Production Networks, *Phys. Rev. Lett.* **124**, 208101 (2020).
- [11] V. Bonifaci, K. Mehlhorn, and G. Varma, *Physarum* can compute shortest paths, *J. Theor. Biol.* **309**, 121 (2012).
- [12] F. Santambrogio, Optimal channel networks, landscape function and branched transport, *Interfaces Free Boundaries* **9**, 149 (2007).
- [13] E. Facca, F. Cardin, and M. Putti, Towards a stationary Monge-Kantorovich dynamics: The *Physarum polycephalum* experience, *SIAM J. Appl. Math.* **78**, 651 (2016).
- [14] E. Facca, S. Daneri, F. Cardin, and M. Putti, Numerical solution of Monge-Kantorovich equations via a dynamic formulation, *J. Sci. Comput.* **82**, 68 (2020).
- [15] E. Facca, F. Cardin, and M. Putti, Branching structures emerging from a continuous optimal transport model, *J. Comput. Phys.* **447**, 110700 (2021).
- [16] D. Baptista, D. Leite, E. Facca, M. Putti, and C. De Bacco, Network extraction by routing optimization, *Sci. Rep.* **10**, 088702 (2020).
- [17] V. Bonifaci, E. Facca, F. Folz, A. Karrenbauer, P. Kolev, K. Mehlhorn, G. Morigi, G. Shahkarami, and Q. Vermande, *Physarum* multi-commodity flow dynamics, [arXiv:2009.01498](https://arxiv.org/abs/2009.01498).
- [18] K. Salimifard and S. Bigharaz, The multicommodity network flow problem: State of the art classification, applications, and solution methods, *Oper. Res. Int. J.* (2020).
- [19] C. H. Yeung, D. Saad, and K. Y. M. Wong, From the physics of interacting polymers to optimizing routes on the London Underground, *Proc. Nat. Acad. Sci.* **110**, 13717 (2013).
- [20] C. H. Yeung and D. Saad, Networking - A statistical physics perspective, *J. Phys. A: Math. Theor.* **46**, 103001 (2011).
- [21] T. Nelson and N. Dengler, Leaf vascular pattern formation, *Plant Cell* **9**, 1121 (1997).
- [22] C. B. Schaffer, B. Friedman, N. Nishimura, L. F. Schroeder, P. S. Tsai, F. F. Ebner, P. D. Lyden, and D. Kleinfeld,

- Two-photon imaging of cortical surface microvessels reveals a robust redistribution in blood flow after vascular occlusion, *PLoS Biol.* **4**, e22 (2006).
- [23] A. Nardini, G. Pedà, and N. L. Rocca, Trade-offs between leaf hydraulic capacity and drought vulnerability: Morpho-anatomical bases, carbon costs and ecological consequences, *New Phytol.* **196**, 788 (2012).
- [24] J. R. Banavar, A. Maritan, and A. Rinaldo, Size and form in efficient transportation networks, *Nature (London)* **399**, 130 (1999).
- [25] Q. Xia, Optimal paths related to transport problems, *Commun. Contemp. Math.* **5**, 251 (2003).
- [26] A. A. Ibrahim, A. Lonardi, and C. De Bacco, Optimal transport in multilayer networks for traffic flow optimization, *Algorithms* **14**, 189 (2021).
- [27] <https://github.com/aleable/McOpt>.
- [28] V. Bonifaci, A Laplacian approach to  $\ell_1$ -norm minimization, *Comput. Optim. Appl.*, **79**, 441 (2021).
- [29] W. L. Briggs, V. E. Henson, and S. F. McCormick, *A Multigrid Tutorial* (SIAM, Philadelphia, 2000).



# OPEN Multicommodity routing optimization for engineering networks

Alessandro Lonardi<sup>1</sup>, Mario Putti<sup>2</sup> & Caterina De Bacco<sup>1</sup>


Optimizing passengers routes is crucial to design efficient transportation networks. Recent results show that optimal transport provides an efficient alternative to standard optimization methods. However, it is not yet clear if this formalism has empirical validity on engineering networks. We address this issue by considering different response functions—quantities determining the interaction between passengers—in the dynamics implementing the optimal transport formulation. Particularly, we couple passengers' fluxes by taking their sum or the sum of their squares. The first choice naturally reflects edges occupancy in transportation networks, however the second guarantees convergence to an optimal configuration of flows. Both modeling choices are applied to the Paris metro. We measure the extent of traffic bottlenecks and infrastructure resilience to node removal, showing that the two settings are equivalent in the congested transport regime, but different in the branched one. In the latter, the two formulations differ on how fluxes are distributed, with one function favoring routes consolidation, thus potentially being prone to generate traffic overload. Additionally, we compare our method to Dijkstra's algorithm to show its capacity to efficiently recover shortest-path-like graphs. Finally, we observe that optimal transport networks lie in the Pareto front drawn by the energy dissipated by passengers, and the cost to build the infrastructure.

Finding optimal flow configurations in transport networks is an important problem in many real-world applications. While natural systems like river basins<sup>1–5</sup>, leaf venations<sup>6–9</sup>, or slime molds<sup>10–17</sup> involve transport of one type of mass only, e.g. water, this may not be the case in several engineering systems. For instance, routing data packets in communication networks, or passengers in urban transportation networks, requires multicommodity approaches where mass of different types interacts in a shared infrastructure, contributing to minimize one unique cost.

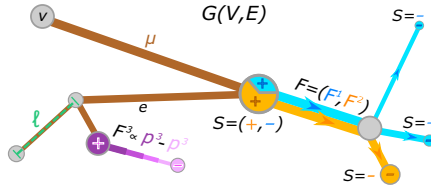
Despite their practical significance, multicommodity algorithms based on optimization routines are burdened by high computational complexity, caused by the simultaneous assignment of multiple commodities. Therefore, practitioners often rely on heuristics and approximations that lead to suboptimal solutions<sup>18</sup>. Distributed approaches like message-passing algorithms have demonstrated encouraging results<sup>19–24</sup>, but remain computationally costly in scenarios where there is a large number of origin-destination pairs to be routed, or when the network is not sparse.

A promising approach is that of optimal transport theory. Recent studies<sup>25,26</sup> have shown that this theoretical formalism can be adapted to address multicommodity scenarios, generalizing well-established results for unicommodity models<sup>27–33</sup>. The works of Lonardi *et al.*<sup>25</sup> and Bonifaci *et al.*<sup>26</sup> focus on a theoretical characterization of the problem, drawing a formal connection between optimal transport and an equivalent dynamical system that is formulated in terms of physical quantities like conductivities and fluxes. While preliminary results on multilayer transportation networks<sup>34</sup> suggest an empirical validity of this choice, questions remain open about its applicability in settings involving the transport of passengers.

In this work, we address this concern by studying the behavior of optimal transport approaches for multicommodity routing on urban transportation networks, with an empirical analysis on the Paris metro network. Our goal is to evaluate how different cost functions impact the distribution of passenger flows. In detail, we search for stationary solutions of a dynamics where edge capacities—conductivities—grow as an increasing function of the total amount of passengers traveling on the edges. We investigate numerically the cases where the dependence between conductivities and fluxes is either the sum of the passengers traveling on an edge (its 1-norm), or the sum of their squares (its 2-norm). The first choice is more intuitive, since counting the total number of users in

<sup>1</sup>Max Planck Institute for Intelligent Systems, Cyber Valley, Tübingen 72076, Germany. <sup>2</sup>Department of Mathematics "Tullio Levi-Civita", University of Padua, Via Trieste 63, Padua, Italy. email: alessandro.lonardi@tuebingen.mpg.de





**Figure 1.** Schematic problem visualization. In brown we draw the edge capacities, green is used to highlight the length of one edge. From the central node a positive inflow of two commodities enters (orange and light blue), these move towards their destinations—the colored minuses—sharing an edge. Thus, multiple colored fluxes generate traffic congestion. In pink we represent the pressure potentials of a third commodity. Differences of pressure along an edge trigger  $F^3$ .

a network is a natural metric to evaluate its occupancy. However, in the second case it is possible to prove that the companion gradient flow used in the numerical solver admits a unique stationary solution<sup>25,26</sup>.

We design several experiments to investigate the main properties of optimal network configurations in the two cases. First, we observe that the 2-norm tends to dilute more substantially passengers on the network, avoiding heavily trafficked routes. Second, we compare our model with Dijkstra’s algorithm<sup>35</sup>, a popular approach for shortest-path minimization. We find that our method is a robust and efficient alternative to reproduce shortest-path-like networks. Furthermore, we test resilience to infrastructural failures, i.e., node and edge removal. Results show that the geographical locations of stations together with their degree, are decisive factors. Finally, we observe that optimal networks lie in the Pareto front drawn by two fundamental driving forces: the energy dissipated by passengers’ flows and the network infrastructural cost.

**Results and discussion**

**Multicommodity routing on networks.** We design a routing optimization problem on a network  $G(V, E)$ , where  $V$  and  $E$  are the sets of nodes and edges, and each edge has length  $\ell_e > 0$ . The edges are given a conventional orientation stored in a signed incidence matrix, with elements  $B_{ve} = \{+1, -1, 0\}$  if  $v$  is the head, the tail, or neither of them for edge  $e$ , respectively. We model transportation of  $M \geq 1$  commodities through the network, each identified by an index  $i$ . We use them to differentiate passengers entering the network from different stations ( $i \in V$ ), so that multiple users sharing the same path catalyze traffic congestion. Suppose that a commodity  $i$  has a mass rate  $S_v^i$  flowing into node  $v$  and outflows  $S_u^i \forall u \neq v$ , with  $\sum_v S_v^i = 0, \forall i \in V$ , to ensure that the system is isolated.

The main quantities of interest are the edge conductivities  $\mu_e \geq 0$ , which can be thought of as capacities. These regulate how passengers flow on the network, as higher conductivity is allocated to edges that are more utilized, while low-conductivity edges are those with fewer passengers. Hence, determining the values of  $\mu_e, \forall e \in E$ , implies determining the flows of passengers, and therefore of traffic on the network. The distribution of conductivities is regulated by the following dynamics and main equations of our model:

$$\sum_u L_{vu} p_u^i = S_v^i \quad \forall v \in V, \forall i = 1, \dots, M \tag{1}$$

$$\frac{d\mu_e}{dt} = \mu_e^{\beta-2} f(F_e) - \mu_e \quad \forall e \in E. \tag{2}$$

Here  $L$  is the weighted Laplacian matrix of the network, with entries  $L_{vu} := \sum_e (\mu_e / \ell_e) B_{ue} B_{ve}$ ;  $p_u^i$  are pressure potentials generated by a commodity  $i$  on the nodes;  $f$  is a non-negative function of the fluxes  $F_e$ ,  $M$ -dimensional vectors with entries  $F_e^i := \mu_e (p_u^i - p_v^i) / \ell_e$ , for  $e = (u, v)$ . A visualization of the main model’s variables is shown in Fig. 1.

Equation (1) is Kirchhoff’s law, expressing conservation of mass; Eq. (2) regulates the time evolution of conductivities by means of a feedback mechanism where the higher the flux on an edge, the larger its conductivity  $\mu_e$ . All commodities share one unique infrastructure, so we follow<sup>25</sup> and assume that  $\mu_e$  is the same for all  $i$ . This particular modeling choice corresponds to not prioritize any commodity in particular, i.e. having all users sharing the metro infrastructure without any hierarchy. However, one could consider imposing a set of rules for traffic regulation by explicitly accounting for different  $\mu_e^i$  terms.

The growth of  $\mu_e$  is governed by the function  $f$ , that is typically an increasing and differentiable function of some norm of the fluxes<sup>25,26</sup>. The aim of our work is to investigate how different expressions of  $f$  result in different distributions of passengers flows, thus we focus on the following two choices: (i)  $f(x) = \|x\|_2^2$  (2-norm), and (ii)  $f(x) = \|x\|_1$  (1-norm). The first captures intuition in contexts as plant biology, where nutrients travel independently in conduits which are held together in fibers, contributing to growth of branches. However, it may not be the most appropriate one in applications involving transport of passengers, as the 2-norm does not have a straightforward interpretation. On the contrary, the latter is arguably a more natural choice, backed up by the

intuition that edge capacities are controlled by the *number* of passengers traveling on them (instead of the sum of squares). Both norms are taken squared, this is motivated by an analogy between our dynamics and Joule's law in electrical circuits, that we discuss in "Connection with optimal transport" section. We remark that other possible choices of  $f$  can be used, e.g. the complete spectrum of  $p$ -norms, or a tunable sigmoid profile as in<sup>13</sup>, these can be interesting subjects for future work. The effect of  $f$  is balanced by a negative linear term in the conductivities, determining their exponential decay in time if no mass is moving through an edge. Note that our dynamics is highly non-linear in  $\mu_e$ , since least-square solutions of Kirchhoff's law are of the form  $p_v^i = \sum_u \Gamma_{vu}^i S_u^i$ , with  $\Gamma$  denoting the Moore-Penrose inverse. Finally, the role of the free parameter  $0 < \beta < 2$  is to capture different transportation mechanisms:  $\beta > 1$  consolidates passengers on fewer edges, following a principle of economy of scale;  $\beta < 1$  enforces passengers to distribute more broadly along the network;  $\beta = 1$  is shortest-path like.

**Connection with optimal transport.** The dynamics introduced in "Multicommodity routing on networks" section has a strong connection with optimal transport theory. In fact, in<sup>25</sup> it is shown that *stationary solutions* of Eqs. (1) and (2) are also stationary points of the minimization problem:

$$\min_{F \in \mathbb{R}^{|E| \times M}} J := \frac{1}{2} \sum_e \frac{\ell_e}{\mu_e} f(F_e) \tag{3}$$

$$\text{s.t. } \sum_e \ell_e \mu_e^{2-\beta} = K^{2-\beta} \tag{4}$$

$$\sum_e B_{ve} F_e^i = S_v^i \quad \forall v \in V, \forall i = 1, \dots, M, \tag{5}$$

for a fixed constant  $K > 0$  and where  $J$  is the dissipation cost. This is also equivalent to minimizing  $J_\Gamma := \sum_e \ell_e f_\Gamma(F_e)^\Gamma$ , with  $\Gamma = (2 - \beta)/(3 - \beta)$ , generalizing Banavar *et al.*<sup>26</sup>.

The crucial distinction between the 1-norm and 2-norm dynamics is that the latter admits the Lyapunov function

$$\mathcal{L}_\beta(\{\mu_e\}) := \frac{1}{2} \sum_{i,v} p_v^i S_v^i + \frac{1}{2(2-\beta)} \sum_e \ell_e \mu_e^{2-\beta}, \tag{6}$$

which enables to prove that *asymptotics* of the dynamics minimize  $J^{25}$  for  $\beta \leq 1$ . We remark that for  $\beta \leq 1$  the Lyapunov admits a unique minimum although possibly multiple minimizers, while for  $\beta > 1$  the functional has several local minima. Noticeably, the first sum in Eq. (6) is equivalent to  $J = (1/2) \sum_e \ell_e \|F_e\|_2^2 / \mu_e$  (see Methods "Lyapunov and dissipation cost equivalence").

The second term in Eq. (6) is  $W := (\sum_e \ell_e \mu_e^\gamma) / 2\gamma$  with  $\gamma = 2 - \beta$ , interpretable as the cost to build the network. With this in mind, the Lyapunov functional becomes the sum of dissipation and infrastructural costs.

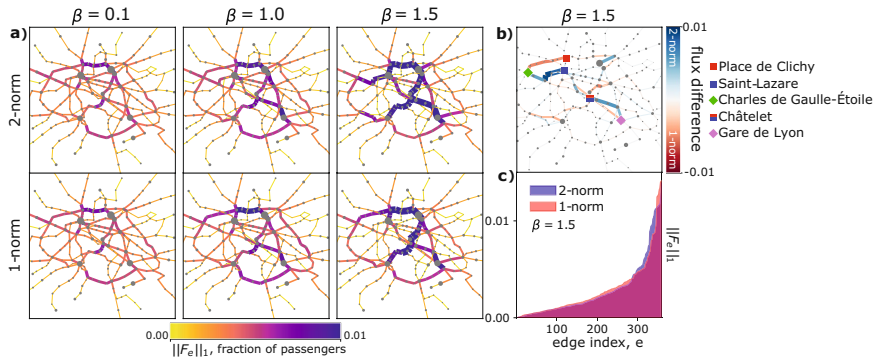
As mentioned before theoretical guarantees cannot be recovered for the 1-norm dynamics, where a Lyapunov functional is not straightforward to derive. While solving the dynamics may still result in meaningful flows, we cannot guarantee that these solutions minimize the cost  $J_\Gamma$ , i.e. to have optimal transport.

However, we find empirically that on the metro network of Paris—our case of study— $J_\Gamma$  decreases along solution trajectories of the dynamics, with stationary solutions lying in a basin of the cost. This empirical result is valid here, but this may not be true for other configurations of the network or initial conditions of the dynamics, hence practitioners should first validate their model (see Methods "Preprocessing", "Validation" for a more detailed explanation; a listing of the variables introduced in "Multicommodity routing on networks", "Connection with optimal transport" sections can be found enclosed as Supplementary Information).

**Results on the Paris metro network.** In this work, we investigate the applicability of the dynamics in Eqs. (1) and (2) on the Paris metro. Topology data are taken from<sup>37</sup>, the network is preprocessed to have a total of  $|V| = 302$  nodes and  $|E| = 359$  edges, coherently with the observed metro of Paris (Methods "Preprocessing"). As anticipated, we define commodities as stations where passengers enter. This means that each vector  $S^i$  has only one positive element in  $v = i$  (where the passengers of type  $i$  enter), while the remaining elements of  $S^i$  contain the outflows of passengers who travel from  $v$ . Other choices can also be made based on the application, but this will not impact the validity of the model. Lastly, we introduce the parameter  $0 \leq \rho \leq 1$ . This averages the passenger inflows as  $S_{v=i}^i(\rho) = S_{v=i}^i - \rho(S_{v=i}^i - \langle S^i \rangle)$ , with  $\langle \cdot \rangle$  average over the nodes. When  $\rho$  tends to 1 passengers distribute uniformly on the network, while  $\rho$  approaching 0 means passengers enter and exit station more heterogeneously, see Methods "Validation".

We test the two response functions  $f$ . Optimal fluxes resulting in the two cases can be seen in Fig. 2a, where the thickness of each edge is proportional to the fraction of passengers traveling through it. As expected, for  $\beta < 1$  optimal transport networks are loopy, with many densely connected edges having fairly uniform fluxes. On the contrary, for  $\beta > 1$  optimal topologies are more tree-like, with few central arteries where traffic is highly concentrated. This applies to both cases.

We notice two distinct behaviors, depending on  $\beta$ . For  $\beta < 1$  ( $\beta = 0.1$  in Fig. 2a), solutions cannot be distinguished. This is explained by the Lyapunov functionals  $\mathcal{L}_\beta$  being strictly convex in this case, with stationary conductivities that are their only minimum. This observation suggests that in the congested transportation regime ( $0 < \beta < 1$ ), where one aims at minimizing traffic congestions, using the 2-norm is equivalent to the



**Figure 2.** Optimal transport networks panel. **(a)** Optimal transport networks with  $\beta = 0.1, 1.0, 1.5$  for the 1-norm and 2-norm. Edge thickness and color are proportional to  $\|F_e\|_1$ , normalized to sum to 1; node sizes are proportional to the number of passengers entering them. All the quantities are averaged over 100 runs of the dynamics with  $\mu_e(0) \sim U(0, 1)$ . **(b)** Network colored using the difference of the fluxes obtained with the 1-norm and with the 2-norm. Results are displayed for  $\beta = 1.5$ , and using the data of **(a)**. Widths of edges are proportional to the absolute value of the flux difference, so that by matching the color and size information it is possible to distinguish differences in networks generated by the two response functions. Marked stations are those discussed in “Results on the Paris metro network” section. **(c)** Sorted flux distribution over the edges for  $\beta = 1.5$ . All quantities have been computed with  $\rho = 0.0$ , i.e.  $S(\rho = 0.0) = S$  (see Methods “Preprocessing”, “Validation”). Similar panels for  $\rho = 0.5$  and  $\rho = 1.0$  can be found in Supplementary Figs. 1 and 2.

more intuitive 1-norm formulation. This is not the case for  $\beta > 1$ , where the two dynamics favor different local minima. These correspond to optimal networks with distinct central arteries where passengers are directed. The differences are further accentuated in Fig. 2b, where edges are colored with flux differences in these two cases, and where we highlight with markers instances of highly traversed stations. In detail, we can see that two routes branch from Charles de Gaulle-Étoile, the upper one passing by Place de Clichy is favored by the 1-norm, and the lower one reaching Saint-Lazare is preferred by the 2-norm. As for the connection between Châtelet and Gare de Lyon, we observe that the 1-norm tends to favor the shortest path between the two stations, with most passengers travelling in a straight line. On the contrary, the path selected by the 2-norm has a deflection.

Stimulated by these qualitative differences, we investigate different metrics for an in-depth quantitative evaluation for the case  $\beta = 1.5$ . First, analyzing the sorted distributions of the fluxes  $\|F_e\|_1$  in Fig. 2c, we notice that the 1-norm dynamics has a more pronounced fat-tailed distribution with a sharper and higher peak. This means that the 1-norm tends to concentrate fluxes on fewer edges. Such effect becomes starker for more homogeneous distributions of passengers entering the stations, i.e. setting  $\rho = 0.5, 1.0$  (see Supplementary Figs. 1 and 2).

We quantify this using the Gini coefficient<sup>38</sup>:

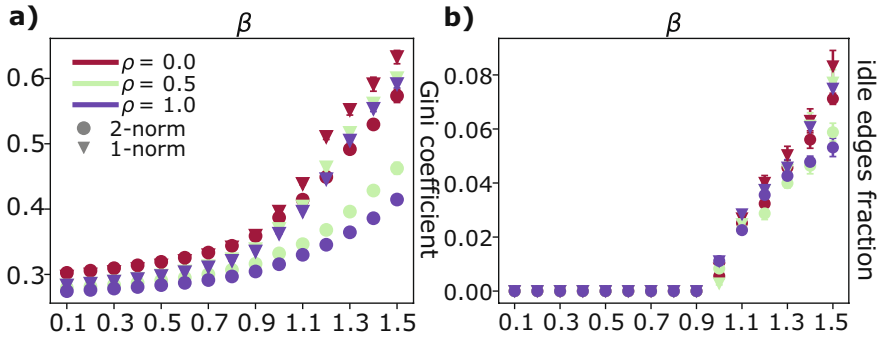
$$\text{Gini coefficient}(x) := \frac{1}{2|E|^2 \bar{x}} \sum_{m,n} |x_m - x_n| \tag{7}$$

for a quantity  $x$ , with  $\bar{x} = \sum_e x_e / |E|$  being its mean, and  $m, n$  denoting edges. In our analysis we set  $x_e = \|F_e\|_1$ . Results are shown in Fig. 3a, where the Gini coefficient is plotted against  $\beta$  for different values of  $\rho$ .

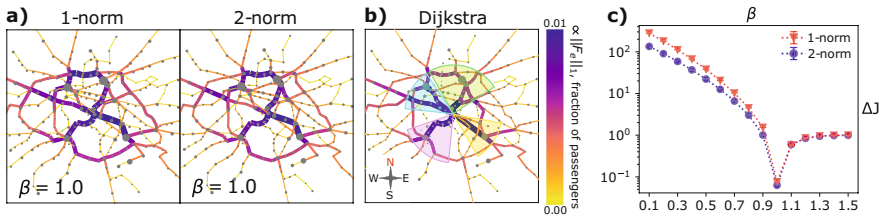
As expected, the Gini coefficient increases with  $\beta$ , as users’ paths are more concentrated along fewer edges. The values for the two dynamics are similar for  $\beta < 1$ , for the reasons mentioned above. Instead, for  $\beta > 1$ , markers progressively separate as  $\beta$  increases. The 2-norm has always smaller values than their counterparts, further demonstrating the tendency of the 2-norm to dilute fluxes on a larger area of the network.

We study the behavior of the fraction of idle edges, i.e. the number of edges with negligible fluxes, divided by the total number of edges  $|E|$ , see Fig. 3b. This quantity manifests a sudden phase transition at  $\beta = 1$ , where the dynamics switches from an homogeneous distribution of passengers on the entire network infrastructure, to a distribution progressively more concentrated on a smaller fraction of edges, as  $\beta$  increases. Finally, the 2-norm dynamics returns fewer idle edges than the 1-norm, as paths are less concentrated. Notably, such abrupt phase transitions are typical of capacitated models on networks<sup>36</sup>, and emerging in routing strategies involving a critical exponent regulating efficient transportation<sup>39</sup>.

To summarize, we observe two main findings. First, we noticed that in the regime of  $\beta < 1$  the 1-norm and the 2-norm produce identical optimal networks. This result does not hold for  $\beta > 1$ , where many local minima of  $\mathcal{L}_\beta$  generate different optimal paths. Second, analyzing the fraction of idle edges, the Gini coefficient of the fluxes, and their distribution, we found that in the regime of branched transportation ( $1 < \beta < 2$ ), the 2-norm tends to limit more traffic congestion, as paths are less consolidated into fewer edges compared to the 1-norm.



**Figure 3.** Evaluation metrics for the optimal transport networks. (a) Gini coefficient vs.  $\beta$ . (b) Idle edges fraction vs.  $\beta$ . Each point is averaged over 100 runs of the dynamics with random initializations of the conductivities.

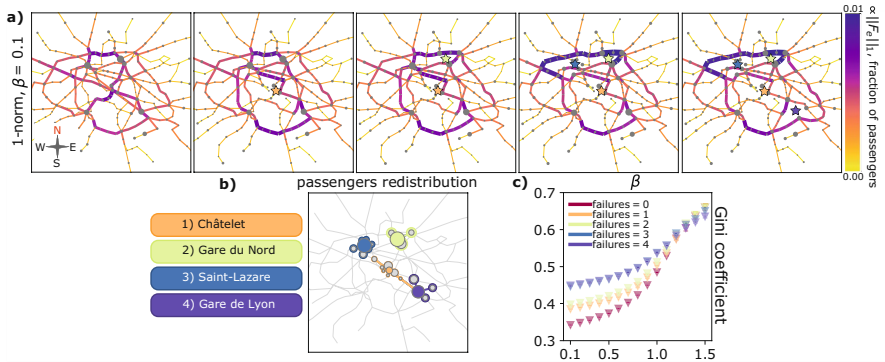


**Figure 4.** Comparison between our methods and Dijkstra's algorithm. Lengths of edges are  $\tilde{\ell}_e = \ell_e/n_e$ , with  $n_e$  number of vehicles passing through  $e$  as in<sup>37</sup>. This rescaling has been performed following the intuition that metro users moving along the network may travel using the fastest route (that for paths of the same length is the one with more frequent trains) to reach their destination; this may not correspond to the geographically shortest one. (a) Optimal transport networks for our methods. Quantities are computed over 100 runs of the dynamics with random initialization of the conductivities,  $\mu_e \sim U(0, 1)$ . (b) Optimal transport network computed with Dijkstra's algorithm. For all three networks, edge widths and colors are  $\|F_e\|_1$ , and the size of each node is proportional to the number of passengers entering in it. (c) Relative energy difference between our methods and Dijkstra's, taken in absolute value. Errorbars are standard deviations over 100 realizations.

**Comparison with Dijkstra algorithm.** As discussed, the main property connecting our 2-norm dynamics with optimal transport is that its stationary solutions are minimizers of the cost  $J_\Gamma = \sum_e \ell_e \|F_e\|_2^\Gamma$ , with  $\Gamma = (2 - \beta)/(3 - \beta)^{35}$ . This cost, for  $\beta = 1$  and  $M = 1$  is equivalent to that of<sup>5,16</sup> and has optimal fluxes taking the shortest path from their source to their sink. A theoretical generalization of this result to the multicommodity setup is not trivial. In fact, for the 2-norm case the cost reads  $J_\Gamma = \sum_e \ell_e \sqrt{\sum_i (F_i^e)^2}$ , that is not linear in the commodities, i.e. searching for its minimizer does not correspond to solving  $M$  unicommodity problems, one for each  $i$ , and then overlapping them. As for the 1-norm, the dissipation cost with  $\beta = 1$  is  $J_\Gamma = \sum_e \sum_i \ell_e |F_i^e|$ , and therefore its unique global minimum corresponds to that obtained overlapping  $M$  shortest paths.

We can numerically compare our methods with a shortest path routine using Dijkstra's algorithm<sup>35</sup>. Precisely, we iterate over the commodities and assign a flux  $F_i^e$  equal to the fraction of passengers moving from the source  $v$  to the sink  $u$ , to each edge belonging to the shortest path between  $v$  and  $u$ —the latter computed with Dijkstra's algorithm.

We compare the optimal transport networks obtained using our methods with  $\beta = 1$  (Fig. 4a) with the networks returned by Dijkstra's algorithm (Fig. 4b). The three graphs are visibly similar but not identical. Particularly, we focus on the four highlighted areas in Fig. 4b, containing the main branches departing from the central area of the city of Paris. We see that the more trafficked routes in the pink South-West region are identical for our methods and for Dijkstra's one. Traffic in the North-West blue region seems to be more diluted for our methods, with the 2-norm optimal network being slightly more similar to Dijkstra's. As for the North green region, both our algorithms concentrate traffic in a curved branch covering a large portion of the Northside of the city. This route is not prioritized in Fig. 4b, as traffic in the green portion is more distributed. Finally, in the South-East



**Figure 5.** Traffic rerouting after network structural failures. (a) We plot the optimal transport networks after nodes trimming. Edge widths and color are normalized fluxes, the size of each node is proportional to the number passengers entering in it. All quantities are averaged over 100 runs of the dynamics with random initialization of the conductivities,  $\mu_e \sim U(0, 1)$ . Stars highlight positions of removed stations, following the same scheme of (b). (b) Network showing which stations have been removed, these correspond to the fully colored nodes, with colors chosen according to the legend on the left. Colored edges, and nodes with colored borders are those where the passengers get redirected. The colored borders are proportional to the passengers' growth. (c) Gini coefficients vs.  $\beta$ , errorbars are standard deviations. Points are colored following the scheme used in the rest of the panel. Similar results for the 2-norm dynamics are in Supplementary Fig. 3.

yellow area, there is only one main route branching from the city center, while its shape is straight for Dijkstra's, our methods favor a slight deflection.

We attribute these differences in the optimal topologies to the high complexity of the energy landscape of  $J_\Gamma$ . In fact, while Dijkstra's algorithm computes and overlaps each source-sink shortest path separately, our methods treat all the commodities at once. This may lead to convergence in suboptimal points, in particular around  $\beta = 1$ , where the cost transitions from being strictly convex ( $\beta \leq 1$ ) to strongly non-convex ( $\beta > 1$ ). While our method in this case may not always reach an optimal solution, it has the practical advantage of having a worst-case complexity of  $O(M|V|^2)^{35}$ . In principle, this can be further reduced using a backward Euler scheme combined with the inexact Newton-Raphson method for the discretization of Eq. (2), and using a Multigrid solver for the solution of Eq. (1) in  $O(M|V|)$  steps<sup>40</sup>. By contrast, Dijkstra's has a worst-case complexity of  $O(|E| + |V| \log |V|)$ <sup>41</sup>, with the algorithms that needs to be executed  $M^2$  (in our application to the Paris metro there are  $M^2 = |V|^2$  source-destination pairs) times to solve a multicommodity problem.

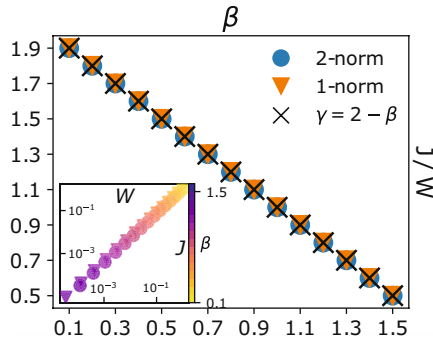
Lastly, we test the deviation of the cost of our methods from Dijkstra's one. In Fig. 4c we plot the relative cost difference taken in absolute value, that is  $\Delta J := |J_\Gamma - J_{\text{Dijkstra}}|/J_{\text{Dijkstra}}$  with Dijkstra's network cost calculated as  $J_{\text{Dijkstra}} = \sum_e \ell_e ||F_e||_1$ . This has a sharp drop at  $\beta = 1$ , where traffic is not favored nor penalized, with the cost of our network that is similar to the one of the shortest path returned by Dijkstra's algorithm. For  $\beta < 1$  we have  $J_\Gamma > J_{\text{Dijkstra}}$ , showing that penalizing traffic congestion has the drawback of producing more expensive infrastructures. We observe the opposite behavior for  $\beta > 1$ , where  $J_\Gamma < J_{\text{Dijkstra}}$ , with congested networks that are progressively cheaper as  $\beta$  increases.

**Network robustness to failures.** We now showcase a possible relevant application of our model by analyzing network's robustness to structural failures as nodes removal. Network managers interested in finding which stations are crucial for alleviating potential traffic overload can look at the congested transportation regime (we set  $\beta = 0.1$  to favor homogeneous fluxes) and investigate how fluxes resulting from our model distribute along the network.

In detail, we remove sequentially a total of four stations from the network: Châtelet, Gare du Nord, Saint-Lazare, and Gare de Lyon. The last three are those with the largest number of inflowing passengers, while Châtelet has a central position and a high node degree  $d = 8$ . Once each station was removed, its passengers were redirected to its neighboring nodes, and then solutions of the dynamics were found with this setting, as depicted in Fig. 5.

In Fig. 5a we display the 1 + 4 networks obtained removing none, and the stations indicated in Fig. 5b. In Fig. 5c we plot the Gini coefficients of the optimal transport networks against  $\beta$ . We notice that for  $\beta > 1$  all the points collapse together, regardless of the number of failures. This scenario, however, is of little interest for the situation we want to address, being flux aggregation already favored by  $\beta > 1$ . As for  $\beta < 1$ , the difference in Gini coefficient gets wider the lower the  $\beta$ , with the largest gap at  $\beta = 0.1$ , we thus investigate this case.

Removing Châtelet from the network causes a considerable jump in the Gini coefficient, thus increasing the possibility of traffic jams. In fact, as we see from the second plot in Fig. 5a, all the passengers who were traveling



**Figure 6.** Pareto front. We plot the dissipation/infrastructure cost ratio vs.  $\beta$ . Different points are averaged over 100 runs with random initialization of the conductivities,  $\mu_e(0) \sim U(0, 1)$ . Inset: infrastructure cost,  $W$ , vs. dissipation cost,  $J$ . Marker shapes are identical to those of the main plot, colors follow the colorbar over  $\beta$ .

on the South-West route branching from the city center are redirected in a way that congests southern arteries of the network. Removing Gare du Nord is not as crucial for traffic rerouting. Indeed, the main difference between the second and the third network of Fig. 5a is that passengers who were departing from Gare du Nord move to its southern neighboring station, Gare de l'Est, and modify only slightly their path. A large jump in the Gini coefficient is visible after removing Saint-Lazare, which seems to be fundamental for connecting the central area of Paris to its north side. In the fourth plot in Fig. 5a, we can see that traffic becomes highly congested in the northern branch directed from east to west. Gare de Lyon causes a negligible change in the Gini coefficient, associated to a modest traffic rerouting in the South-East part of the network.

**Pareto front.** To conclude our analysis of the multicommodity routing problem it is possible to verify that stationary solutions of Eqs. (1) and (2) lie in the Pareto front (Fig. 6), which can be expressed in closed form as:

$$\frac{J}{W} = \gamma, \quad (8)$$

with  $\gamma = 2 - \beta$  (see Methods "Pareto front derivation").

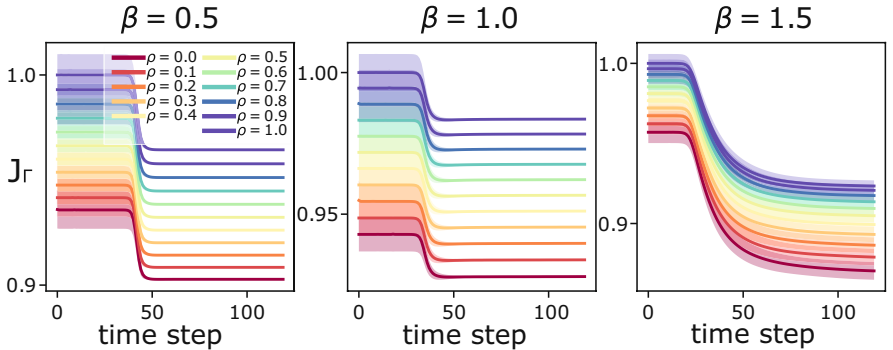
The emergence of a Pareto front between  $J$  and  $W$  is not limited to engineering networks like the ones studied here. A similar trade-off has been observed in the widened pipe model for plants of Koçillari<sup>32</sup>, where minimization of hydraulic resistance and of carbon cost compete for natural selection.

Moreover, looking at the inset of Fig. 3 and Fig. 6 we can observe that the Gini coefficient and the fraction of idle edges can be interpreted as driving forces responsible for the design of the optimal transport network, counterbalancing its cost. In fact, congested transport networks obtained for low values of beta  $\beta$  have a high cost, but are more resilient to damage—low Gini and no idle edges—being their infrastructure densely connected. On the contrary, setting  $\beta$  large has the effect of producing sparse networks. These infrastructures have the benefit of being cheaper, but they are less resistant to node and edge failures, as mentioned in "Network robustness to failures" section.

## Conclusions

Multicommodity routing is a powerful tool to model optimal network configurations in transportation systems<sup>18</sup>. In this work, we developed a robust and efficient model able to perform this task by finding stationary solutions of a dynamical system controlling fluxes and conductivities of edges. Our dynamics extends previous works focusing solely on the unicommodity<sup>15,27,28,30,43</sup>, and on the multicommodity setup<sup>25,26,34</sup>.

Precisely, we propose two different response functions regulating the growth of conductivities, whose evolution is dictated by the passengers moving in the metro. We performed a thorough empirical study of the optimal transport networks resulting in the two cases. Using metrics like the fraction of idle edges and the Gini coefficient of the fluxes, we found that the two functions behave similarly in the congested transportation regime, but differently in the branched transportation one. In this case, the 1-norm dynamics produces flows that are more concentrated on fewer edges, potentially leading to traffic overload. We addressed the capability of our method to recover shortest path networks by comparing it with Dijkstra's routine. Such comparison showed that our approach is a viable computational alternative to perform this task, achieving accurate results and being, in principle, scalable for large networks. Additionally, we performed an experiment to measure network robustness to infrastructural failures, revealing that the stations of Châtelet and Saint-Lazare are crucial to ease congestion of metro routes. Finally, we showed that solutions of our model lie in the Pareto front drawn by the energy dissipated during transport and the network infrastructural cost.



**Figure 7.** Validation of the dynamics with the 1-norm. We show the dissipation cost evolution along solution trajectories of Eqs. (1) and (2). Results are displayed for different combinations of  $\rho$  and  $\beta$ , and are averaged over 100 runs with random initializations of  $\mu_e(0) \sim U(0, 1)$ . The curves are normalized in (0, 1). Shaded areas denote standard deviations, these are thicker for  $\beta > 1$  since the cost is concave, with a rich landscape a local minima.

Altogether, our findings extend the current research in multicommodity routing problems using optimal transport principles and help to understand the mechanism underlying passenger flows in transportation systems.

Our formalism can be further extended to other possible applications related to the flow of passengers in transport networks. An example could be to incorporate time dependences in the passengers' inflows, modeling scenarios where stations are subject to different loads during a day, thus generalizing<sup>44</sup>. One could also compare the extent of traffic jams in multicommodity settings to that of other routing strategies for urban transport displaying notable phase transitions and scaling laws<sup>39,45,46</sup>.

We would like to remark that our approach is applicable to a variety of practical problems unrelated to transportation systems. A practitioner may then consider response functions for the dynamics alternative to those studied in this work. The analysis performed in this work show how such a problem can be addressed and paves the way for further research beyond urban transportation networks.

### Methods

**Lyapunov and dissipation cost equivalence.** Here we show that the first term of the Lyapunov functional in Eq. (6) is identical to the 2-norm dissipation cost  $J = (1/2) \sum_e \ell_e \|F_e\|_2^2 / \mu_e$ , we follow<sup>25</sup>. Multiplying both sides of Eq. (1) for  $p_v^i$  and summing over  $i$  and  $v$  yields

$$\sum_{i,v,u,e} (\mu_e / \ell_e) B_{ue} B_{ve} p_u^i p_v^i = \sum_{i,v} p_v^i S_v^i \tag{9}$$

$$\sum_e \frac{\ell_e}{\mu_e} \|F_e\|_2^2 = \sum_{i,v} p_v^i S_v^i, \tag{10}$$

where we made explicit the network Laplacian entries  $L_{vu} := \sum_e (\mu_e / \ell_e) B_{ue} B_{ve}$ , and we used the definition of the fluxes  $F_e^i := \mu_e (p_u^i - p_v^i) / \ell_e$ , for  $e = (u, v)$ , and  $\forall i$ . Equation (10) is the identity we wanted to prove.

**Preprocessing.** The original dataset in<sup>37</sup> is provided as a multilayer network embedded with different transportation types, thus we performed a preprocessing to extract the metro network. First, we trimmed nodes belonging to other layers and then merged redundant stations having the same name by collapsing them together. This redundancy was due to the presence of stations with two entrances located in slightly different geographical positions; their coordinates displacement was always negligible compared to the physical extension of the whole network. The trimmed graph reflects consistently the real topology of the Paris metro. For convenience, the longitude and the latitude of nodes are rescaled within the range [0, 1].

We did not have access to the exact travel routes data, so we assigned the entries of  $S$  based on the ‘‘importance’’ of each station. In fact the number of users validating their tickets when entering a station, the only data at disposal, is easier to track than the number of exiting users together with their entrance station. In practice, we assigned  $N - 1$  positive ‘‘influence factors’’ to each station  $i$ , one for each node  $u \neq i$  where the users entering in  $i$  can potentially exit:  $r_u^i = g^u / \sum_{w \neq i} g^w$ , instead  $r_{v=i}^i = 0$ , where  $g^v$  is the amount of users entering the metro from  $v = i$ . Note that  $0 \leq r_v^i \leq 1$  for all  $v$  nodes, and  $\sum_{v \neq i} r_v^i = 1$ . Thus, we can estimate the number of

people exiting from a station  $u \neq i$  by assigning  $S_u^i = -r_u g^{v=i}$ , while  $S_{v=i}^i = g^v$ . The intuition is that a station with a high entering volume of passengers, i.e. high  $g^v$ , should have a large amount of exiting users, thus its “influence” value  $r$  should be high.

**Validation.** We validate Eqs. (1) and (2) with the 1-norm for several combinations of  $\beta$  and of the input loads  $S(\rho)$ , with  $0 \leq \rho \leq 1$  progressively smoothing the passengers inflows data collected in<sup>47</sup>. In detail, users entering stations are regulated as  $g^v(\rho) = g^v - \rho(g^v - \langle g \rangle)$  where  $g^v$  are inflows in  $v$  as in<sup>47</sup>, and  $\langle \cdot \rangle$  averages over the nodes. Using this procedure, we build  $S(\rho)$  following the “influence assignment” described in Methods “Preprocessing”. Thus,  $S(\rho = 0) = S$  corresponds to the originally extrapolated mass matrix, while  $S_{v=i}^i(\rho = 1) = \langle g \rangle$  and  $S_{u \neq i}^i(\rho = 1) = -\langle g \rangle / (|V| - 1)$  for all  $u \neq v = i$ . Meaning that for  $\rho = 1$  passengers move with uniform rates from each—and to—all stations. In Fig. 7 we plot the time-evolution of  $J_{F_r}$ , which decreases over time.

**Pareto front derivation.** To obtain the Pareto form in closed form as in Eq. (8) it is sufficient to exploit the scaling  $\mu_e \sim (F_e)^\delta$ ,  $\delta = 3 - \beta$ , valid for stationary solutions of the multicommodity dynamics<sup>25</sup>. In particular, it is immediate to recover Eq. (8) by rewriting  $J$  in Eq. (3) as a function of the conductivities  $\mu_e$ .

### Data availability

All data used for the experiments on the Paris metro network are publicly available<sup>37,47</sup>.

### Code availability

An open-source implementation of the code is accessible at <https://github.com/aleable/McOpt>.

Received: 20 October 2021; Accepted: 21 April 2022

Published online: 06 May 2022

### References

- Sinclair, K. & Ball, R. C. Mechanism for global optimization of river networks from local erosion rules. *Phys. Rev. Lett.* **76**, 3360–3363. <https://doi.org/10.1103/PhysRevLett.76.3360> (1996).
- Rinaldo, A. *et al.* Minimum energy and fractal structures of drainage networks. *Water Resour. Res.* **28**, 2183–2195. <https://doi.org/10.1029/92WR00801> (1992).
- Rinaldo, A., Rodriguez-Iturbe, I., Rigon, R., Ijjasz-Vasquez, E. & Bras, R. L. Self-organized fractal river networks. *Phys. Rev. Lett.* **70**, 822–825. <https://doi.org/10.1103/PhysRevLett.70.822> (1993).
- Sun, T., Meakin, P. & Jossang, T. Minimum energy dissipation model for river basin geometry. *Phys. Rev. E* **49**, 4865–4872. <https://doi.org/10.1103/PhysRevE.49.4865> (1994).
- Konkol, A., Schwenk, J., Katifori, E. & Shaw, J. B. Interplay of river and tidal forcings promotes loops in coastal channel networks (2021). [arXiv:2108.04151](https://arxiv.org/abs/2108.04151).
- Ronellenfötsch, H. & Katifori, E. Global optimization, local adaptation, and the role of growth in distribution networks. *Phys. Rev. Lett.* <https://doi.org/10.1103/PhysRevLett.117.138301> (2016).
- Ronellenfötsch, H. & Katifori, E. Phenotypes of vascular flow networks. *Phys. Rev. Lett.* <https://doi.org/10.1103/PhysRevLett.123.248101> (2019).
- Katifori, E., Szöllösi, G. J. & Magnasco, M. O. Damage and fluctuations induce loops in optimal transport networks. *Phys. Rev. Lett.* <https://doi.org/10.1103/PhysRevLett.104.048704> (2010).
- Xia, Q. The formation of a tree leaf. *ESAIM Control Optim. Calc. Var.* **13**, 359–377. <https://doi.org/10.1051/cocv:2007016> (2007).
- Tero, A. *et al.* Rules for biologically inspired adaptive network design. *Science* **327**, 439–442. <https://doi.org/10.1126/science.1177894> (2010).
- Tero, A., Yumiki, K., Kobayashi, R., Saigusa, T. & Nakagaki, T. Flow-network adaptation in Physarum amoebae. *Theory Biosci.* **127**, 89–94. <https://doi.org/10.1007/s12064-008-0037-9> (2008).
- Tero, A., Kobayashi, R. & Nakagaki, T. Physarum solver: a biologically inspired method of road-network navigation. *Phys. A* **363**, 115–119. <https://doi.org/10.1016/j.physa.2006.01.053> (2006).
- Tero, A., Kobayashi, R. & Nakagaki, T. A mathematical model for adaptive transport network in path finding by true slime mold. *J. Theor. Biol.* **244**, 553–564. <https://doi.org/10.1016/j.jtbi.2006.07.015> (2007).
- Yamada, H., Toth, A. & Nakagaki, T. Intelligence: maze-solving by an amoeboid organism. *Nature* **407**, 470–470. <https://doi.org/10.1038/35035159> (2000).
- Bonifaci, V., Mehlhorn, K. & Varma, G. Physarum can compute shortest paths. *J. Theor. Biol.* **309**, 121–133. <https://doi.org/10.1016/j.jtbi.2012.06.017> (2012).
- Bonifaci, V. Physarum can compute shortest paths: a short proof. *Inf. Process. Lett.* **113**, 4–7. <https://doi.org/10.1016/j.ipl.2012.09.005> (2013).
- Bonifaci, V. A revised model of fluid transport optimization in Physarum polyccephalum. *J. Math. Biol.* **74**, 567–581. <https://doi.org/10.1007/s00285-016-1036-y> (2017).
- Salimifard, K. & Bigharaz, S. The multicommodity network flow problem: state of the art classification, applications, and solution methods. *Oper. Res. Int. Journal* <https://doi.org/10.1007/s12351-020-00564-8> (2020).
- Yeung, C. H. & Saad, D. Networking - a statistical physics perspective. *J. Phys. A Math. Theor.* <https://doi.org/10.1088/1751-8113/46/10/103001> (2011).
- Yeung, C. H., Saad, D. & Wong, K. Y. M. From the physics of interacting polymers to optimizing routes on the London Underground. *Proc. Natl. Acad. Sci.* **110**, 13717–13722. <https://doi.org/10.1073/pnas.1301111110> (2013).
- Altarelli, F., Braunstein, A., Dall’Asta, L., De Bacco, C. & Franz, S. The edge-disjoint path problem on random graphs by message-passing. *PLoS ONE* **10**, 1–18. <https://doi.org/10.1371/journal.pone.0145222> (2016).
- De Bacco, C., Franz, S., Saad, D. & Yeung, C. H. Shortest node-disjoint paths on random graphs. *J. Stat. Mech. Theory Exp.* **2014**, P07009 (2014).
- Xu, Y., Po, H. F., Yeung, C. H. & Saad, D. Scalable node-disjoint and edge-disjoint multi-wavelength routing (2021). [arXiv:2107.00609](https://arxiv.org/abs/2107.00609).
- Yeung, C. H. & Saad, D. Competition for shortest paths on sparse graphs. *Phys. Rev. Lett.* <https://doi.org/10.1103/PhysRevLett.108.208701> (2012).
- Lonardi, A., Facca, E., Putti, M. & De Bacco, C. Designing optimal networks for multicommodity transport problem. *Phys. Rev. Res.* <https://doi.org/10.1103/PhysRevResearch.3.043010> (2021).



26. Bonifaci, V. *et al.* Physarum-inspired multi-commodity flow dynamics. *Theor. Comput. Sci.* <https://doi.org/10.1016/j.tcs.2022.02.001> (2022).
27. Facca, E., Cardin, F. & Putti, M. Towards a stationary Monge-Kantorovich dynamics: the physarum polyecephalum experience. *SIAM J. Appl. Math.* **78**, 651–676. <https://doi.org/10.1137/16M1098383> (2016).
28. Facca, E., Daneri, S., Cardin, F. & Putti, M. Numerical solution of Monge-Kantorovich equations via a dynamic formulation. *J. Sci. Comput.* **82**, 68. <https://doi.org/10.1007/s10915-020-01170-8> (2020).
29. Facca, E., Cardin, F. & Putti, M. Physarum dynamics and optimal transport for basis pursuit (2020). [arXiv:1812.11782](https://arxiv.org/abs/1812.11782).
30. Facca, E., Cardin, F. & Putti, M. Branching structures emerging from a continuous optimal transport model. *J. Comput. Phys.* <https://doi.org/10.1016/j.jcp.2021.110700> (2021).
31. Baptistista, D., Leite, D., Facca, E., Putti, M. & De Bacco, C. Network extraction by routing optimization. *Sci. Rep.* <https://doi.org/10.1038/s41598-020-77064-4> (2020).
32. Baptistista, D. & De Bacco, C. Principled network extraction from images. *R. Soc. Open Sci.* <https://doi.org/10.1098/rsos.210025> (2021).
33. Baptistista, D. & De Bacco, C. Convergence properties of optimal transport-based temporal networks. In Benito, R. M. *et al.* (eds.) *Complex Networks and Their Applications X*, 578–592. [https://doi.org/10.1007/978-3-030-93409-5\\_48](https://doi.org/10.1007/978-3-030-93409-5_48) (Springer International Publishing, Cham, 2022).
34. Ibrahim, A. A., Lonardi, A. & De Bacco, C. Optimal transport in multilayer networks for traffic flow optimization. *Algorithm* <https://doi.org/10.3390/a14070189> (2021).
35. Dijkstra, E. A note on two problems in connexion with graphs. *Numer. Math.* **1**, 269–271. <https://doi.org/10.1007/BF01386390> (1959).
36. Banavar, J. R., Colaiori, F., Flammini, A., Maritan, A. & Rinaldo, A. Topology of the Fittest Transportation Network. *Phys. Rev. Lett.* **84**, 4745–4748. <https://doi.org/10.1103/PhysRevLett.84.4745> (2000).
37. Kujala, R., Weckström, C., Darst, R. K., Mladenović, M. N. & Saramäki, J. A collection of public transport network data sets for 25 cities. *Sci. Data* <https://doi.org/10.1038/sdata.2018.89> (2018).
38. Dixon, P. M., Weiner, J., Mitchell-Olds, T. & Woodley, R. Bootstrapping the gini coefficient of inequality. *Ecology* **68**, 1548–1551 (1987).
39. Yan, G., Zhou, T., Hu, B., Fu, Z.-Q. & Wang, B.-H. Efficient routing on complex networks. *Phys. Rev. E* <https://doi.org/10.1103/PhysRevE.73.046108> (2006).
40. Facca, E. & Benzi, M. Fast iterative solution of the optimal transport problem on graphs. *SIAM J. Sci. Comput.* **43**, A2295–A2319. <https://doi.org/10.1137/20M137015X> (2021).
41. Fredman, M. L. & Tarjan, R. E. Fibonacci heaps and their uses in improved network optimization algorithms. *J. ACM* **34**, 596–615. <https://doi.org/10.1145/28869.28874> (1987).
42. Koçillari, L. *et al.* The widened pipe model of plant hydraulic evolution. *Proc. Natl. Acad. Sci.* <https://doi.org/10.1073/pnas.2100314118> (2021).
43. Hu, D. & Cai, D. Adaptation and optimization of biological transport networks. *Phys. Rev. Lett.* <https://doi.org/10.1103/PhysRevLett.111.138701> (2013).
44. Lonardi, A., Facca, E., Putti, M. & De Bacco, C. Infrastructure adaptation and emergence of loops in network routing with time-dependent loads (2021). [arXiv:2112.10620](https://arxiv.org/abs/2112.10620).
45. Toroczkai, Z. & Bassler, K. E. Jamming is limited in scale-free systems. *Nature* **428**, 716–716. <https://doi.org/10.1038/428716a> (2004).
46. Zhao, L., Lai, Y.-C., Park, K. & Ye, N. Onset of traffic congestion in complex networks. *Phys. Rev. E* <https://doi.org/10.1103/PhysRevE.71.026125> (2005).
47. "Trafic annuel entrant par station du réseau ferré 2019", accessed: 2020-08-28 (2019).

## Acknowledgements

We thank Enrico Facca for helpful discussions and for insightful comments improving the manuscript. We acknowledge the help of Daniela Leite for data preprocessing. The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Alessandro Lonardi.

## Author contributions

All authors contributed to developing the models. A.L. and C.D.B. conceived the experiments, analyzed the results and reviewed the manuscript. A.L. conducted the experiments. All authors have read and agreed to the final version of the manuscript.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-11348-9>.

**Correspondence** and requests for materials should be addressed to A.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

## Infrastructure adaptation and emergence of loops in network routing with time-dependent loads

Alessandro Lonardi <sup>1,\*</sup> Enrico Facca <sup>2</sup> Mario Putti,<sup>3</sup> and Caterina De Bacco <sup>1,†</sup><sup>1</sup>Max Planck Institute for Intelligent Systems, Cyber Valley, 72076 Tübingen, Germany<sup>2</sup>Laboratoire Paul Painlevé, UMR No. 8524, CNRS, Inria, Université Lille, 59000 Lille, France<sup>3</sup>Department of Mathematics “Tullio Levi-Civita,” University of Padua, Via Trieste 63, 35131 Padua, Italy

(Received 27 September 2022; accepted 24 January 2023; published 3 February 2023)

Network routing approaches are widely used to study the evolution in time of self-adapting systems. However, few advances have been made for problems where adaptation is governed by time-dependent inputs. In this work we study a dynamical systems where the edge conductivities of a network are regulated by time-varying mass loads injected on nodes. Motivated by empirical observations, we assume that conductivities adapt slowly with respect to the characteristic time of the loads. Furthermore, assuming the loads to be periodic, we derive a dynamics where the evolution of the system is controlled by a matrix obtained with the Fourier coefficients of the input loads. Remarkably, we find a sufficient condition on these coefficients that determines when the resulting network topologies are trees. We show an example of this on the Bordeaux bus network where we tune the input loads to interpolate between loopy and tree topologies. We validate our model on several synthetic networks and provide an expression for long-time solutions of the original conductivities.

DOI: [10.1103/PhysRevE.107.024302](https://doi.org/10.1103/PhysRevE.107.024302)

### I. INTRODUCTION

Optimized transport of resources is a pivotal contributing factor in determining the structural evolution of real-world networks. Archetypes for self-organizing systems that ramify into networks in order to optimize energy expenditure rates are xylem conduits in leaves [1–4], river basins [5–9], and slime molds [10–19]. These formations are not only restricted to the natural realm but can also be generated by anthropogenic processes. A prominent example is that of transportation networks such as railway and metro systems, which are designed to jointly optimize traffic overload and infrastructural cost [20–22].

Typically, optimal transport of mass in networks is set as a minimization problem where resources moving through the edges have to satisfy a set of constraints, e.g., conservation of mass, while minimizing a suitable transportation cost [1,3,19,23–30]. Several efficient methods have been proposed to solve this problem. A popular approach is that of message-passing algorithms [31], where sources of mass are matched in sender-receiver pairs and messages encode mass transfer between them [22,32–36]. Promising results have also been obtained with optimal transport theory [2,15,20,27,37–42], the approach we consider in this work. The general idea behind this method is to describe the transport of mass as a

process being regulated by edge capacities, quantities evolving with a dynamical system to allocate mass fluxes.

Despite their usage in modeling transportation problems across domains, a common drawback of all these methods is to consider only stationary loads, i.e., resources that are injected and travel through the network do not change with time. This assumption may not be valid in certain scenarios. For instance, blood vessels are known for adapting their structure continuously to meet changing metabolic demands [25,43–45]. Similarly, passengers in transportation systems enter stations with hourly, weekly, and seasonal time-varying rates [46].

A viable approach to model these systems is to control the network evolution considering an ensemble average of the stress generated by the loads [24,25,37]. This relies on assuming stationary loads on nodes but with their positions varied stochastically. The ensemble average over the loads' locations is then computed as a proxy of a system with loads of fixed locations but time-varying amounts. This technique has also been employed to study network resilience to edge cutting [3] or for routing problems with spatially correlated loads [2].

Remarkably, adding stochasticity in the loads may lead to the emergence of loops in the resulting optimal networks topologies [2,3,24,25,37]. This result is complementary to the hierarchical formation of trees since loops provide alternative routes to accommodate fluctuations or guarantee robustness against broken links. Recently, loop formation has also been observed in multicommodity setups [27,28], where the loads are deterministic inputs of the problem. In this case loop generation is a consequence of having different commodities interacting in a unique shared infrastructure.

In all these works, the time-varying character of the transport network loads is neglected because the main problem

\*alessandro.lonardi@tuebingen.mpg.de

†caterina.debacco@tuebingen.mpg.de

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by the Max Planck Society.

variables are taken on average. Here we develop a model that considers the explicit time dependence of the mass inflows and investigate, both analytically and numerically, the long-term behavior of time-varying transport networks. This allows us to show that it is not the process uncertainty, inherent in any stochastic framework, but the nonstationarity of the loads that promotes loops, which is fundamentally different from what can be concluded using stochastic formulations.

In particular, we generalize the routing problem in the work of Facca *et al.* [41] by considering periodic mass loads on nodes. We postulate an analytical relationship connecting physical quantities as the edge conductivities and the coefficients of the Fourier series expansion of the loads. We then define a dynamics that rapidly converges to the long-run average solutions of the original dynamics.

Our model relies on the idea of distinguishing slow-varying variables from fast ones. The first are capacities of edges that are regulated by the Fourier coefficients of the forcing; the second are, for example, loads of passengers entering and exiting network nodes. The physical intuition is that, while fluxes of passengers in a transportation system have the same rate of change of the network loads, roads do not. In fact, it is reasonable to assume that a network manager has a coarser observation scale of a transportation system than the users, whose paths rapidly fluctuate. In practice, this means that modifications in the network infrastructure occur on a much larger timescale than that of daily passengers' fluctuations.

Remarkably, we find that the Fourier decomposition of the loads yields a sufficient condition to determine whether the resulting optimal networks will contain loops or be a tree. Performing a numerical validation of our dynamics on synthetic networks, we are also able to provide an analytical expression for the long-run conductivities. Precisely, we find that the conductivities start oscillating around constant values at large timescales and at certain frequencies that can be expressed in terms of those of the input loads. Furthermore, we define a Lyapunov functional for our dynamical formulation, allowing us to interpret stationary topologies as optimal networks, i.e., structures minimizing the global cost to build the graph. Finally, we examine a case study with loads that are the sum of decoupled harmonic oscillators, finding that the condition on the Fourier coefficients can be equivalently reformulated in terms of the loads' amplitudes and phases. We numerically investigate this last setup on the Bordeaux bus network.

## II. TIME-VARYING LOADS IN ROUTING OPTIMIZATION ON NETWORKS

Consider a network  $G$  with nodes  $v \in V$  and edges  $e \in E$ , each of length  $\ell_e > 0$ . The orientation of the edges is conventionally assigned by the signed incidence matrix of the graph, with entries  $B_{ve} = \pm 1$  if node  $v$  is the tail or the head of edge  $e$ , and  $B_{ve} = 0$  otherwise. We consider a routing optimization problem on  $G$  setting time-varying mass loads  $S(t) = \{S_v(t)\}$  on nodes being the amount of mass either injected in ( $S_v(t) > 0$ ) or extracted from ( $S_v(t) < 0$ ) node  $v$ . Concretely, one could think of  $S(t)$  as a time-dependent origin-destination vector of passengers moving in a transportation network, where mass entries correspond to the fraction of passenger flowing through stations. This allows us to write Kirchoff's conservation law

as

$$\sum_u L_{vu}(\mu) p_u(t) = S_v(t) \quad \forall v \in V \forall t \geq 0, \quad (1)$$

where  $\mu = \{\mu_e\}$  are the non-negative edge conductivities,  $p(t) = \{p_v(t)\}$  are pressure potentials on nodes, and  $L_{vu}(\mu) := \sum_e B_{ve}(\mu_e/\ell_e) B_{ue}$  are the entries of the weighted Laplacian of the network [47]. The conductivities can be interpreted as the capacities that the edges must have to allocate the mass loads acting on the nodes; thus we can consider them proportional to the edges' sizes. When considering passengers moving along a transportation network,  $\mu$  can be seen as the width of a road or more generally a measure of the infrastructure's resources used to carry traffic flows.

We propose a model in which the forcings  $S(t)$  dictate the time evolution of the conductivities by means of a feedback dynamics. In particular, we couple Eq. (1) with the system of ordinary differential equations (ODEs)

$$\frac{d\mu_e(t)}{dt} = \frac{F_e^2(t)}{\mu_e^2(t)} - \mu_e(t) \quad \forall e \in E, \quad (2)$$

$$\mu_e(0) = m_e \quad \forall e \in E, \quad (3)$$

with  $m_e > 0$  initial values. For a solution trajectory  $\mu(t)$ , we define the fluxes  $F_e(t) \equiv F_e(\mu(t), S(t)) := \mu_e(t)[p_u(t) - p_v(t)]/\ell_e$  for  $e = (u, v)$ , with  $p_v(t) \equiv p_v(\mu(t), S(t)) := \sum_u L_{vu}^{\dagger}(\mu(t)) S_u(t)$  the solution of Eq. (1), where  $L^{\dagger}$  denotes the Laplacian pseudoinverse. We assume that the system is isolated, namely,  $\sum_v S_v(t) = 0 \forall t \geq 0$ , so  $p(t)$  is a well-defined potential (see [47], Lemma 0). Specifically, Klein and Randić [47] showed that  $L$  generally is not invertible, but Eq. (1) can be solved by the pseudoinverse within the subspace orthogonal to the unitary vector, that is, when  $\sum_v S_v(t) = 0$ .

In Eq. (2) the growth in time of the conductivities is proportional to the flux forcing term  $F_e^2(t)$  with  $\mu$  decaying exponentially when no flux flows through an edge. In practice, this corresponds to enlarging a road when many passengers travel along it and reducing it when there is no traffic. We illustrate this intuition with a schematic representation in Fig. 1.

The free parameter  $0 < \gamma < 2$  tunes between different transportation mechanisms [27,38,41]. The case  $\gamma < 1$  encourages mass consolidation on a few edges,  $\gamma = 1$  is shortest-path-like, and  $1 < \gamma < 2$  penalizes traffic congestion.

Our dynamical formulation assumes continuous variables for fluxes and conductivities, but the mass  $S(t)$  could be arbitrarily continuous or discrete. While this is valid in many scenarios (e.g., when modeling a large number of individuals), it may be limiting in cases where a discrete (or atomic) representation is necessary to capture fine-grain differences in the number of passengers. For this, one should consider alternative formulations and approaches, for instance, using message passing or belief propagation as in [22,32–36].

Finally, we remark that Eq. (2) can be made scale independent with respect to the model variables by an opportune nondimensionalization that we describe in detail in Appendix A.

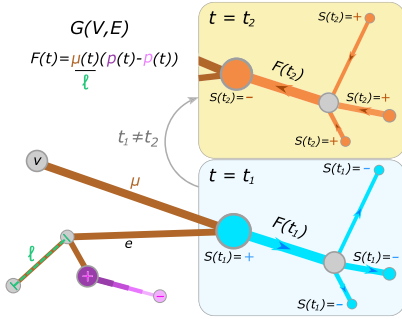


FIG. 1. Schematic visualization of the problem. We highlight conductivities (brown), the length of an edge (green), and the difference of pressure (purple) along an edge that triggers their fluxes. In the rightmost blue and yellow panels we depict two scenarios where the time-dependent loads  $S(t)$  generate fluxes that move from the central part of the network to its periphery ( $t = t_1$ ) and vice versa ( $t = t_2 \neq t_1$ ). Node and edge widths are proportional to  $S(t)$  and  $F(t)$ , respectively.

### III. MODEL CONSTRUCTION

#### A. Slow adaptation of conductivities

In several biological systems the adaptation time of organisms is much slower (weeks) than the characteristic time of the mass injected in the system (seconds) [25,43–45]. In order to describe these organisms, a common approach is that of approximating the fast time-varying input loads with combinations of open and closed switchlike nodes with constant inflows and to assume that the conductivities are regulated by an ensemble average of the pressures over different states of the loads [3,24,25,37].

Instead, here we want to model the evolution of these slow adapting conductivities taking in account the time dependence of the loads. We formalize this hypothesis by assuming (i) the existence of a slow timescale  $\tau$ , with  $\tau = Kt$  and  $K \gg 1$ , and that (ii) in a fixed time window  $\Delta$ , small with respect to the slow variable  $\tau$  but large with respect to the time  $t$ ,

$$\hat{\mu}_e(\tau + t') \approx \hat{\mu}_e(\tau) \quad \forall e \in E \forall t' \in [0, \Delta) \forall \tau \geq 0 \quad (4)$$

holds for some conductivities  $\hat{\mu} = \{\hat{\mu}_e\}$  with the natural time of evolution being  $\tau$ . Timescales are depicted in Fig. 2. We can interpret  $t$  as seconds,  $\Delta$  as days, and  $\tau$  as months. Such distinction between different natural timescales is observed in the interplay of rivers and tide loads in coastal delta formation [9], where the assumption is that tides cycle much faster than the river channel adaptation, a distinction analogous to that between  $t$  and  $\tau$ .

Finally, we assume that (iii) the evolution in  $\tau$  of  $\hat{\mu}$  is determined by the time integral average of the product of the mass loads. Assumptions (i), (ii), and (iii) together lead to the

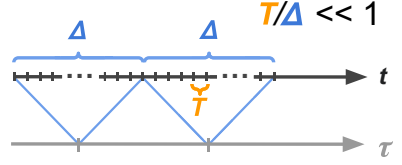


FIG. 2. Schematic representation of the different time variables. The two arrows denote timescales  $t$  and  $\tau$ . The time windows  $\Delta$ , large with respect to  $t$ , are denoted with blue curved brackets and the fast period  $T$  in orange. Each window  $\Delta$  along which we integrate the dynamics (1)–(3) contains a large number of periods  $T$ .

definition of

$$\hat{\Phi}_e(\hat{\mu}, \tau) := \frac{\hat{\mu}_e^{2-\gamma}}{\ell_e^2} \sum_{uv} A_{eu}(\hat{\mu}) A_{ev}(\hat{\mu}) \times \frac{1}{\Delta} \int_{\tau}^{\tau+\Delta} S_u(t) S_v(t) dt - \hat{\mu}_e \quad (5)$$

for all  $e \in E$  and  $\tau \geq 0$ , where we introduced  $A_{ev}(\hat{\mu}) := \sum_u B_{ev} L_{vu}^{\gamma} \forall e \in E \forall v \in V$ . The functional  $\hat{\Phi}$  is the natural approximation of the right-hand side of Eq. (2), as shown in Appendix B.

We define then a trajectory  $\hat{\mu}(\tau)$  as a solution of the dynamics

$$\frac{d\hat{\mu}_e(\tau)}{d\tau} = \hat{\Phi}_e(\hat{\mu}(\tau), \tau) \quad \forall e \in E, \quad (6)$$

$$\hat{\mu}_e(0) = \hat{m}_e \quad \forall e \in E, \quad (7)$$

with  $\hat{m}_e > 0$  initial conditions. In general,  $\hat{\Phi}$  is difficult to manipulate as the loads  $S(t)$  may assume any arbitrary expression, possibly preventing the exact computation of the time integrals. For this reason, we investigate its behavior for a particular class of functions  $S(t)$  that allows for analytical tractability.

#### B. Periodicity of the loads

We consider periodic loads  $S(t)$ , with period  $T$  small with respect to the fixed integration window  $\Delta$  introduced in Sec. III A:

$$S_v(t + T) = S_v(t), \quad T/\Delta \ll 1 \quad \forall v \in V \forall t \geq 0. \quad (8)$$

This allows us to express each  $S_v(t)$  using its Fourier series  $S_v(t) = \sum_{n \in \mathbb{Z}} c_n^v \exp(i\omega n t)$ , with  $\omega = 2\pi/T$ . Substituting this into Eq. (5) yields the pivotal result

$$\frac{1}{\Delta} \int_{\tau}^{\tau+\Delta} S_u(t) S_v(t) dt = C_{uv} + O(\Delta), \quad (9)$$

holding for all  $u, v \in V$  and  $\tau \geq 0$ . The matrix  $C$  has entries  $C_{uv} := \sum_n (c_n^u)^* c_n^v \forall u, v \in V$ , with  $c^*$  denoting the complex conjugation of  $c$ . The term  $O(\Delta)$  contains all negligible contributions  $\varepsilon$ , decaying as  $\varepsilon/\Delta \rightarrow 0$  for  $\Delta \rightarrow +\infty$ . For a detailed derivation of this result, one can refer to Appendix C.

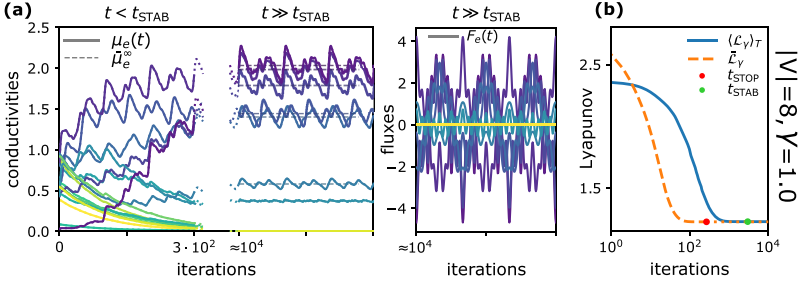


FIG. 3. Characterization of the fast conductivities  $\mu(t)$ . All results are computed on a synthetic network with  $|V| = 8$  and setting  $\gamma = 1.0$ . (a) Fast conductivities  $\mu(t)$  and fluxes  $F(t)$  are drawn with solid lines, stationary solutions  $\bar{\mu}^\infty$  are dashed. Labels on the  $x$  axis correspond to the number of iterations of the numerical discretization of Eqs. (1)–(3). Conductivities are depicted in two time windows, before and after their stabilization time  $t_{\text{STAB}}$ . Fluxes are drawn only for  $t \gg t_{\text{STAB}}$ . Colors denote different edges. (b) Evolution of  $\bar{L}_\gamma$  and of  $\langle L_\gamma \rangle_T$  in time. The green and the red circles denote  $t_{\text{STAB}}$  and  $t_{\text{STOP}}$ , respectively.

### C. Periodic-load dynamics

Combining Eqs. (5) and (9), we can build a dynamics for some new conductivities  $\bar{\mu} = \{\bar{\mu}_e\}$ , which evolve in the long timescale  $\tau$ . Precisely, we ignore negligible contributions in Eq. (9) and define

$$\frac{d\bar{\mu}_e(\tau)}{d\tau} = \bar{\Phi}_e(\bar{\mu}_e(\tau)) \quad \forall e \in E, \quad (10)$$

$$\bar{\mu}_e(0) = \bar{m}_e \quad \forall e \in E, \quad (11)$$

with  $\bar{m}_e > 0$  initial conditions. The right-hand side of Eq. (10) is such that  $\bar{\Phi} \simeq \Phi$  for  $\Delta \gg 1$  and reads

$$\bar{\Phi}_e(\bar{\mu}) := \frac{\bar{\mu}_e^{2-\gamma}}{\ell_e^2} \sum_{uv} A_{eu}(\bar{\mu}) A_{ev}(\bar{\mu}) C_{uv} - \bar{\mu}_e \quad \forall e \in E. \quad (12)$$

An important point is that the problem in Eqs. (10) and (11) is *not* equivalent to the dynamics Eqs. (1)–(3) with each  $S_v(t)$  integrated over  $T$ . The latter case would imply that  $C_{uv}$  had the form  $C_{uv} = \bar{S}_u \bar{S}_v$ , where  $\bar{S}_v$  is the integral of  $S_v(t)$  over the period. This is only a particular case of the dynamics in Eqs. (10) and (11).

Noticeably, in this case the condition  $\text{rank}(C) = 1$  holds, i.e., since  $C$  is symmetric, there exists a vector  $y \in \mathbb{R}^{|V|}$  such that  $C_{uv} = y_u y_v \forall u, v \in V$ . This is a sufficient condition for Eqs. (10) and (11) to return a loopless network at convergence (see Appendix D for a proof) and confirms previous results observed for constant loads [23,26,29].

However, this condition does not hold generally for any arbitrary choice of the loads, as  $C$  may have a more general expression, in particular  $\text{rank}(C) > 1$ . Moreover, the case of constant loads is not the only one where  $\text{rank}(C) = 1$ . We provide an example of this in Sec. VB, where we explore the case of each  $S_v(t)$  being the sum of a finite number of harmonic oscillators.

## IV. CHARACTERIZATION OF THE FAST DYNAMICS

Finding an analytical expression for the fast conductivities  $\mu(t)$  solutions of Eqs. (1)–(3) cannot be done by directly solving the dynamics, because of the nonlinear dependence

on  $\mu(t)$  in the Laplacian pseudoinverse. Nevertheless, here we propose an argument to characterize their long-time behavior.

We support our findings with an empirical validation on synthetic networks built taking the Delaunay triangulation of  $|V|$  nodes placed at random in the unit square. In our experiments, we set  $|V| = 2^i$ , with  $i = 3, \dots, 9$ . The vector of loads  $S(t)$  is  $S(t) := 20 S_1(t) + 10 S_4(t) + 5 S_8(t)$ , where each factor is defined as  $S_n(t) := q_n \cos(\omega t)$ , with amplitudes extracted at random from a  $|V|$ -dimensional Dirichlet distribution as  $q_n \sim \mathcal{D}(\alpha = 1) - 1/|V|$  [so that  $\sum_n S_n(t) = 0 \forall t \geq 0$ ] and  $n = 1, 4, 8$ . The period has been conventionally set to have  $\omega = 2\pi$ .

We observe that the evolution of the fast conductivities is typically divided in two phases, as shown in Fig. 3(a). First, the conductivities undergo a stabilization transient for  $t < t_{\text{STAB}}$ , where they strongly depend on their initial conditions  $m_e$  and significantly change their mean values. Then, when  $t > t_{\text{STAB}}$ , the conductivities reach a plateau and oscillate around fixed values. More precisely, either they move around mean values that are far from zero and preserve their oscillatory nature for all times or they decay to zero with negligible oscillations that are progressively damped as  $t$  increases. These experimental observations suggest the following ansatz for the stabilized solutions, for all  $t > t_{\text{STAB}}$  and  $e \in E$ :

$$\mu_e(t) = a_e + b_e(t) \text{ s.t. } a_e = \text{const}, \quad b_e(t+T) = b_e(t). \quad (13)$$

We compare solutions of the new dynamics (10) and (11) with those of Eqs. (1)–(3) [see Fig. 3(a) for an example]. In the figure the conductivities  $\mu(t)$  are oscillating around the constant values of  $\bar{\mu}(\tau)$  reached at convergence, which we denote with  $\bar{\mu}^\infty = \{\bar{\mu}_e^\infty\}$ . Motivated by this empirical observation, we set

$$a_e = \bar{\mu}_e^\infty \quad \forall e \in E. \quad (14)$$

We experimentally notice that also the fluxes start to oscillate around a constant value after a first stabilization time

interval [see Fig. 3(a)]. We use this evidence to deduce (see Appendix E) that the main oscillatory modes of the conductivities are resonant with the squared fluxes and have the form

$$b_e(t) = \sum_{n,m \in \mathcal{N}} b_e^n b_e^m \exp[i\omega(n+m)t] \quad \forall e \in E, \quad (15)$$

with  $\mathcal{N} := \{n_e\}$  the set of the Fourier modes of the loads. Hence, the conductivities oscillate with modes determined by those of the loads. This result is supported by several numerical experiments (see Appendix E for details).

Remarkably, these numerical experiments also serve as a validation for hypothesis (ii) in Sec. III A. In fact, for any sufficiently slow time  $\tau$ , the conductivities fluctuate around a constant value, suggesting the possibility of neglecting their fast oscillatory nature when studying asymptotics of Eqs. (1)–(3).

### Candidate Lyapunov functional

We empirically observe [see Fig. 3(b)] that our new dynamics (10) and (11) admits a candidate Lyapunov functional reading

$$\bar{\mathcal{L}}_Y(\bar{\mu}) := \frac{1}{2} \sum_e \frac{\ell_e}{\bar{\mu}_e} F_e^2(\bar{\mu}) + \frac{1}{2\gamma} \sum_e \ell_e \bar{\mu}_e^\gamma, \quad (16)$$

where for each edge  $e$  we define the squared slow fluxes  $F_e^2 := (\bar{\mu}_e^2/\ell_e^2) \sum_u A_{eu}(\bar{\mu})A_{ev}(\bar{\mu})C_{uv}$ . Noticeably, if  $\text{rank}(C) = 1$  holds, it is possible to formally prove that  $\bar{\mathcal{L}}_Y(\bar{\mu})$  is a well-defined Lyapunov functional for Eqs. (10) and (11) (see Appendix D for detailed derivations). In addition, we can interpret the functional as in Ref. [27] for multicommodity optimal transport. Namely, the Lyapunov is the sum of a dissipation cost, the first addend in Eq. (16), with an infrastructural cost, the price needed to build the transport network.

We notice empirically that the functional reaches a plateau at  $t_{\text{STOP}}$ , defined as the time for which  $\Delta \bar{\mathcal{L}}_Y/\delta t < \varepsilon$  is satisfied, with  $\Delta \bar{\mathcal{L}}_Y := |(\bar{\mathcal{L}}_Y)^{r+1} + (\bar{\mathcal{L}}_Y)^r|/(\bar{\mathcal{L}}_Y)^{r+1}$ , where the upper indices are consecutive iterations in the finite-difference discretization of Eqs. (10) and (11). In all our experiments, we set  $\delta t = 0.1$  as time step of a forward Euler method and the convergence threshold to  $\varepsilon = 10^{-5}$ .

Additionally, we observe that the candidate Lyapunov functional  $\bar{\mathcal{L}}_Y$  converges to a value that is the same achieved by the running average functional over the period  $T$ :

$$(\mathcal{L}_Y)_T := \frac{1}{T} \int_t^{t+T} \left( \frac{1}{2} \sum_v p_v(\mu) S(t') + \frac{1}{2\gamma} \sum_e \ell_e \mu_e^\gamma \right) dt', \quad (17)$$

with  $\mu$  that is evaluated along solution trajectories of Eqs. (1)–(3). The functional (17) reaches a plateau at the stabilization time  $t_{\text{STAB}}$ , when the fast conductivities  $\mu(t)$  start oscillating around constant values. Remarkably, in Fig. 3(b) we see that  $t_{\text{STAB}} \gg t_{\text{STOP}}$ , which is due to the fact that the time step  $\delta t$  for the numerical discretization of Eqs. (1)–(3) has to be set much lower than  $\delta t$  in order to capture the oscillatory nature of the loads. In our experiments we set it to  $\delta t = \delta t/10$ . A practical consequence of this is that the discretization of Eqs. (10) and (11) is a fast and scalable alternative to extract the conductivities around which long-run solutions of Eqs. (1)–(3) stabilize.

Because of this analogy between an optimal transport (functional minimization) setup and the solutions of our dynamical system, we can interpret the networks determined from the dynamics in Eqs. (10) and (11) as optimal topologies minimizing the infrastructural and dissipation cost. These networks can also be obtained by averaging long-run solutions of the original dynamics in Eqs. (1)–(3). In fact, as discussed in Sec. IV, long-run trajectories of Eqs. (1)–(3) oscillate around asymptotics of the newly defined dynamical system in Eqs. (10) and (11).

## V. GENERATION OF LOOPS

### A. Conditions for the generation of loops in closed form

If  $C$  has  $\text{rank}(C) = 1$ , i.e.,  $C_{uv} = y_u y_v$  for some  $y \in \mathbb{R}^{|V|}$ , the dynamics (10) and (11) produces trees at convergence. One trivial case where this holds is when the loads  $S(t)$  are static, i.e., constant for all times. However, this is not the only setting where  $\text{rank}(C) = 1$  is satisfied. In particular, there are cases where such a condition holds but  $S$  do change in time.

Here we explore a case of study proposing an ansatz where the loads are the sum of decoupled harmonic oscillators

$$S_v(t) = \sum_{i=1}^{N_v} A_v^i \cos(\omega n_v^i t + \phi_v^i) + d_v \quad \forall v \in V, \quad (18)$$

with  $\omega = 2\pi/T$ ,  $n_v^i, N_v \in \mathbb{N}$ , and  $A_v^i, d_v \in \mathbb{R}$ . By construction, these loads are periodic in  $T$ ; hence we compare them with their Fourier series representation  $S_v(t) = a_v^0/2 + \sum_{n_v \geq 1} a_{n_v}^{v+} \cos(\omega n_v t + \phi_v^{n_v})$ . Equating this expression with Eq. (18) yields

$$c_{n_v}^{v+} = \frac{A_v^i}{2} \exp(i\phi_v^i) \delta_{n_v, n_v^i} \quad \forall n_v \in \mathbb{N}, \quad (19)$$

where we conventionally set  $\phi_v^0 = 0 \forall v \in V$  and where only a finite number of Fourier coefficients are different from zero, given that the sum in Eq. (18) is finite.

The goal here is to express  $\text{rank}(C) = 1$  in terms of  $\{A_v^i\}$ ,  $\{n_v^i\}$ , and  $\{\phi_v^i\}$ , amplitudes, modes, and phases of the harmonic oscillators, respectively. To do that, we start by noting that  $\text{rank}(C) = 1$  is satisfied if and only if  $C_{uv} = y_u y_v \forall u, v \in V$ , with  $y_v = \pm \sqrt{C_{vv}}$ , and where the plus or minus signs have to be determined among  $2^{|V|}$  possible choices in such a way that  $\sum_v y_v = 0$  (see Appendix F).

Defining the complex vectors  $v_v = \{c_{n_v}^{v+}\}$  with entries of the Fourier coefficients in Eq. (19), we rewrite  $C_{uv} = \pm \sqrt{C_{uv}} \sqrt{C_{vv}}$  as  $v_u \cdot v_v = \pm \|v_u\| \|v_v\|$ , where the centered dot denotes the complex dot product and  $\|\cdot\|$  is its correspondent norm. Thus, the rank condition on  $C$  can be reformulated in terms of an equivalent linear dependence condition of the form  $v_v = \lambda v_u$  between the vectors  $v_v, v \in V$ , and for  $\lambda \neq 0$ . Finally, substituting Eq. (19) in this linear dependence condition leads to the following main result.

**Proposition 1.** Let the time-dependent loads  $S(t)$  injected in the network nodes be as in Eq. (18). If the following hold, then, for any  $\gamma \leq 1$ , a stationary solution of Eqs. (10) and (11) is a tree: (a)  $\phi_v^k = \phi_u^k + k\pi$ ,  $k \in \mathbb{Z}$ , i.e., sources and sinks are in phase, and (b)  $A_v^k \delta_{n_v, n_v^k} = \lambda(-1)^k A_u^k \delta_{n_u, n_u^k}$  (implying that  $N_v = N$  for all  $v$ ).

For a formal justification of this result see Appendix F.



FIG. 4. Bordeaux bus optimal transport network. (a) Network visualization. Input loads have been built as described in Sec. VB. The tree network originated by  $C$  with rank 1 is plotted in blue, the loopy topology in orange. The yellow and the magenta stars denote the geographical location of the two loads, the green squares those of the sinks. Here the width of edges corresponds to slow conductivities at convergence  $\bar{\mu}_e^\infty$ . Results are plotted for  $\gamma = 0.9$ . (b) Basis loop fraction against rank( $C$ ). Points correspond to averages over 100 runs of the experiments where positions of the sources and sinks are extracted at random. Shaded regions denote their standard deviations. Results are displayed for  $\gamma = 0.5$ .

### B. Numerical tests on the Bordeaux bus

In order to test the rank condition on  $C$  we design two experiments on the real network of the buses of Bordeaux. The network topology has been constructed focusing on a central region of the city and using data collected from [48]. Here we assume that the loads, representing passengers entering or exiting the network, vary much faster than the conductivities. These latter quantities can be thought of as the size of the roads that a network manager needs to design; thus we can safely assume their evolution to happen on a larger timescale with respect to that of  $S(t)$ .

First, we design a simulation with two source nodes  $v_1$  and  $v_2$  [the stars in Fig. 4(a)] and five sinks [the green squares in Fig. 4(a)] that we extract at random among the nodes of the network. Then we consider two cases where the sources are built in such a way that (i) rank( $C$ ) = 1 and (ii) rank( $C$ ) = 2. These are, respectively, (i)  $S_{v_1}(t) = S_{v_2}(t) = 100 \cos(\omega t)$ , with  $\omega = 2\pi$ , and (ii)  $S_{v_1}(t) = 100 \cos(\omega_1 t)$ , with  $\omega_1 = 2\pi$ , and  $S_{v_2}(t) = 100 \cos(\omega_2 t)$ , with  $\omega_2 = 4\pi$ . All the sinks  $u \neq v_1, v_2$  have loads  $S_u(t) = -[S_{v_1}(t) + S_{v_2}(t)]/5$  in both cases, to ensure conservation of mass.

We expect that in the first case the network extracted from Eqs. (10) and (11) with  $\gamma \leq 1$  is a tree. In the second case the network can possibly contain loops. We run the dynamics setting  $\gamma = 0.9$  and we display our findings in Fig. 4(a). The empirical results reflect our predictions: the blue network (the first case) is a tree. In contrast, in the orange network (the second case) loops emerge.

We further validate our results on the bus network of Bordeaux with a second experiment. We assign the loads  $S_v(t) = \sum_{i=1}^n S_i^v(t)$ , with  $S_i^v = (100/|Q_n|) \cos(\omega_i t)$ , to a set  $Q_n$  of randomly extracted nodes and  $S_u^v = -[100/(|V| - |Q_n|)] \cos(\omega_i t)$  to the remaining ones. The modes are  $n = 1, \dots, 6$ , while the number of nodes which are randomly extracted for each  $n$  are  $Q := |Q_n| = \{1, 5, 10, 20\}$ . We set again  $\omega = 2\pi$ .

Exploiting the exact relation that the matrix  $C$  has with the modes of the loads (see Sec. VA), it is possible to see that our particular construction of  $S(t)$  gives ranks ranging in

$1 \leq \text{rank}(C) \leq 6$ . We show our results in Fig. 4(b), where we plot the fraction of basis loops of the network against the rank of  $C$ . The dynamics is executed for  $\gamma = 0.5$  and the random extraction of the forcings has been varied over 100 runs. In the plot, it is clearly visible that for all values of  $Q$ , the fraction of basis loops is zero at rank( $C$ ) = 1. Moreover, we can see that when we increase the complexity of the problem, i.e., when rank( $C$ ) grows, the values attained on the  $y$  axis also increase. This suggests that the rank of the  $C$  can be used as a qualitative proxy to predict the number of loops in the optimal transport network. Finally, as one could intuitively expect, the basis loop fraction increases with  $Q$ , i.e., with the number of nodes where mass is injected or extracted.

### VI. CONCLUSION AND OUTLOOK

Routing models on networks are relevant to study many real-world problems. While most of the works in the current literature consider stationary setups, i.e., the inflows injected in the network do not change in time [2,15,20,27,37–42], few recent works investigate time-varying loads and the majority of these models study solely the averaged evolution of the networks' variables [3,9,24,37].

In this work we analyzed a dynamical system where the conductivities are regulated by time-varying mass inflows. Motivated by empirical observations [25,43–45], we assumed the existence of auxiliary conductivities that have response times which are much slower than those of the loads. Furthermore, in order to make the problem analytically tractable, we supposed that all the loads injected in nodes are periodic, in a period that is substantially smaller than the adaptation time of the new conductivities. These two hypothesis together allowed us to deduce a dynamics where the evolution of the systems is solely regulated by an input matrix constructed using the Fourier series expansion of the loads.

The resulting dynamics allowed us to derive the main findings of our work. In detail, combining theoretical arguments with empirical evidence on synthetic networks, we found an expression for the long-run solutions of the original



dynamics, which cannot otherwise be obtained by simply solving the original dynamics. These long-run solutions are the sum of stationary components, equal to the asymptotics of the dynamics we constructed, and an oscillatory one. This second contribution can be expressed as the sum of periodic signals, with modes related to those of the loads. Moreover, we discussed a sufficient condition on the loads that determines when optimal transport networks can be loopless. Such a condition was numerically validated on the Bordeaux bus network. Finally, our dynamics can be connected to an optimization setup, as shown by the proposed candidate Lyapunov functional. As a result, asymptotic trajectories of our dynamics minimize the total cost needed to build the network infrastructure.

Importantly, the numerical discretization of the dynamics we proposed in this work can be used as an efficient method to rapidly converge to average long-run solutions of the original dynamics.

Our results can be extended in several ways. For instance, it would be interesting to investigate different types of input loads that relax the periodicity hypothesis and use this to analyze the behavior of the conductivities in different problems' settings. Similarly, it would be interesting to explore how this formalism adapts to multilayer networks, where passengers can enter different stations corresponding to different transportation modes [20]. Another relevant application could be that of integrating our findings with the recent work of Baptista and De Bacco [19], where the authors studied how topological properties of the transport network change in time, as we approach stationary configurations, and how these reflect on the shape of the conductivities.

While our work constitutes a step towards extending the formalism of capacitated networks to time-dependent loads, it is important to remark that our findings are valid in a particular time limit. Specifically, this is the scenario where conductivities slowly evolve with the integral average of periodic forcings, as introduced in Sec. III. It is not clear how the theoretical analysis presented in this work could be adapted to scenarios where loads and conductivities evolve with the *same* timescale. This could be an interesting avenue for future work. Another interesting direction could be that of considering additional constraints on the evolution of the conductivities, which are not currently included in our model. For instance, one could introduce a threshold capacity above which the edge traffic saturates, causing blockage of roads.

Altogether, we believe that our results enrich the current knowledge on network routing problems with time-varying input loads and have immediate practical implications. In fact, our model is deterministic, since there is only one single realization of the inputs and thus adequate to model real-world scenarios where time-dependent loads are measured quantities, e.g., the amount of passengers traveling in a metro (which can be easily tracked), without the need of stochastic formulations that require the introduction of probability distributions that are hard to characterize.

To facilitate practitioners in using our model, we have made the algorithmic implementation publicly available [49].

## ACKNOWLEDGMENT

The authors thank the International Max Planck Research School for Intelligent Systems for supporting A.L.

## APPENDIX A: NONDIMENSIONALIZATION OF THE MODEL

Here we show how our model can be made dimensionless, i.e., constants can be removed by appropriately rescaling dimension-dependent quantities. We start from the dimension-dependent ODEs

$$\frac{d\tilde{\mu}_e(\tilde{t})}{d\tilde{t}} = a \frac{\tilde{F}_e^2(\tilde{t})}{\tilde{\mu}_e^\gamma(\tilde{t})} - b\tilde{\mu}_e(\tilde{t}) \quad \forall e \in E, \quad (\text{A1})$$

with  $a$  and  $b$  coefficients with appropriate dimensions. We then choose the nondimensionalization

$$t := \tilde{t}/t_c, \quad (\text{A2})$$

$$\mu_e := \tilde{\mu}_e/\mu_c \quad \forall e \in E, \quad (\text{A3})$$

$$S_v(t) := \tilde{S}_v(\tilde{t})/S_c \quad \forall v \in V, \quad (\text{A4})$$

where  $S_c$  is the characteristic unit of  $S$ . Substituting Eqs. (A2)–(A4) in Kirchhoff's law yields  $F_e(t) = \tilde{F}_e(\tilde{t})/S_c \forall e \in E$ , with  $F(t)$  adimensional fluxes.

Recasting all adimensional variables in Eq. (A1), we get

$$\frac{d\mu_e(t)}{dt} = a \left( \frac{t_c S_c^2}{\mu_c^{\gamma+1}} \right) \frac{F_e^2(t)}{\mu_e^\gamma(t)} - b t_c \mu_e(t) \quad \forall e \in E, \quad (\text{A5})$$

showing that, to recover Eq. (2), we can set

$$t_c = 1/b, \quad (\text{A6})$$

$$\mu_c^{\gamma+1}/S_c^2 = a/b. \quad (\text{A7})$$

We note that a procedure for the nondimensionalization of a model similar to ours can be found in [2] (Supplemental Material Sec. II).

## APPENDIX B: DERIVATION OF (5)

In order to define Eq. (5), we perform the calculations on the right-hand side of Eq. (2),

$$\int_{\tau}^{\tau+\Delta} \Phi_e(\mu(t), t) dt := \frac{1}{\Delta} \int_{\tau}^{\tau+\Delta} \mu_e^{-\gamma}(t) F_e^2(t) - \mu_e(t) dt \quad (\text{B1})$$

$$= \frac{1}{\Delta} \int_{\tau}^{\tau+\Delta} \left( \frac{\mu_e^{-2-\gamma}(t)}{\ell_e^2} \sum_{uvmm} B_{me} B_{ne} L_{um}^{\dagger}(\mu(t)) L_{vn}^{\dagger}(\mu(t)) \right. \\ \left. \times S_u(t) S_v(t) - \mu_e(t) \right) dt \quad (\text{B2})$$

$$\stackrel{t'=\tau-t}{=} \frac{1}{\Delta} \int_0^{\Delta} \left( \frac{\mu_e^{-2-\gamma}(\tau+t')}{\ell_e^2} \sum_{uvmm} B_{me} B_{ne} L_{um}^{\dagger}(\mu(\tau+t')) \right. \\ \left. \times L_{vn}^{\dagger}(\mu(\tau+t')) S_u(\tau+t') S_v(\tau+t') - \mu_e(\tau+t') \right) dt' \quad (\text{B3})$$

$$\begin{aligned} &\stackrel{\text{(ii)}}{\approx} \frac{\mu_e^{2-\gamma}(\tau)}{\ell_e^2} \sum_{uv} A_{eu}(\mu(\tau)) A_{ev}(\mu(\tau)) \frac{1}{\Delta} \int_0^\Delta S_u(\tau + t') \\ &\quad \times S_v(\tau + t') - \mu_e(\tau + t') dt' \end{aligned} \quad (\text{B4})$$

$$\begin{aligned} &\stackrel{t'=t+\tau}{=} \frac{\mu_e^{2-\gamma}(\tau)}{\ell_e^2} \sum_{uv} A_{eu}(\mu(\tau)) A_{ev}(\mu(\tau)) \frac{1}{\Delta} \int_\tau^{\tau+\Delta} S_u(t) \\ &\quad \times S_v(t) - \mu_e(t) dt \end{aligned} \quad (\text{B5})$$

$$=: \hat{\Phi}_e(\mu(\tau), \tau), \quad (\text{B6})$$

which are valid for all  $e \in E$ . In detail, in Eq. (B2) we used the definition of the fluxes  $F_e(t) := \mu_e(t)[p_u(t) - p_v(t)]/\ell_e \forall e \in E$

$E$  and evaluated the pressure solving Kirchhoff's law, i.e.,  $p_v(t) := \sum_u L_{vu}^\dagger(\mu(t)) S_u(t) \forall v \in V$ . The second important step is in Eq. (B4), where we used hypothesis (ii) in Sec. III A, namely, the approximation in Eq. (4), to carry the conductivities out of the time integral, and we introduced  $A_{ev}(\mu(t)) := \sum_u B_{ev} L_{vu}^\dagger(\mu(t)) \forall e \in E \forall v \in V$ .

### APPENDIX C: DERIVATION OF (9)

We enforce the hypothesis of periodicity of the loads, i.e.,  $S_v(t) = S_v(t + T)$ , with  $T/\Delta \ll 1$ , and we parametrize the integration window  $\Delta$  as  $\Delta = KT$ ,  $K \gg 1$ . This allows us to split the integral in Eq. (9) into two separate contributions. In detail, making the reasonable hypothesis that  $S_u(t) S_v(t)$  is bounded by  $M < +\infty$  for all  $t \geq 0$  and for all  $u, v \in V$ , we can write

$$\frac{1}{\Delta} \int_\tau^{\tau+\Delta} S_u(t) S_v(t) dt = \frac{1}{\Delta} \int_\tau^{\tau+\Delta} \sum_{n_u n_v} (c_u^{n_u})^* c_v^{n_v} \exp[i\omega(n_v - n_u)t] dt \quad (\text{C1})$$

$$= \sum_{n_u n_v} (c_u^{n_u})^K c_v^{n_v} \left( \sum_{k=1}^{[K]} \mathcal{I}_k(n_u, n_v) + \mathcal{I}_K(n_u, n_v) \right), \quad (\text{C2})$$

$$\mathcal{I}_k(n_u, n_v) := \frac{1}{\Delta} \int_{\tau+(k-1)T}^{\tau+kT} \exp[i\omega(n_v - n_u)t] dt \quad \forall k = 1, \dots, [K], \quad (\text{C3})$$

$$\mathcal{I}_K(n_u, n_v) := \frac{1}{\Delta} \int_{\tau+KT}^{\tau+KT} \exp[i\omega(n_v - n_u)t] dt. \quad (\text{C4})$$

Hence, we separate the first  $[K]$  integrals over the period  $T$  from the last one in  $([K]T, KT)$ . Since  $K \gg 1$ , the first  $[K]$  contributions can be evaluated as

$$\sum_{n_u n_v} (c_u^{n_u})^* c_v^{n_v} \sum_{k=1}^{[K]} \mathcal{I}_k(n_u, n_v) = \frac{[K]}{K} \sum_{n_u n_v} (c_u^{n_u})^* c_v^{n_v} \delta_{n_u n_v} \quad (\text{C5})$$

$$= \sum_{n_c} (c_u^{n_c})^* c_v^{n_c} + O(\Delta), \quad (\text{C6})$$

with  $\delta_{ij}$  being the Kronecker delta for two indices  $i$  and  $j$ . As for the second term, in the limit  $K \gg 1$  we can write

$$\left| \sum_{n_u n_v} (c_u^{n_u})^* c_v^{n_v} \mathcal{I}_K(n_u, n_v) \right| \leq \frac{K - [K]}{K} M \sim O(\Delta), \quad (\text{C7})$$

showing that integrals over the small interval  $([K]T, KT)$  are negligible for a large integration window.

### APPENDIX D: SUFFICIENT CONDITION ON THE RANK FOR OPTIMAL TREES

We discuss in detail the sufficient condition  $\text{rank}(C) = 1$  to obtain loopless optimal networks running the dynamics (10) and (11). Our argument proceed as follows.

The matrix  $C$  is symmetric by construction; thus if its rank is 1 its eigenvalue decomposition is of the form  $C =$

$\sum_{i=1}^N \lambda_i x_i x_i^\top$ , with all the eigenvalues equal to zero except one. We conventionally choose it to be  $\lambda_1 = \sum_v C_{vv} > 0$ , with a unit norm eigenvector  $x_1$ . Defining  $y := \sqrt{\lambda_1} x_1$  and substituting the eigendecomposition of  $C$  in Eq. (12), we get that  $\hat{\Phi}_e$  is proportional to  $\hat{F}_e := (\bar{\mu}_e/\ell_e) \sum_v B_{ev} \hat{p}_v$ , with  $\hat{p}_v := \sum_u L_{vu}^\dagger(\bar{\mu}) y_u$ . In order to conclude, we need to show that  $\hat{p}$  is a well-defined solution of Kirchhoff's law  $\sum_u L_{uv}(\bar{\mu}) \hat{p}_u = y_v$ , i.e.,  $y$  is a zero-sum vector [47]. This comes as a consequence of conservation of mass. Indeed, since for all times  $\sum_v S_v(t) = 0$  holds, we have  $\sum_v S_u(t) S_v(t) = 0 \forall u \in V$ . Using Eq. (9) and ignoring negligible terms of  $O(\Delta)$ , this yields  $\sum_v C_{uv} = 0 \forall u \in V$ . Finally, substituting the eigendecomposition of  $C$  in this last relation gives  $\sum_v y_u y_v = 0 \forall u \in V$ . This is satisfied only if  $\sum_v y_v = 0$ , i.e.,  $y$  is a zero-sum vector. In this case, Eqs. (10) and (11) correspond to the standard dynamics (1)–(3) with constant loads, which are  $S(t) = y \forall t \geq 0$ , and we recover the well-known result that optimal networks are trees for  $\gamma \leq 1$  [23,26,29].

Noticeably, if  $\text{rank}(C) = 1$ , it is possible to prove that the functional  $\tilde{\mathcal{L}}_\gamma(\bar{\mu})$  proposed in Eq. (16) is a well-defined Lyapunov functional. This means that for any  $\bar{\mu}(\tau)$  solution trajectory of Eqs. (10) and (11), we have  $d\tilde{\mathcal{L}}_\gamma(\bar{\mu}(\tau))/d\tau \leq 0$ , with stationarity achieved only by asymptotics of the dynamics. Having established that  $\hat{p}$  is a well-defined potential, we can write the Lyapunov functional as  $\tilde{\mathcal{L}}_\gamma(\bar{\mu}) = (1/2) \sum_v \hat{p}_v(\bar{\mu}) S_v + (1/2\gamma) \sum_e \ell_e \bar{\mu}_e^\gamma$ . This last expression is useful to conclude the proof, which follows that in Ref. [27].

**APPENDIX E: DERIVATION OF (15)**

In order to discern the nature of the fast oscillating component  $b_e(t)$  of the stabilized solutions, we need to investigate further the original dynamics Eqs. (1)–(3). From our numerical validation we observe that the fluxes start to oscillate around a constant value after a first stabilization time interval [see Fig. 3(a)], analogously to the conductivities. This suggests the ansatz  $F_e(t) = \sum_{n_e \in \mathcal{Z}} F_e^{n_e} \exp(i\omega n_e t) \forall e \in E$  for all times  $t$  sufficiently larger than  $t_{\text{STAB}}$  and with the terms  $F_e^{n_e}$  amplitudes of the Fourier series decomposition. We argue that pairing this expression with Kirchhoff's law, i.e.,  $\sum_e B_{ve} F_e(t) = S_v(t)$ , yields

$$F_e(t) = \sum_{n_e \in \mathcal{N}} F_e^n \exp(i\omega n t) \quad \forall e \in E, \quad (\text{E1})$$

with  $\mathcal{N} := \{n_v\}$  the set of Fourier modes of the loads injected in the network. Our argument is the following.

Assuming the ansatz  $F_e(t) = \sum_{n_e \in \mathcal{Z}} F_e^{n_e} \exp(i\omega n_e t) \forall e \in E$ , we separate the contributions

$$F_e(t) = \varphi_e(t) + \psi_e(t), \quad (\text{E2})$$

$$\varphi_e(t) = \sum_{n_e \in \mathcal{N}} F_e^{n_e} \exp(i\omega n_e t), \quad (\text{E3})$$

$$\psi_e(t) = \sum_{n_e \notin \mathcal{N}} F_e^{n_e} \exp(i\omega n_e t). \quad (\text{E4})$$

Substituting Eqs. (E2)–(E4) in Kirchhoff's law returns the conditions

$$\sum_e B_{ve} \varphi_e(t) = S_v(t), \quad (\text{E5})$$

$$\sum_e B_{ve} \psi_e(t) = 0, \quad (\text{E6})$$

valid for all  $v \in V$ . Now, in order to guarantee that the fluxes  $\{\varphi_e(t), \psi_e(t)\}$  are well defined, we suppose the existence of two time-dependent potentials  $\alpha(t) = \{\alpha_v(t)\}$  and  $\beta(t) = \{\beta_v(t)\}$ . These are defined on the network nodes and such that for all  $e \in E$  we have

$$\varphi_e(t) := \frac{\mu_e}{\ell_e} \sum_v B_{ve} \alpha_v(t), \quad (\text{E7})$$

$$\psi_e(t) := \frac{\mu_e}{\ell_e} \sum_v B_{ve} \beta_v(t). \quad (\text{E8})$$

Note that these definitions lead to  $F_e(t)$  being a potential-based flux and yield  $p_v(t) = \alpha_v(t) + \beta_v(t) \forall v \in V$ . Substituting Eqs. (E7) and (E8) in Eqs. (E5) and (E6), respectively, implies that  $\psi_e(t) = 0 \forall e \in E$  and for sufficiently large times. Hence, the only nonzero terms in the Fourier decomposition of  $F_e(t)$  have modes in  $\mathcal{N}$ .

This result is particularly useful to describe the behavior of  $\mu(t)$  at large times. First, we recall that  $\mu_e(t) = \bar{\mu}_e^\infty + b_e(t) \forall e \in E$ , as discussed in Sec. IV. Moreover, we observe that in our numerical experiments [see Fig. 3(a)] the size of the amplitude of the oscillatory term  $b_e(t)$  is negligible in size with respect to  $\bar{\mu}_e^\infty$ , unless  $\mu_e(t)$  decays to zeros. This allows us to approximate Eq. (2) as

$$\frac{d\mu_e(t)}{dt} \simeq \frac{F_e^2(t)}{(\bar{\mu}_e^\infty)^\gamma} - \mu_e(t) \quad \forall e \in E. \quad (\text{E9})$$

Finally, substituting Eq. (E1) in Eq. (E9), we get the desired results, i.e., the main oscillatory modes of the conductivities, hence of  $b_e(t)$ , are resonant with the squared fluxes. Thus we obtain Eq. (15).

**Validation on synthetic networks**

We test these expressions numerically on networks generated as described in Sec. IV. We compute  $P_e := \int_{\mathbb{R}} |\mathcal{F}[b_e](f)|^2 df$ , the total spectral density of the oscillatory components  $b_e(t)$ , after the conductivities  $\mu(t)$  stabilize. Here  $\mathcal{F}[\cdot](f)$  is the Fourier transform operator. Additionally, we calculate  $P_{\mathcal{N}}$ , obtained summing the atomic contributions of the spectral density on the modes  $k \in \mathcal{K} := \{k \text{ s.t. } k = n + m \text{ for } n, m \in \mathcal{N}\}$ , namely,  $P_{\mathcal{N},e} := \sum_{k \in \mathcal{K}} \int_{\mathbb{R}} |\mathcal{F}[b_e](f)|^2 \delta(f - k) df \forall e \in E$ .

From Eq. (15) we expect to have most of the spectral density of  $b_e(t)$  concentrated on the modes in  $\mathcal{K}$ , i.e., the ratio  $P_e/P_{\mathcal{N},e}$  should be close to 1 for each edge. In Fig. 5(a) we plot  $P = \{P_e\}$  versus  $P_{\mathcal{N}} = \{P_{\mathcal{N},e}\}$  for the example network considered in Fig. 3. The plot supports Eq. (15); indeed, the elementwise ratio  $P/P_{\mathcal{N}}$  is close to 1 for all points (each correspondent to a different edge) with a slight deviation only for small (thus negligible) values of the conductivities.

We further validate this result on an additional synthetic example network. We construct the Delaunay networks described in Sec. IV considering 100 combinations of seeds for the nodes' positions and for the random input loads. Then we compute the spectral densities  $P$  and plot them against  $\delta P$ , with entries  $\delta P_e := (P_{\mathcal{N},e} - P_e)/P_e$ . We show in Fig. 5(b) results for  $\gamma = 0.5, 1, 1.5$  on 100 random graphs of size  $|V| = 8$ . Here we clearly see that  $\delta P$  are negligible for any edge with  $P$  larger than a threshold  $\alpha$  (in our experiments we set  $\alpha = 10^{-3}$ ), further supporting the result in Fig. 5(a).

It is worth mentioning how the points cluster in different regions of the plot for different values of  $\gamma$ . The green points, corresponding to  $\gamma = 0.5$ , are divided into two clusters: one around  $P$  small and  $\delta P = 1$  and another with  $P$  large and  $\delta P$  negligible. This reflects the tendency of  $\gamma < 0.5$  to aggregate fluxes on few edges. The blue points, corresponding to  $\gamma = 1.5$ , are instead concentrated around a region with  $P$  large and  $\delta P$  small, since in this case fluxes are distributed on more edges. Finally, the orange points, corresponding to  $\gamma = 1$ , represent a transition between the two cases and are located in a cluster placed in between the other two. This result is consistent with the behavior of  $\gamma$  mentioned in Sec. II.

We test the scalability of our result by running the same validation just described, but increasing the graphs sizes. We plot our results in Fig. 5(c). Here we show the compatibility of  $\delta P_e := \sum_e \delta P_e \mathbb{I}(\delta P_e > \alpha)/E'$  with zero. Here  $\mathbb{I}(\cdot)$  is the indicator function and  $E'$  the number of the edges that do not get trimmed by  $\alpha$ . We see that all values attain values close to  $\delta P_e = 0$  and all error bars (expressing standard deviations over 100 random graph realizations) are always intersecting the line highlighting  $\delta P_e = 0$ . The decreasing trend of  $\delta P_e$  for  $\gamma = 0.5$  can be attributed to the fact that we fixed the cutoff threshold  $\alpha$  a priori and thus we do not have a precise trim for  $P$  for larger networks.

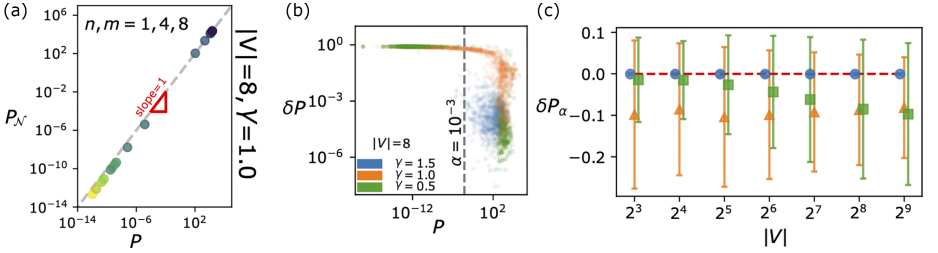


FIG. 5. Spectral density validation. (a) Plot of  $P_N$  versus  $P$  for the example network of Fig. 3. Each point corresponds to an edge; the color scale is that of Fig. 3(a). (b) Plot of  $P$  versus  $\delta P$ . Each point corresponds to an edge; marker color denotes  $\gamma = 0.5, 1, 1.5$ . The dashed line is the cutoff used to build  $\delta P_\alpha$ . (c) Compatibility of  $\delta P_\alpha$  with  $\delta P_\alpha = 0$  for different networks' sizes. Markers and bars correspond to averages and standard deviations over 100 random configurations of the problem, respectively. The networks have been obtained pairing ten seeds for node coordinate generation and ten seeds for mass and conductivity initialization  $\mu_e \sim U(0, 1)$ .

#### APPENDIX F: HARMONIC-OSCILLATOR CONDITIONS

We already established that if  $\text{rank}(C) = 1$ , then there exists a zero-sum vector  $y$  such that  $C_{uv} = y_u y_v \forall u, v \in V$  (see Appendix D). Inspecting the diagonal elements of  $C$ , it is immediate to get  $y_v = \pm \sqrt{C_{vv}} \forall v \in V$ . Here the choice of the plus or minus sign is constrained among  $2^{|V|}$  possibilities, to those for which  $\sum_u y_u = 0$  holds. The right-to-left implication comes naturally from the definition of  $C$ . Namely, if we suppose that  $C_{uv} = y_u y_v \forall u, v \in V$ , we are imposing that all the columns of  $C$  are scalar multipliers of each other, i.e.,  $\text{rank}(C) = 1$ .

Substituting Eq. (19) in  $v_v = \lambda v_u$  leads to

$$A_v^i \exp(i\phi_v^i) \delta_{n_u, n_v} = \lambda A_u^i \exp(i\phi_u^i) \delta_{n_u, n_v}, \quad (\text{F1})$$

which needs to be satisfied for each pair of  $n_u, n_v \in \mathbb{N}$ . This is valid if the phases are such that  $\phi_v^i = \phi_u^i + k\pi$ ,  $k \in \mathbb{Z}$ , i.e., condition (i) in Sec. V A holds. Substituting this last equality in Eq. (F1), we get

$$A_v^i \exp(i\phi_v^i) \delta_{n_u, n_v} = \lambda A_u^i \exp(i\phi_u^i) (-1)^k \delta_{n_u, n_v}, \quad (\text{F2})$$

$$A_v^i \delta_{n_u, n_v} = \lambda A_u^i (-1)^k \delta_{n_u, n_v}, \quad (\text{F3})$$

which is precisely (ii) in Sec. V A. In conclusion, fixing the input loads in such a way that (i) and (ii) hold leads to  $\text{rank}(C) = 1$ , which is sufficient to get optimal tree topologies, as shown in Appendix D.

- 
- [1] H. Ronellenfitch and E. Katifori, *Phys. Rev. Lett.* **117**, 138301 (2016).
- [2] H. Ronellenfitch and E. Katifori, *Phys. Rev. Lett.* **123**, 248101 (2019).
- [3] E. Katifori, G. J. Szöllősi, and M. O. Magnasco, *Phys. Rev. Lett.* **104**, 048704 (2010).
- [4] Q. Xia, *ESAIM: Contr. Optim. Calc. Var.* **13**, 359 (2007).
- [5] K. Sinclair and R. C. Ball, *Phys. Rev. Lett.* **76**, 3360 (1996).
- [6] A. Rinaldo, I. Rodriguez-Iturbe, R. Rigon, R. L. Bras, E. Ijjasz-Vasquez, and A. Marani, *Water Resour. Res.* **28**, 2183 (1992).
- [7] A. Rinaldo, I. Rodriguez-Iturbe, R. Rigon, E. Ijjasz-Vasquez, and R. L. Bras, *Phys. Rev. Lett.* **70**, 822 (1993).
- [8] T. Sun, P. Meakin, and T. Jøssang, *Phys. Rev. E* **49**, 4865 (1994).
- [9] A. Konkol, J. Schwenk, E. Katifori, and J. B. Shaw, *Geophys. Res. Lett.* **49**, e2022GL098284 (2022).
- [10] A. Tero, S. Takagi, T. Saigusa, K. Ito, D. P. Bebber, M. D. Fricker, K. Yumiki, R. Kobayashi, and T. Nakagaki, *Science* **327**, 439 (2010).
- [11] A. Tero, K. Yumiki, R. Kobayashi, T. Saigusa, and T. Nakagaki, *Theory Biosci.* **127**, 89 (2008).
- [12] A. Tero, R. Kobayashi, and T. Nakagaki, *Physica A* **363**, 115 (2006).
- [13] A. Tero, R. Kobayashi, and T. Nakagaki, *J. Theor. Biol.* **244**, 553 (2007).
- [14] T. Nakagaki, H. Yamada, and Á. Tóth, *Nature (London)* **407**, 470 (2000).
- [15] V. Bonifaci, *Inf. Process. Lett.* **113**, 4 (2013).
- [16] V. Bonifaci, *J. Math. Biol.* **74**, 567 (2017).
- [17] D. Baptista, D. Leite, E. Facca, M. Putti, and C. De Bacco, *Sci. Rep.* **10**, 20806 (2020).
- [18] D. Baptista and C. De Bacco, *R. Soc. Open Sci.* **8**, 210025 (2021).
- [19] D. Baptista and C. De Bacco, in *Complex Networks & Their Applications X*, edited by R. M. Benito, C. Cherifi, H. Cherifi, E. Moro, L. M. Rocha, and M. Sales-Pardo (Springer International, Cham, 2022), pp. 578–592.
- [20] A. A. Ibrahim, A. Lonardi, and C. De Bacco, *Algorithms* **14**, 189 (2021).
- [21] V. Bonifaci, E. Facca, F. Folz, A. Karrenbauer, P. Kolev, K. Mehlhorn, G. Morigi, G. Shahkarami, and Q. Vermande, *Theor. Comput. Sci.* **920**, 1 (2022).

- [22] C. H. Yeung, D. Saad, and K. Y. M. Wong, *Proc. Natl. Acad. Sci. USA* **110**, 13717 (2013).
- [23] S. Bohn and M. O. Magnasco, *Phys. Rev. Lett.* **98**, 088702 (2007).
- [24] F. Corson, *Phys. Rev. Lett.* **104**, 048703 (2010).
- [25] D. Hu, D. Cai, and A. V. Rangan, *PLoS ONE* **7**, e45444 (2012).
- [26] J. B. Kirkegaard and K. Sneppen, *Phys. Rev. Lett.* **124**, 208101 (2020).
- [27] A. Lonardi, E. Facca, M. Putti, and C. De Bacco, *Phys. Rev. Res.* **3**, 043010 (2021).
- [28] A. Lonardi, M. Putti, and C. De Bacco, *Sci. Rep.* **12**, 7474 (2022).
- [29] J. R. Banavar, F. Colaiori, A. Flammini, A. Maritan, and A. Rinaldo, *Phys. Rev. Lett.* **84**, 4745 (2000).
- [30] Q. Xia, *Commun. Contemp. Math.* **05**, 251 (2003).
- [31] M. Mezard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, Oxford, 2009).
- [32] C. H. Yeung and D. Saad, *Phys. Rev. Lett.* **108**, 208701 (2012).
- [33] C. H. Yeung and D. Saad, *J. Phys. A: Math. Theor.* **46**, 103001 (2013).
- [34] F. Altarelli, A. Braunstein, L. Dall'Asta, C. De Bacco, and S. Franz, *PLoS ONE* **10**, e0145222 (2016).
- [35] C. D. Bacco, S. Franz, D. Saad, and C. H. Yeung, *J. Stat. Mech.* (2014) P07009.
- [36] Y.-Z. Xu, H. F. Po, C. H. Yeung, and D. Saad, *Phys. Rev. E* **105**, 044316 (2022).
- [37] D. Hu and D. Cai, *Phys. Rev. Lett.* **111**, 138701 (2013).
- [38] V. Bonifaci, K. Mehlhorn, and G. Varma, *J. Theor. Biol.* **309**, 121 (2012).
- [39] E. Facca, F. Cardin, and M. Putti, *SIAM J. Appl. Math.* **78**, 651 (2018).
- [40] E. Facca, S. Daneri, F. Cardin, and M. Putti, *J. Sci. Comput.* **82**, 68 (2020).
- [41] E. Facca, F. Cardin, and M. Putti, *J. Comput. Phys.* **447**, 110700 (2021).
- [42] E. Facca, F. Cardin, and M. Putti, [arXiv:1812.11782](https://arxiv.org/abs/1812.11782).
- [43] B. Folkow, O. Lundgren, and I. Wallentin, *Acta Physiol. Scand.* **57**, 270 (1963).
- [44] D. N. Granger, P. R. Kvietyts, and M. A. Perry, *Am. J. Physiol.* **242**, G570 (1982).
- [45] R. J. Widmer, R. H. Stewart, M. F. Young, J. E. Laurinec, G. A. Laine, and C. M. Quick, *Am. J. Physiol.* **292**, R2312 (2007).
- [46] Transport for London open data, <https://tfl.gov.uk/info-for/open-data-users/> (accessed 20 May 2021).
- [47] D. J. Klein and M. Randić, *J. Math. Chem.* **12**, 81 (1993).
- [48] R. Kujala, C. Weckström, R. K. Darst, M. N. Mladenović, and J. Saramäki, *Sci. Data* **5**, 180089 (2018).
- [49] N-STARK: OS code implementation, <https://github.com/aleable/N-STARK>.



## OPEN ACCESS

EDITED BY  
Adriano Tiribocchi,  
National Research Council (CNR), Italy

REVIEWED BY  
Giovanni Franzina,  
Istituto per le Applicazioni del Calcolo  
(IAC), Italy  
Pablo Villegas,  
Enrico Fermi Center for Study and  
Research, Italy

\*CORRESPONDENCE  
Alessandro Lonardi,  
✉ alessandro.lonardi@tuebingen.mpg.de  
Diego Baptista,  
✉ diego.theuerkauf@tuebingen.mpg.de

<sup>†</sup>These authors have contributed equally to  
this work and share first authorship

SPECIALTY SECTION  
This article was submitted to Complex  
Systems, a section of the journal  
Frontiers in Physics

RECEIVED 03 November 2022  
ACCEPTED 20 January 2023  
PUBLISHED 27 February 2023

CITATION  
Lonardi A, Baptista D and De Bacco C  
(2023), Immiscible color flows in optimal  
transport networks for  
image classification.  
*Front. Phys.* 11:1089114.  
doi: 10.3389/fphy.2023.1089114

COPYRIGHT  
© 2023 Lonardi, Baptista and De Bacco.  
This is an open-access article distributed  
under the terms of the Creative Commons  
Attribution License (CC BY). The use,  
distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in  
this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Immiscible color flows in optimal transport networks for image classification

Alessandro Lonardi<sup>\*†</sup>, Diego Baptista<sup>\*†</sup> and Caterina De Bacco

Physics for Inference and Optimization Group, Max Planck Institute for Intelligent Systems, Cyber Valley, Tübingen, Germany

In classification tasks, it is crucial to meaningfully exploit the information contained in the data. While much of the work in addressing these tasks is focused on building complex algorithmic infrastructures to process inputs in a black-box fashion, little is known about how to exploit the various facets of the data before inputting this into an algorithm. Here, we focus on this latter perspective by proposing a physics-inspired dynamical system that adapts optimal transport principles to effectively leverage color distributions of images. Our dynamics regulates immiscible fluxes of colors traveling on a network built from images. Instead of aggregating colors together, it treats them as different commodities that interact with a shared capacity on the edges. The resulting optimal flows can then be fed into standard classifiers to distinguish images in different classes. We show how our method can outperform competing approaches on image classification tasks in datasets where color information matters.

## KEYWORDS

network flow optimization, image classification, network optimization, optimal transport, self-adapting dynamical systems

## 1 Introduction

Optimal transport (OT) is a powerful method for computing the distance between two data distributions. This problem has a cross-disciplinary domain of applications, ranging from logistics and route optimization [1–3] to biology [4, 5] and computer vision [6–10], among others. Within this broad variety of problems, OT is largely utilized in machine learning [11] and deployed for solving classification tasks, where the goal is to optimally match discrete distributions that are typically learned from data. Relevant usage examples are also found in multiple fields of physics, as in protein fold recognition [12], stochastic thermodynamics [13], designing transportation networks [14, 15], routing in multilayer networks [16], or general relativity [17]. A prominent application is image classification [18–23], where the goal is to measure the similarity between two images. OT solves this problem by interpreting image pairs as two discrete distributions and then assessing their similarity *via* the Wasserstein ( $W_1$ ) distance ([24], Definition 6.1), a measure obtained by minimizing the cost needed to transform one distribution into the other. Using  $W_1$  for image classification carries many advantages over other similarity measures between histograms. For example,  $W_1$  preserves all properties of a metric [9, 24], it is robust over domain shift for train and test data [22], and it provides meaningful gradients to learn data distributions on non-overlapping domains [25]. Because of these and several other desirable properties, much research effort has been put into speeding up algorithms to calculate  $W_1$  [12, 19, 20, 26, 27]. However, all these methods overlook the potential of effectively using image colors directly in the OT formulation. As a result, practitioners have access to increasingly efficient algorithms, but those do not necessarily

improve accuracy in predictions, as we lack a framework that fully exploits the richness of the input information.

Colored images originally encoded as three-dimensional histograms—with one dimension per color channel—are often compressed into lower dimensional data using feature extraction algorithms [9, 23]. Here, we propose a different approach that maps the three distinct color histograms to multicommodity flows transported in a network built using images' pixels. We combine recent developments in OT with the physics insights of capacitated network models [1, 5, 28–31] to treat colors as masses of different types that flow through the edges of a network. Different flows are coupled together with a shared conductivity to minimize a unique cost function. This setup is reminiscent of the distinction between modeling the flow of one substance, e.g., water, and modeling the flows of multiple substances that do not mix, e.g., immiscible fluids, which share the same network infrastructure. By virtue of this multicommodity treatment, we achieve stronger classification performance than state-of-the-art OT-based algorithms in real datasets where color information matters.

## 2 Problem formulation

### 2.1 Unicommodity optimal transport

Given two  $m$ - and  $n$ -dimensional probability vectors  $g$  and  $h$  and a positive-valued ground cost matrix  $C$ , the goal of a standard—unicommodity—OT problem is to find an optimal transport path  $P^*$  satisfying the conservation constraints  $\sum_j P_{ij} = g_i \forall i$  and  $\sum_i P_{ij} = h_j \forall j$ , while minimizing  $J(g, h) = \sum_{ij} P_{ij} C_{ij}$ .

Entries  $P_{ij}^*$  can be interpreted as the mass transported from  $g_i$  to  $h_j$  when paying a cost  $C_{ij}$ , while  $J^*$ , i.e.,  $J$  evaluated at  $P^*$ , encodes the minimum effort needed to transport  $g$  to  $h$ . Notably, if all entries  $C_{ij}$  are distances between  $i$  and  $j$ , then  $J^*$  is the  $W_1$  distance between  $g$  and  $h$  (see [24] for a standard proof and [9] for derivations focusing on the discrete case).

### 2.2 Physics-inspired multicommodity optimal transport

Interpreting colors as masses traveling along a network built from images' pixels (as we define in detail below), unicommodity OT could be used to capture the similarity between grayscale images. However, it may not be ideal for colored images, when color information matters. The limitation of unicommodity OT in Section 2.1 is that it does not fully capture the variety of information contained in different color channels as it is not able to distinguish them. Motivated by this, we tackle this challenge and move beyond this standard setting by incorporating insights from the dynamics of immiscible flows into physics. Specifically, we treat the different pixels' color channels as masses of different types that do not mix but rather travel and interact on the same network infrastructure, while optimizing a unique cost function. By assuming capacitated edges with conductivities that are proportional to the amount of mass traveling through an edge, we can define a set of ODEs that regulate fluxes and conductivities. These are optimally distributed along a network to better account for color information while satisfying physical conservation laws. Similar ideas

have been successfully used to route different types of passengers in transportation networks [2, 16, 32].

Formally, we couple together the histograms of  $M = 3$  color channels, the *commodities*, indexed with  $a = 1, \dots, M$ . We define  $g^a$  and  $h^a$  as  $m$ - and  $n$ -dimensional probability vectors of mass of type  $a$ . More compactly, we define the matrix  $G$  with entries  $G_{ia} = g_i^a$  (respectively,  $H$  for  $h$ ), each containing the intensity of color channel  $a$  in pixel  $i$  of the first (respectively, second) image. These regulate the sources and sinks of mass in our setting. We then enforce the conservation of mass for each commodity index  $a$   $\sum_i g_i^a = \sum_j h_j^a$ . This ensures that all the color mass in the first image is accounted for in the second image, and vice versa. This should be valid for each mass type.

Moreover, we define the set  $\Pi(G, H)$  containing  $(m \times n \times M)$ -dimensional tensors  $P$  with entries  $P_{ij}^a$  being transport paths between  $g^a$  and  $h^a$ . These regulate how fluxes of colors of different types travel along a network. We enforce the interaction between transport paths for different commodities by introducing a *shared cost*.

$$J_\Gamma(G, H) = \sum_{ij} \|P_{ij}\|_2^\Gamma C_{ij}, \tag{1}$$

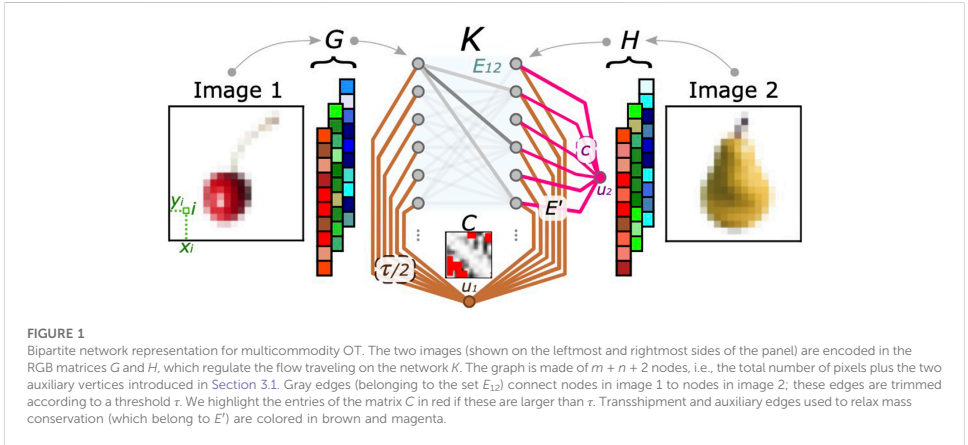
where  $\|P_{ij}\|_2 = (\sum_a P_{ij}^a)^{1/2}$  is the 2-norm of the vector  $P_{ij} = (P_{ij}^1, \dots, P_{ij}^M)$  and  $0 < \Gamma < 4/3$  is a regularization parameter. We take  $\Gamma > 0$  since a negative exponent would favor the proliferation of loops with infinite mass [28]. Instead, we conventionally consider  $\Gamma < 4/3$  (see Section 3.2) since the cost  $J_\Gamma$  exhibits the same convexity properties for any  $\Gamma > 1$ , i.e., it is strictly convex, and OT paths do not change substantially with  $\Gamma$  in this regime [2]. We can thus formulate its corresponding multicommodity OT problem as that of finding a tensor  $P^*$  solution of

$$J_\Gamma^*(G, H) = \min_{P \in \Pi(G, H)} J_\Gamma(G, H). \tag{2}$$

It should be noted that for  $M = 1$  and  $\Gamma = 1$ , we recover the standard unicommodity OT setup.

The problem in Eq. 2 admits a precise physical interpretation. In fact, it can be recast as a constrained minimization problem with the objective function being the energy dissipated by the multicommodity flows (Joule's law) and a constant total conductivity. Furthermore, transport paths follow Kirchhoff's law enforcing conservation of mass [2, 32, 33] (see [Supplementary Material](#) for a detailed discussion).

Noticeably,  $J_\Gamma$  is a quantity that takes into account all the different mass types, and the OT paths  $P^*$  are found through a unique optimization problem. We emphasize that this is fundamentally different from solving  $M$ -independent unicommodity problems, where different types of mass are not coupled together as in our setting, and then combining their optimal costs to estimate images' similarity. Estimating  $J_\Gamma^*(G, H)$  directly gives a quantitative and principled measure of the similarity between two images  $G$  and  $H$ . The lower this cost, the higher the similarity of the two images. While this is valid also for the unicommodity cost in Section 2.1, the difference here is that we account differently for the color information as we distinguish different colors *via* the  $M$ -dimensional vector  $P_{ij}$ . The cost in Eq. 2 then properly couples colors by following physical laws regulating immiscible flows. The idea is that if this information matters for the given classification task, incorporating it into the minimization problem would output a cost that helps to distinguish images better, e.g., with higher accuracy.



### 3 Materials and methods

#### 3.1 Optimal transport network on images

Having introduced the main ideas and intuitions, we now explain in detail how to adapt the OT formalism to images. Specifically, we introduce an auxiliary bipartite network  $K_{m,n}(V_1, V_2, E_{12})$ , which is the first building block of the network where the OT problem is solved. A visual representation of this is shown in Figure 1. The images 1 and 2 are represented as matrices ( $G$  and  $H$ ) of sizes  $m \times M$  and  $n \times M$ , respectively, where  $M$  is the number of color channels of the images ( $M = 3$  in our examples). The sets of nodes  $V_1$  and  $V_2$  of the network  $K_{m,n}$  are the pixels of images 1 and 2, respectively. The set of edges  $E_{12}$  contains a subset of all pixel pairs between the two images, as detailed further. We consider the cost of an edge  $(i, j)$  as

$$C_{ij}(\theta, \tau) = \min\{(1 - \theta)\|v_i - v_j\|_2 + \theta\|G_i - H_j\|_1, \tau\}, \quad (3)$$

where the vector  $v_i = (x_i, y_i)$  contains the horizontal and vertical coordinates of pixel  $i$  of image 1 (similarly  $v_j$  for image 2). The quantity  $\theta \in [0, 1]$  is a hyperparameter that is given in input and can be chosen with cross-validation. It acts as a weight for a convex combination between the Euclidean distance between pixels and the difference in their color intensities, following the intuition in [9, 23]. When  $\theta = 0$ , the OT path  $P^*$  is the one that minimizes only the geometrical distance between pixels. Instead, when  $\theta = 1$ , pixels' locations are no longer considered, and transport paths are only weighted by color distributions. The parameter  $\tau$  is introduced following [22, 23] with the scope of removing all edges with cost  $C_{ij}(\theta, \tau) = \tau$ , i.e., those for which  $(1 - \theta)\|v_i - v_j\|_2 + \theta\|G_i - H_j\|_1 > \tau$ . These are substituted by  $m + n$  transshipment edges  $e \in E'$ , each of which has a cost of  $\tau/2$  and is connected to one unique auxiliary vertex  $u_1$ . Thresholding the cost decreases significantly the computational complexity of OT, making it linear with the number of nodes  $|V_1| + |V_2| + 2 = m + n + 2$  (see Supplementary Material).

Furthermore, we relax the conservation of mass by allowing  $\sum_i G_{ia} \neq \sum_j H_{ja}$ . The excess mass  $m^* = \sum_j H_{ja} - \sum_i G_{ia}$  is assigned to a second

auxiliary node,  $u_2$ . We connect it to the network with  $n$  additional transshipment edges,  $e \in E'$ , each penalizing the total cost by  $c = \max_{ij} C_{ij}/2$ . This construction improves classification when the histograms' total masses largely differ [22]. Intuitively, this can happen when comparing “darker” images against “brighter” images more precisely, when entries of  $g^a$  and  $h^a$  are further apart in the RGB color space.

Overall, we obtain a network  $K$  with nodes  $V = V_1 \cup V_2 \cup \{u_1, u_2\}$  and edges  $E = E_{12} \cup E'$ , i.e., the original bipartite graph  $K_{m,n}$ , together with the auxiliary transshipment links and nodes. It should be noted that in its entirety, the system is isolated, i.e., the total mass is conserved. See Supplementary Material for a detailed description of the OT setup.

Given this auxiliary graph, the OT problem is then solved by injecting the color mass contained in image 1 in nodes  $i \in V_1$ , as specified by  $G$ , and extracting it from nodes  $j \in V_2$  of image 2, as specified by  $H$ . This is carried out by transporting mass using either i) an edge in  $E_{12}$  or ii) a transshipment one in  $E'$ . In the following section, we describe how this problem is solved mathematically.

#### 3.2 Optimizing immiscible color flows: The dynamics

We solve the OT problem by proposing the following ODEs for controlling mass transportation:

$$\sum_{j \in \partial i} L_{ij}[x] \phi_j^a = S_i^a \quad \forall i \in V, \quad a = 1, \dots, M, \quad (4)$$

$$\frac{dx_e}{dt} = x_e^a \frac{\|\phi_i - \phi_j\|_2^2}{C_e^2} - x_e, \quad \forall e = (i, j) \in E, \quad (5)$$

which constitute the pivotal equations of our model. Here, we introduce the *shared conductivities*  $x_e \geq 0$  and define  $S_i^a = G_{ia} - H_{ia}$ , taking values  $S_{u_1}^a = 0$  and  $S_{u_2}^a = m^*$  on the auxiliary nodes. With  $L_{ij}[x] = \sum_e (x_e/C_e) B_{ie} B_{je}$ , we denote the weighted Laplacian of  $K$ , where  $B$  is its signed incidence matrix and  $\partial i$  is the neighborhood of node  $i$ . Lastly,  $\phi_i^a$  is the scalar potential acting on nodes for a given



commodity  $a$ . The least-square solutions of Eq. 4 are  $\phi_e^a[x] = \sum_j L_{ij}^\dagger[x] S_j^a$ , where  $\dagger$  denotes the Moore–Penrose inverse. The critical exponent  $0 < \beta < 2$  [ $\Gamma = 2(2 - \beta)/(3 - \beta)$ ] is a hyperparameter that needs to be chosen before solving Eqs 4 and 5. Depending on the modeling task, its value can be fixed *a priori* (e.g.,  $\beta = 1$  for the shortest path problem [34],  $\beta = 5/3$  for river networks [35], and  $\beta \rightarrow 2^-$  for the Steiner tree problem [36]) or cross-validated as we do here for image classification. The exponent aggregates paths using the principle of economy of scale if  $1 < \beta < 2$ . It dilutes them along the network otherwise, with the goal of reducing traffic congestion. This behavior is a direct consequence of the subadditivity of  $J_\Gamma$  in Eq. 2 for  $\beta > 1$  ( $\Gamma < 1$ ), and, respectively, superadditivity for  $\beta < 1$  ( $\Gamma > 1$ ). It has been theoretically discussed and empirically observed, for example, in [32, 37, 38].

The feedback mechanism of Eq. 5 defines multicommodity fluxes ( $P_e^a$ ) that are admissible for the minimization problem introduced in Eq. 2. Particularly, for color of type  $a$  on edges  $e = (i, j)$ , we couple potentials ( $\phi_e^a$ ) that are the solutions of Eq. 4 and shared conductivities ( $x_e$ ) to define

$$P_e^a(t) = x_e(t) \frac{\phi_i^a[x(t)] - \phi_j^a[x(t)]}{C_e}, \quad \forall e \in E, \quad a = 1, \dots, M. \quad (6)$$

This also highlights another physical interpretation; i.e., by interpreting the  $\phi_e^a$  as pressure potentials, the fluxes are seen to arise from a difference in pressure between two nodes as in hydraulic or electrical networks. Crucially, this allocation is governed by *one unique conductivity* for all commodities, whose dynamics depends on the 2-norm over  $a$  of differences in potentials, as in Eq. 5. In analogy with immiscible flows, this ensures that flows of different types share the same infrastructure, and in practice, it couples them into a unique optimization problem.

In the case of only one commodity ( $M = 1$ ), variants of this dynamics have been used to model transport optimization in various physical systems [1, 5, 29–31].

The salient result of our construction is that the asymptotic trajectories of Eqs 4 and 5 are equivalent to the minimizers of Eq. 2, i.e.,  $\lim_{t \rightarrow \infty} P(t) = P^*$  (see [Supplementary Material](#) for derivations following [32, 33]). Therefore, numerically integrating our dynamics solves the multicommodity OT problem. In other words, this allows us to estimate the optimal cost in Eq. 2 and use that to compute similarities between images. A pseudo-code of the algorithmic implementation is shown in [Algorithm 1](#).

### 3.3 Computational complexity

In principle, our multicommodity method has a computational complexity of order  $O(M|V|^2)$  for complete transport network topologies, i.e., when edges in the transport network  $K$  are assigned to all pixel pairs. Nonetheless, we substantially reduce this complexity to  $O(M|V|)$  by sparsifying the graph with the trimming procedure of [22, 23]. More details are given in [Supplementary Material](#). Empirically, we observe that by running Eqs 4 and 5, most of the entries of  $x$  decay to zero after a few steps, producing a progressively sparser weighted Laplacian  $L[x]$ . This allows for faster computation of the Moore–Penrose inverse  $L^\dagger[x]$  and least-square potentials  $\phi_e^a = \sum_j L_{ij}^\dagger[x] S_j^a$ . A thorough experimental analysis of the convergence properties of the OT dynamics has been carried out in [39].

## 4 Results and discussion

### 4.1 Classification task

We provide empirical evidence that our multicommodity dynamics outperforms competing OT algorithms on classification tasks. As anticipated previously, we use the OT optimal cost  $J_\Gamma^*$  as a measure of similarity between two images and perform supervised classification with a  $k$ -nearest neighbor ( $k$ -NN) classifier as described in [20]. Alternative methods (e.g., SVM as in [19]) could also be used for this task. However, these may require the cost  $J_\Gamma^*$  to satisfy the distance axioms to properly induce a kernel. While it is not straightforward to verify these conditions for the OT cost in Eq. 2, this is not necessary for the  $k$ -NN classifier, which requires looser conditions on  $J_\Gamma^*$ .

We compare the classification accuracy of our model against i) the Sinkhorn algorithm [19, 40] (utilizing the more stable Sinkhorn scheme proposed in [41]); ii) a unicommodity dynamics executed on grayscale images, i.e., with color information compressed into one single commodity ( $M = 1$ ); and iii) the Sinkhorn algorithm on grayscale images. All methods are tested on the following two datasets: the Jena Flowers 30 Dataset (JF30) [42] and the Fruit Dataset (FD) [43]. The first consists of 1,479 images of 30 wild-flowering angiosperms (flowers). Flowers are labeled with their species, and inferring them is the goal of the classification task. The second dataset contains 15 fruit types and 163 images. Here, we want to classify fruit types. The parameters of

- 
- 1: **Input:** Image 1 ( $G \in \mathbb{R}^{m \times M}$ ), Image 2 ( $H \in \mathbb{R}^{n \times M}$ ),  $0 \leq \theta \leq 1$ ,  $\tau \geq 0$ ,  $0 \leq \beta \leq 2$
  - 2: **Initialize:**  $x(0) = \bar{x} > 0$  ▷ e.g.  $\bar{x}_e \sim U(0, 1)$
  - 3: Construct a bipartite network  $K_{m,n}$  between  $G$  and  $H$  ▷ complexity  $O(m \cdot n)$
  - 4: Assign  $C_{ij}(\theta, \tau) = \min\{(1 - \theta) \|v_i - v_j\|_2 + \theta \|G_i - H_j\|_1, \tau\}$  to every edge  $(i, j)$  in  $K_{m,n}$ , as in Eq. (3) ▷ complexity  $O(m + n)$
  - 5: Remove from  $K_{m,n}$  all edges s.t.  $C_{ij} > \tau$
  - 6: Add  $u_1$  to  $K_{m,n}$  and it  $m + n$  auxiliary links, each costing  $\tau/2$
  - 7: Balance mass: add  $u_2$ , with inflowing mass  $m^a = \sum_i H_{ia} - \sum_j G_{ja}$
  - 8: **while** convergence is False **do**
  - 9:     Solve Kirchoff’s law, Eq. (4)  $\rightarrow \phi \in \mathbb{R}^{|V| \times M}$
  - 10:     Update  $x$  with discretization of Eq. (5)
  - 11: **end while**
  - 12: Compute  $P$  as in Eq. (6)
  - 13: **Return:**  $J_\Gamma^*(G, H)$  as in Eq. (2)
- 

**Algorithm 1.** Multicommodity dynamics.

**TABLE 1** Classification task results. With multicommodity, Sinkhorn RGB, unicommodity, and Sinkhorn GS, we label methods on colored images (the first two) and grayscale images (the second two). The optimal parameters in the central columns are selected with a 4-fold cross-validation;  $k$  is the number of nearest neighbors used in the classifier. The rightmost column shows the fraction (in percentage) of correctly classified images. Results are ordered by performance, and we highlight the best ones in bold.

	Algorithm	Hyperparameters					Class accuracy [%] (↑)
		$\theta$	$\tau$	$\beta$	$\varepsilon$	$k$	
JF30	Multicommodity	0.25	0.125	1	—	1	62.2
	Sinkhorn RGB	0.25	0.05	—	100	1	58.4
	Sinkhorn GS	0.25	0.05	—	500	1	54.3
	Unicommodity	0.25	0.125	1.25	—	1	53.6
FD	Multicommodity	0	0.04	1.5	—	2	75.0
	Sinkhorn RGB	0.5	0.06	—	750	1	69.6
	Unicommodity	0	0.06	1.5	—	5	64.3
	Sinkhorn GS	0.25	0.06	—	500	4	60.7

the OT problem setup ( $\theta$  and  $\tau$ ) and regularization parameters ( $\beta$  and  $\varepsilon$ , which enforce the entropic barrier in the Sinkhorn algorithm [19]), have been cross-validated for both datasets (see Section 3 and Section 4 in Supplementary Material). All methods are then tested in their optimal configurations (see Supplementary Material for implementation details).

Classification results are shown in Table 1. In all cases, leveraging colors leads to higher accuracy (about an 8% increase) with respect to classification performed using grayscale images. This signals that in the datasets under consideration, color information is a relevant feature for differentiating image samples. Remarkably, we get a similar increase in performance (about 7%–8%) on both colored datasets when comparing our multicommodity dynamics against the Sinkhorn algorithm. As the two algorithms use the same (colored) input, we can attribute this increment to the effective usage of color that our approach is capable of.

In addition, by analyzing results in more detail, we first observe that on JF30, all methods perform best when  $\theta = 0.25$ , i.e., 25% of the information used to build  $C$  comes from colors. This trend does not recur on the FD, where both dynamics favor  $\theta = 0$  (Euclidean  $C$ ). Hence, our model is able to leverage color information *via* the multicommodity OT dynamical formulation.

Second, on JF30, both dynamics perform best with  $\tau = 0.125$ , contrary to Sinkhorn-based methods that prefer  $\tau = 0.05$ . Thus, Sinkhorn’s classification accuracy is negatively affected both by low  $\tau$ —many edges of the transport network are cut—and by large  $\tau$ —noisy color information is used to build  $C$ . We do not observe this behavior in our model, where trimming fewer edges is advantageous. All optimal values of  $\tau$  are lower on the FD since the color distributions in this dataset are naturally light-tailed (see Supplementary Material).

Lastly, we investigate the interplay between  $\theta$  and  $\beta$ . We notice that  $\theta = 0$  (FD) corresponds to higher  $\beta = 1.5$ . Instead, for larger  $\theta = 0.25$  (JF30), the model prefers lower  $\beta$  ( $\beta = 1$  and  $1.25$  for the multicommodity and unicommodity dynamics, respectively). In the former case ( $\theta = 0$ ,  $C_{ij}$  is the Euclidean distance), the cost is equal to zero for pixels with the same locations. Thus, consolidation of transport paths—large  $\beta$ —is favored on cheap links. As instead, increasing  $\theta$  leads to more edges with comparable costs as colors distribute smoothly over images. In this second scenario, better

performance is achieved with distributed transport paths, i.e., lower  $\beta$  (see Supplementary Material).

## 4.2 Performance in terms of sensitivity

We assess the effectiveness of our method against benchmarks by comparing the sensitivity of our multicommodity dynamics and that of the Sinkhorn algorithm on the colored JF30 dataset. Specifically, we set all algorithm parameters to their best configurations, as shown in Table 1. Then, for each of the 30 classes in JF30, we compute its one-to-all sensitivity, i.e., the true positive rate. This is defined for any class  $c$  as

$$S(c) = \frac{TP(c)}{TP(c) + FN(c)} \tag{7}$$

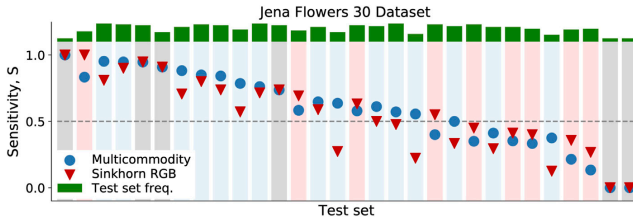
where  $TP(c)$  is the true positive rate, i.e., the number of images in  $c$  that are correctly classified;  $FN(c)$  is the false negative rate, i.e., the number of  $c$ -samples that are assigned a label different from  $c$ . Hence, Eq. 7 returns the probability that a sample is assigned label  $c$ , given that it belongs to  $c$ .

We find that our method robustly outperforms the Sinkhorn algorithm. Specifically, the multicommodity dynamics has the highest sensitivity 50% of the times—15 classes out of a total of 30—as shown in Figure 2. For nine classes, Sinkhorn has higher sensitivity, and for six classes, both methods give the same values of  $S$ . Furthermore, we find that in 2/3 (20 out of 30) of the classes, the multicommodity dynamics returns  $S(c) \geq 1/2$ . This means that our model predicts the correct label more than 50% of the time. In only three out of these 20 cases, Sinkhorn attains higher values of  $S$ , while in most instances where Sinkhorn outperforms our method, it has a lower sensitivity of  $S < 1/2$ . Hence, this is the case in classes where both methods have difficulty distinguishing images.

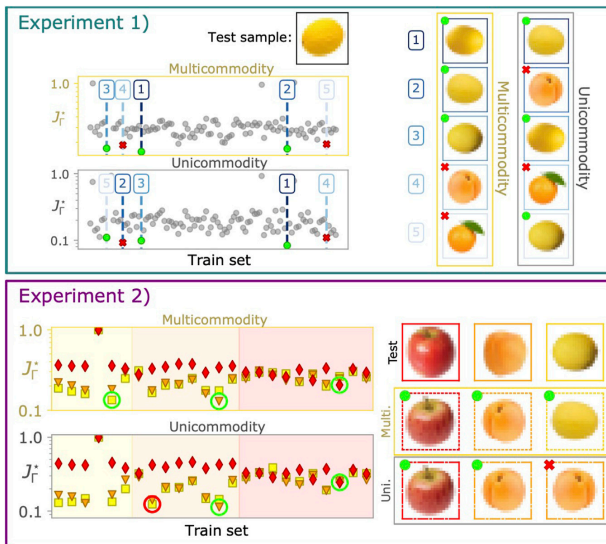
## 4.3 The impact of colors

To further assess the significance of leveraging color information, we conduct three different experiments that highlight both qualitatively and quantitatively various performance differences between the unicommodity and multicommodity approaches. As the two share the same principled dynamics based on OT with the main difference being that multicommodity does not compress the color information, we can use this analysis to better understand how fully exploiting the color information drives better classification.

*Experiment 1: Landscape of optimal cost.* Here, we focus on a qualitative comparison between the cost landscapes obtained with the two approaches. We consider the example of an individual image taken from the FD test set and plot the landscape of optimal costs  $J^*$  when comparing it to the train set. Results for the multicommodity dynamics ( $M = 3$ ) and the unicommodity dynamics ( $M = 1$ ) on grayscale images are shown in Figure 3. Here, we highlight the five lowest values of the cost and mark them in green if they correspond to correctly classified train samples and in red otherwise. At first glance, one may conclude that their performance is identical (as both dynamics classify correctly three samples out of five), and we notice how the multicommodity dynamics consistently clusters them at the bottom of the cost landscape, thus ranking them in a better order. This may explain why the cross-validated best value of  $k$  (the number of nearest neighbors in the  $k$ -NN classifier) is higher for unicommodity methods in this dataset. On a larger sample of data, this results in better overall classification performance, as shown in Table 1.



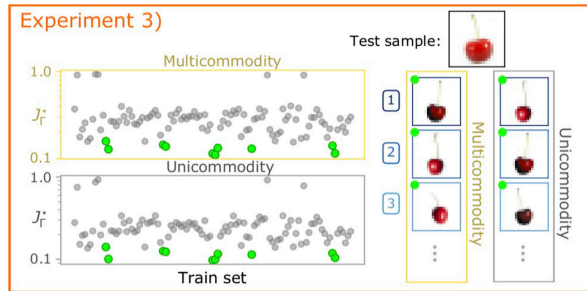
**FIGURE 2**  
Sensitivity on the JF30 dataset. Sensitivity values are shown for the multicommodity dynamics (blue circles) and for Sinkhorn RGB (red triangles). Markers are sorted in descending order of  $S$ , regardless of the method. Background colors are blue, red, and gray, when  $S$  is higher for the multicommodity method, the Sinkhorn algorithm, or none of them, respectively. In green, we plot frequency bars for all classes in the test set.



**FIGURE 3**  
Evaluating the effect of colors. Experiment 1: The top black-framed image is the one to be classified. Predictions given by the multicommodity and unicommodity dynamics (those with lower  $J_r^*$ ) are shown on the right side of the panel and are displayed in a sorted fashion from worst to best (from bottom to top). Experiment 2: The top right samples are the three test images to be classified. Middle and bottom rows are predictions given by the two dynamics. Markers, backgrounds, and test images shared a color code: red for apples, orange for apricots, and yellow for melons. In both panels, green circles and red crosses are used to highlight classified and misclassified images, respectively. All algorithms are executed with their optimal configurations listed in Table 1.

*Experiment 2: Controlling for shape.* We further mark this tendency with a second experiment where we select a subset of the FD composed of images belonging to three classes of fruits that have similar shapes but different colors such as red apples, orange apricots, and yellow melons. As we expect shape to be less informative than colors in this custom set, we can assess the extent to which color plays a crucial role in the classification process. Specifically, the test set is made of three random samples, each drawn from one of these classes (top row of the rightmost panel) in

Figure 3, while the train set contains the remaining instances of the classes. We plot the cost landscape  $J_r^*$  for the train set and draw in the red, orange, and yellow values of  $J_r^*$  that correspond to the samples that are compared against the test apple, apricot, and melon, respectively. We also sort the train samples so that they are grouped in three regions (highlighted by the background color in Figure 3), which correspond to train melons, apricots, and apples. With this construction, if the minimum cost among the yellow markers falls in the yellow region, it



**FIGURE 4** Evaluating the importance of colors: when shapes matter most. Experiment 3: The top black-framed image is the one to be classified. The best three (out of 10) predictions returned by the two dynamics are shown on the right. We mark the training samples belonging to the same class as the test image with green circles. All algorithms are executed with their optimal configurations listed in Table 1.

will correspond to a correctly classified sample (respectively, for orange and red). We further mark the yellow, orange, and red minima in green if the test and train labels correspond, i.e., the marker's and background colors are the same, and in red otherwise. Train and test samples are also in Figure 3. The multicommodity dynamics correctly label each test image. In contrast, unicommodity dynamics fails at this task, labeling a melon as an apricot. This suggests that the multicommodity approach is able to use the color information in datasets where this feature is more informative than others, e.g., shape.

*Experiment 3: When shape matters.* Having shown results on a custom dataset where shape was controlled to matter less, we now do the opposite and select a dataset where this feature should be more informative. The goal is to assess whether a multicommodity approach helps in this case as well, as its main input information may not be as relevant anymore. Specifically, we select as a test sample a cherry, whose form is arguably distinguishable from that of many other fruits in the dataset. One can expect that comparing it against the train set of the FD will result in having both unicommodity and multicommodity dynamics able to assign low  $J_r^*$  to train cherries and higher costs to other fruits. This intuition is confirmed by the results in Figure 4. Here, train cherries (in green) strongly cluster in the lower portion of the cost landscape, whereas all the other fruits have higher costs. In Figure 4, we also plot some of the correctly classified train samples. These results suggest that when color information is negligible compared to another type of information (e.g., shape), unicommodity and multicommodity formulations perform similarly. In light of this, we reinforce the claim that our multicommodity formulation can boost classification in contexts where color information does matter but may not give any advantage when other types of information are more informative. We encourage practitioners to evaluate when this is the case based on domain knowledge when available.

## 5 Conclusion

We propose a physics-informed multicommodity OT formulation for effectively using color information to improve image classification. We model colors as immiscible flows traveling on a capacitated network and propose equations for its dynamics, with the goal of optimizing flow distribution on edges. Color flows are regulated by a shared conductivity

to minimize a unique cost function. Thresholding the ground cost as in [22, 23] makes our model computationally efficient.

We outperform other OT-based approaches such as the Sinkhorn algorithm on two datasets where color matters. Our model also assigns a lower cost to correctly classified images than its unicommodity counterpart, and it is more robust on datasets where items have similar shape. Thus, color information is distinctly relevant. We note that for some datasets, color information may not matter as much as another type of information (e.g., shape), which has stronger discriminative power. However, while we focused here on different color channels as the different commodities in our formulation, the ideas of this study can be extended to scenarios where other relevant information can be distinguished into different types. For instance, one could combine several features together, e.g., colors, contours, and objects' orientations when available.

Our model can be further improved. While it uses the thresholding of [22, 23] to speed up convergence (as mentioned in Section 3.1), it is still slower than Sinkhorn-based methods. Hence, investigating approaches aimed at improving its computational performance is an important direction for future work. Speed-up can be achieved, for example, with the implementation of [39], where the unicommodity OT problem on sparse topologies is solved in  $O(|E|^{0.36})$  time steps. This bound has been found using a backward Euler scheme combined with the inexact Newton–Raphson method for the update of  $x$  and solving Kirchhoff's law using an algebraic multigrid method [44].

Our main goal is to frame an image classification task into that of finding optimal flows of masses of different types in networks built from images. We follow physics principles to assess whether using colors as immiscible flows can give an advantage compared to other standard OT-based methods that do not incorporate such insights. The increased classification performance observed in our experiments stimulates the integration of similar ideas into deep network architectures [45] as a relevant avenue for future work. Combining their prediction capabilities with our insights on how to better exploit the various facets of the input data has the potential to push the performance of deep classifiers even further. For example, one could extend the state-of-the-art architecture of Eisenberger et al. [45], which efficiently computes implicit gradients for generic Sinkhorn layers within a neural network, by including edge, shape, and contour information for Wasserstein barycenter computation or image clustering.

## Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: <https://doi.org/10.7910/DVN/QDHYST>, <https://github.com/daniloeler/fruitdataset>.

## Author contributions

All authors contributed to developing the models, conceiving the experiments, analyzing the results, and reviewing the manuscript. AL and DB conducted the experiments. All authors read and agreed to the published version of the manuscript.

## Acknowledgments

The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting AL and DB.

## References

- Kaiser F, Ronellenfitch H, Without D. Discontinuous transition to loop formation in optical supply networks. *Nat Commun* (2020) 11:5796–11. doi:10.1038/s41467-020-19567-2
- Lonardi A, Putti M, De Bacco C. Multicommodity routing optimization for engineering networks. *Scientific Rep* (2022) 12:7474. doi:10.1038/s41598-022-11348-9
- Lonardi A, Facca E, Putti M, De Bacco C. *Infrastructure adaptation and emergence of loops in network routing with time-dependent loads*. arXiv (2021). doi:10.48550/ARXIV.2112.10620
- Demetci P, Santorella R, Sandstede B, Noble WS, Singh R. *Gromov-Wasserstein optimal transport to align single-cell multi-omics data*. bioRxiv (2020). doi:10.1101/2020.04.28.066787
- Katifori E, Szöllösi GJ, Magnasco MO. Damage and fluctuations induce loops in optimal transport networks. *Phys Rev Lett* (2010) 104:048704. doi:10.1103/PhysRevLett.104.048704
- Werman M, Peleg S, Rosenfeld A. A distance metric for multidimensional histograms. *Comput Vis Graphics, Image Process* (1985) 32:328–36. doi:10.1016/0734-189X(85)90055-6
- Peleg S, Werman M, Rom H. A unified approach to the change of resolution: Space and gray-level. *IEEE Trans Pattern Anal Machine Intelligence* (1989) 11:739–42. doi:10.1109/34.192468
- Rubner Y, Tomasi C, Guibas LJ. A metric for distributions with applications to image databases. In: *31st International Conference on Computer Vision (IEEE Cat. No.98CH36271)*. Bombay, India: IEEE (1998). p. 59–66. doi:10.1109/ICCV.1998.710701
- Rubner Y, Tomasi C, Guibas LJ. The earth mover's distance as a metric for image retrieval. *Int J Comput Vis* (2000) 40:99–121. doi:10.1023/A:1026543900054
- Baptista D, De Bacco C. Principled network extraction from images. *R Soc Open Sci* (2021) 8:210025. doi:10.1098/rsos.210025
- Peyré G, Cuturi M. Computational optimal transport: With applications to data science. *Foundations Trends® Machine Learn* (2019) 11:355–607. doi:10.1561/22000000073
- Koehl P, Delarue M, Orland H. Optimal transport at finite temperature. *Phys Rev E* (2019) 100:013310. doi:10.1103/PhysRevE.100.013310
- Aurell E, Mejía-Monasterio C, Muratore-Ginanneschi P. Optimal protocols and optimal transport in stochastic thermodynamics. *Phys Rev Lett* (2011) 106:250601. doi:10.1103/PhysRevLett.106.250601
- Leite D, De Bacco C. *Revealing the similarity between urban transportation networks and optimal transport-based infrastructures*. arXiv preprint arXiv:2209.06751 (2022).
- Baptista D, Leite D, Facca E, Putti M, De Bacco C. Network extraction by routing optimization. *Scientific Rep* (2020) 10:20806. doi:10.1038/s41598-020-77064-4
- Ibrahim AA, Lonardi A, Bacco CD. Optimal transport in multilayer networks for traffic flow optimization. *Algorithms* (2021) 14:189. doi:10.3390/a14070189
- Mondino A, Suhr S. An optimal transport formulation of the Einstein equations of general relativity. *J Eur Math Soc* (2022). doi:10.4171/JEMS/1188
- Grauman K, Darrell T. Fast contour matching using approximate Earth mover's distance. In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer*

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note


All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphy.2023.1089114/full#supplementary-material>

- Vision and Pattern Recognition, 2004. CVPR 2004; 27 June 2004 - 02 July 2004. Washington, DC, USA: IEEE (2004). doi:10.1109/CVPR.2004.1315035
- Cuturi M. Sinkhorn distances: Lightspeed computation of optimal transport. In: *Advances in neural information processing systems*, Vol. 26. Red Hook, NY, USA: Curran Associates, Inc. (2013). p. 2292–300.
- Koehl P, Delarue M, Orland H. Statistical physics approach to the optimal transport problem. *Phys Rev Lett* (2019) 123:040603. doi:10.1103/PhysRevLett.123.040603
- Thorpe M, Park S, Kolouri S, Rohde GK, Slepčev D. A transportation  $L^p$  distance for signal analysis. *J Math Imaging Vis* (2017) 59:187–210. doi:10.1007/s10851-017-0726-4
- Pele O, Werman M. A linear time histogram metric for improved SIFT matching. In: *Computer vision - ECCV 2008*. Berlin, Heidelberg: Springer Berlin Heidelberg (2008). p. 495–508. doi:10.1007/978-3-540-88690-7\_37
- Pele O, Werman M. Fast and robust earth mover's distances. In: *2009 IEEE 12th International Conference on Computer Vision*. Kyoto, Japan: IEEE (2009). p. 460–7. doi:10.1109/ICCV.2009.5459199
- Villani C. In: *Optimal transport: Old and new*, Vol. 338. Berlin, Heidelberg: Springer (2009). doi:10.1007/978-3-540-71050-9
- Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. In: *Deep Recup YW Teh, editors. Proceedings of the 34th international conference on machine learning (PMLR). Proceedings of machine learning research*, Vol. 70 (2017). p. 214–23.
- Lin T, Ho N, Jordan M. On efficient optimal transport: An analysis of greedy and accelerated mirror descent algorithms. In: *Proceedings of the 36th international conference on machine learning (PMLR). In: Proceedings of machine learning research*, Vol. 97 (2019). p. 3982–91.
- Dvurechensky P, Gasnikov A, Kroshnin A. Computational optimal transport: Complexity by accelerated gradient descent is better than by Sinkhorn's algorithm. In: *Proceedings of the 35th international conference on machine learning (PMLR). In: Proceedings of machine learning research*, Vol. 80 (2018). p. 1367–76.
- Banavar JR, Colaiori F, Flammini A, Maritan A, Rinaldo A. Topology of the fittest transportation network. *Phys Rev Lett* (2000) 84:4745–8. doi:10.1103/PhysRevLett.84.4745
- Ronellenfitch H, Katifori E. Global optimization, local adaptation, and the role of growth in distribution networks. *Phys Rev Lett* (2016) 117:138301. doi:10.1103/PhysRevLett.117.138301
- Hu D, Cai D. Adaptation and optimization of biological transport networks. *Phys Rev Lett* (2013) 111:138701. doi:10.1103/PhysRevLett.111.138701
- Corson F. Fluctuations and redundancy in optimal transport networks. *Phys Rev Lett* (2010) 104:048703. doi:10.1103/PhysRevLett.104.048703
- Lonardi A, Facca E, Putti M, De Bacco C. Designing optimal networks for multicommodity transport problem. *Phys Rev Res* (2021) 3:043010. doi:10.1103/PhysRevResearch.3.043010
- Bonifazi V, Facca E, Folz F, Karrenbauer A, Kolev P, Mehlhorn K, et al. Physarum-inspired multi-commodity flow dynamics. *Theor Comput Sci* (2022) 920:1–20. doi:10.1016/j.tcs.2022.02.001

34. Bonifaci V, Mehlhorn K, Varma G. Physarum can compute shortest paths. *J Theor Biol* (2012) 309:121–33. doi:10.1016/j.jtbi.2012.06.017
35. Rinaldo A, Rodriguez-Iturbe I, Rigon R, Ijjasz-Vasquez E, Bras RL. Self-organized fractal river networks. *Phys Rev Lett* (1993) 70:822–5. doi:10.1103/PhysRevLett.70.822
36. Barabási AL. Invasion percolation and global optimization. *Phys Rev Lett* (1996) 76:3750–3. doi:10.1103/PhysRevLett.76.3750
37. Santambrogio F. Optimal channel networks, landscape function and branched transport. *Inter Free Boundaries* (2007) 9:149–69. doi:10.4171/IFB/160
38. Ibrahim AA, Leite D, De Bacco C. Sustainable optimal transport in multilayer networks. *Phys Rev E* (2022) 105:064302. doi:10.1103/PhysRevE.105.064302
39. Facca E, Benzi M. Fast iterative solution of the optimal transport problem on graphs. *SIAM J Scientific Comput* (2021) 43:A2295–A2319. doi:10.1137/20M137015X
40. Flamary R, Courty N, Gramfort A, Alaya MZ, Boisbunon A, Chambon S, et al. POT: Python optimal transport. *J Machine Learn Res* (2021) 22:1–8.
41. Schmitzer B. Stabilized sparse scaling algorithms for entropy regularized transport problems. *SIAM J Scientific Comput* (2019) 41:A1443–A1481. doi:10.1137/16M1106018
42. Seeland M, Rzanny M, Alaqraa N, Wäldchen J, Mäder P. *Jena Flowers 30 dataset* (2017). doi:10.7910/DVN/QDHYST
43. Macanhã PA, Eler DM, Garcia RE, Junior WEM. Handwritten feature descriptor methods applied to fruit classification. In: *Information Technology - new generations*. Cham: Springer International Publishing (2018). p. 699–705. doi:10.1007/978-3-319-54978-1\_87
44. Trottenberg U, Oosterlee CW, Schuller A. *Multigrid*. Amsterdam, Netherlands: Elsevier (2000).
45. Eisenberger M, Toker A, Leal-Taixé L, Bernard F, Cremers D. A unified framework for implicit sinkhorn differentiation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022). p. 509–18.
46. *MODI (Open Source code implementation)* (2022). Available from: <https://github.com/aleable/MODI>.

**Bilevel Optimization for Traffic Mitigation in Optimal Transport Networks**Alessandro Lonardi<sup>✉\*</sup> and Caterina De Bacco<sup>†</sup>*Max Planck Institute for Intelligent Systems, Cyber Valley, Tübingen 72076, Germany* (Received 28 June 2023; revised 21 September 2023; accepted 22 November 2023; published 26 December 2023)

Global infrastructure robustness and local transport efficiency are critical requirements for transportation networks. However, since passengers often travel greedily to maximize their own benefit and trigger traffic jams, overall transportation performance can be heavily disrupted. We develop adaptation rules that leverage optimal transport theory to effectively route passengers along their shortest paths while also strategically tuning edge weights to optimize traffic. As a result, we enforce both global and local optimality of transport. We prove the efficacy of our approach on synthetic networks and on real data. Our findings on the international European highways suggest that thoughtfully devised routing schemes might help to lower car-produced carbon emissions.

DOI: [10.1103/PhysRevLett.131.267401](https://doi.org/10.1103/PhysRevLett.131.267401)

*Introduction.*—Transport networks are ubiquitous in nature and engineering, spanning from living organisms to cities and telecommunications. Many of these systems can be modeled by adaptation rules that follow the principle of minimum energy, regulating edge flows to optimize transportation costs. Examples in biology are plants, whose profiles emerge from a trade-off between minimization of hydraulic resistance and carbon cost [1], and leaves, shaped by the interplay of nutrients' transport efficiency and robustness to damage [2–4].

Similarly, adaptation rules have been employed to model traffic flows in urban transportation by jointly minimizing the energy dissipated by the passengers and the construction cost of the infrastructure [5–11]. While these models set forth a first approach to simulate traffic flows using adaptation, they crucially neglect that passengers in a transportation network do not move cohesively to minimize a unique cost. Instead, they choose their routes greedily to maximize their benefit (Wardrop's first principle) [12–14]. As a consequence, transport networks may be globally inefficient.

In this Letter, we propose a set of adaptation equations to find traffic flows that mitigate congestion, considered as a proxy for global efficiency, while trading off against the shortest routes.

We frame the problem in a bilevel optimization setup, which poses a competition between greedy passengers and a network manager. The passengers minimize their

origin-destination path cost seeking for the user equilibrium [15] (lower-level problem), whereas the network manager guarantees global efficiency by mitigating traffic bottlenecks on edges to achieve the system optimum (upper-level problem), while implicitly accounting for passengers' shortest path. We tackle the optimization problem by alternating optimal transport- (OT) inspired adaptation rules for the lower-level optimization and a projected stochastic gradient descent (PSGD) scheme for the upper-level optimization.

In detail, greedy passenger flows are found by solving a dynamical system that governs the evolution of edge capacities, variables that control passenger allocation, so that these travel on their shortest paths. Adaptation rules are a well-established mechanism for route assignment on networks [3,5–11,16–21] and in continuous domains [22–26]. Classically, user equilibrium greedy flows can be found with the Frank-Wolfe algorithm [27] or, alternatively, with recent methods accounting for passengers' travel budgeting [28]. Here, we propose a model that exploits OT theory to prove that, at convergence, passengers move along the shortest path. Particularly, our dynamical system admits a Lyapunov functional [24] that asymptotically converges to the shortest path (Wasserstein) distance between entry and exit distributions of passengers [16,17,29].

Traffic mitigation is performed by minimizing a quadratic loss function that penalizes edges whose traffic exceeds a prefixed threshold. The minimization problem can be treated analytically by assuming that the network edges are endowed with capacities and weights (resistances) and their flows are the gradient of a scalar potential, as for electrical networks. We derive closed-form gradients for the weights, which can be interpreted as the cost that passengers pay for traveling. In practice, network managers would implement these weights by strategically designing incentives or disincentives, e.g., assigning road tolls, to encourage

---

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by the Max Planck Society.*

passengers to relocate from jammed edges. The task of traffic mitigation has been addressed using several methods. These include belief propagation [30–32], adaptive dynamical networks [33], Markov chain Monte Carlo schemes [34], cellular automata [35,36], and heuristic routing models [37].

A bilevel optimization problem similar to the one studied here was solved using message passing [38]. While the problem’s setting is similar to ours, the methodologies differ since we alternate adaptation rules for the capacities with global descent for the weights, whereas message passing uses local updates for flows. Our approach outputs individual passengers’ optimal paths, whereas the formulation in Li *et al.* [38] can only extract aggregate routes.

We find that our method effectively trades off traffic mitigation against the shortest passenger routes. Namely, both on synthetic topologies and real roads, it returns optimal transport networks where congestion is heavily reduced. We argue that this result is beneficial for reducing the carbon footprint of roads. We also show that the uncoordinated actions of the network manager and passengers can be counterproductive, i.e., they may increase traffic, with an outcome opposite to that intended.

*Problem.*—We take a network  $G(V, E)$  where  $M \geq 1$  groups of greedy passengers  $i$  can travel from origin nodes  $O^i$  (one node per group) to possibly multiple destination nodes  $D^i$ . Stationary numbers of entry and exit passengers are stored in a mass matrix with entries  $\tilde{S}_v^i > 0$  for each  $v = O^i$ ,  $\tilde{S}_v^i < 0$  for  $v \in D^i$ , and  $\tilde{S}_v^i = 0$  otherwise. We assume that the system is isolated, i.e., that passengers entering the network must also exit. This condition is  $\sum_v \tilde{S}_v^i = 0$  for all  $i$ . When traveling along an edge, passengers pay a cost  $\tilde{w}_e > 0$ , and finally, each edge is equipped with a capacity that controls the rate at which passengers  $i$  are allocated along each edge  $e$ ;  $\tilde{c}_e^i \geq 0$ . Intuitively, one could think of capacities as the space occupied by passengers of type  $i$ , i.e., larger space accommodates more passengers. All problem variables have been introduced with units, however, these can be nondimensionalized to derive scale-independent adaptation rules (see Supplemental Material [39]). We denote dimensional quantities with a tilde and dimensionless ones without.

*Lower-level optimization.*—The lower-level problem allows us to find the cheapest routes from  $O^i$  to  $D^i$ . In order to model traffic flows, we introduce the fluxes  $F_e^i$ , specifying the displacement of  $S^i$  along an edge  $e$ . In analogy with electrical networks, we assume that there exists an auxiliary pressure potential  $p_v^i$  on each node  $v$  due to index  $i$ . We interpret them as the travel demand from passengers traveling from  $v$ . With this, we define the potential-based fluxes for all  $e = (u, v)$  and  $i$ , i.e., Poiseuille’s law, as

$$F_e^i = \frac{c_e^i}{w_e} (p_u^i - p_v^i). \quad (1)$$

Fluxes must obey Kirchhoff’s law. We can write it as  $\sum_e B_{ve} F_e^i = S_v^i$ , where  $B$  is a conventionally oriented incidence matrix of the network. Substituting Eq. (1) in Kirchhoff’s law, the potential becomes a function of  $c$  and  $w$ , namely,  $p_v^i = \sum_u (L^{\dagger i})_{vu} S_u^i$ , where  $\dagger$  denotes the Moore-Penrose inverse and  $L_{uv}^i = \sum_e (c_e^i/w_e) B_{ue} B_{ve}$  are entries of the network weighted Laplacian. With this substitution,  $F \equiv F(c, w)$  is also a function of only  $c$  and  $w$ , the only independent problem’s variables.

For any fixed set of weights, we write the lower-level problem as

$$J(c, w) = \sum_{et} w_e |F_e^i|, \quad (2)$$

$$\min_{c \geq 0} J(c, w). \quad (3)$$

The convex OT cost  $J$  in Eq. (2) is the sum over  $M$  indexes of the  $w$  shortest path costs  $J^i = \sum_e w_e |F_e^i|$  [16,17]. Its only minimizer is the overlap of  $M$  shortest paths from all  $O^i$  to  $D^i$ , which are found with  $c$  using Eq. (1) and Kirchhoff’s law.

*Upper-level optimization.*—The upper-level problem formalizes the task of the network manager of tuning  $w$  to mitigate traffic jams triggered by the passengers. We measure traffic by penalizing congested links where  $\sum_i |F_e^i|$  exceeds a threshold  $\theta \geq 0$ , above which infrastructural failures may occur. Conveniently, we introduce  $\Delta_e = \sum_i |F_e^i| - \theta$ .

Analogously to Eqs. (2) and (3), for any set of capacities, the upper-level optimization is

$$\Omega(c, w) = \frac{1}{2} \sum_e \Delta_e^2 H(\Delta_e), \quad (4)$$

$$\min_{w \geq c} \Omega(c, w), \quad (5)$$

where  $H$  is the Heaviside step function. In Eq. (4), other objective functions, e.g., the hinge loss, can be utilized [38,46]; we do not explore this here. Furthermore, the weights are constrained to be larger than a small  $\epsilon > 0$ . This means that passengers cannot profit ( $w < 0$ ) or travel for free ( $w = 0$ ). Practically, this ensures that the Laplacian  $L$  is well defined.

*Bilevel optimization.*—We combine the two optimization problems into one. Suppose that the network manager is regularly informed of the passengers’ routes and, using such information, the weights are tuned to mitigate traffic. After each update, passengers reroute according to the updated weights.

Formally, this translates into the problem

$$\min_{w \geq c} \Omega(w; \hat{c}), \quad (6)$$

$$\text{such that } \hat{c} = \operatorname{argmin}_{c \geq 0} J(c; w), \quad (7)$$



where the equality in Eq. (7) comes from the convexity of  $J$  [16,17]. In Eq. (6) we explicitly state the dependence on  $w$  as a variable and on  $c$  as a parameter [conversely for Eq. (7)].

*Optimal transport dynamics.*—To find the shortest paths required for the lower-level problem, we couple fluxes and capacities with the ordinary differential equations

$$\frac{dc_e^i}{dt} = \frac{F_e^i{}^2}{c_e^i} - c_e^i, \quad (8)$$

where fluxes obey Kirchhoff's law. In Eq. (8), edges with high flux enlarge, whereas those where the negative decaying term prevails shrink. Crucially, asymptotic solutions converge to the minimum OT cost  $J$  in Eq. (2), being the Wasserstein distance between passengers' entry and exit distributions [39] and whose minimizers are origin-destination shortest paths. Beside Kirchhoff's law and positivity, capacities in Eqs. (2) and (3) are otherwise unconstrained. One can potentially add additional constraints, e.g., a limited budget, by employing recent ideas in the context of adaptation equations [40]. We do not explore this here.

*Projected stochastic gradient descent.*—Minimization of Eq. (6) is performed using stochastic gradient descent with a projection step to enforce  $w \geq e$ . Importantly, we can derive a closed-form expression for the gradients  $\Psi_e = \partial\Omega/\partial w_e$  [39]. To explore the nonconvex landscape of the minimization in Eqs. (6) and (7), we update the weights with dropout at each step, i.e., setting to zero  $|E|(1-q)$  random gradients, where  $0 \leq q \leq 1$ . For  $q = 1$  we get vanilla gradient descent.

*Bilevel optimization scheme.*—In order to find the optimal  $c$  and  $w$ , and hence  $F$ , we iterate between Eq. (8) and PSGD recursively. The scheme is repeated until  $J$  and  $\Omega$  converge. A diagram outlining the optimization method is in Fig. 1; we also provide an open-source code (Bilevel routing on networks with optimal transport, BROTON) [47].

*Experimental setup.*—We analyze BROTON's optimal networks against two baselines. The first, referred to as OT, consists of finding passengers' shortest paths without any intervention from the network manager. We assume a unitary cost per unit of length fare, i.e., we set  $w = \ell$  with  $\ell$  the Euclidean lengths of the edges, and numerically integrate Eq. (8). The second, referred to as PSGD, reflects the scenario of a network manager that tunes  $w$  only relying on the shortest paths taken when  $w = \ell$  and that disregards how fluxes redistribute while updating  $w$ . In practice, this corresponds to running PSGD only, with initial conditions being  $w(0) = \ell + \xi$  and  $c_e^i \approx |F_{\text{Dij},e}^i|$  [39], and then to integrating Eq. (8), with  $w = w_{\text{PSGD}}^*$  being the optimal weights returned by the network manager. Here,  $\xi$  is a small zero-sum uniform noise,  $F_{\text{Dij}}^i$  are the shortest path fluxes computed with Dijkstra's algorithm, and the approximation arises because, to avoid numerical instabilities, a small

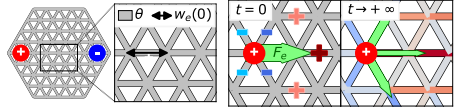


FIG. 1. Bilevel optimization scheme on a lattice. Entry and exit inflows are the red and blue nodes, respectively. Initially, (green) fluxes distribute minimizing the travel cost  $w_e(t=0) = \ell_e$ , being the length of an edge. If they exceed  $\theta$  they get penalized; hence, the network manager tunes the weights to encourage rerouting over more expensive (red) or cheaper (blue) edges (for a companion figure, see Supplemental Material [39]).

nonzero  $c_e^i$  is allocated to all edges. We fix BROTON's initial conditions to  $w(0) = \ell + \xi$  and  $c_e^i(0) = S_{O_i}^i$ .

*Synthetic experiments.*—First, we study a network of size  $|V| = 300$ ,  $|E| = 864$ , with nodes placed uniformly at random in the unitary disk and edges extracted from their Delaunay triangulation. Entry and exit inflows are  $S_{O_i}^i = +1$  on an origin node at the center, and  $S_{D_i}^i = -1/D$ , on  $D = 4, 8$  destinations  $D^i$  on the disk edge. Since  $M = 1$ , there is only a single index  $i$ . Here we discuss results for  $D = 8$ , for experiments with varying  $q$  for  $D = 4, 8$ ; see Supplemental Material [39].

We evaluate  $J$  and  $\Omega$  at convergence for all methods with different  $q$  and ranging  $\theta$  from  $\theta = 0$  to a large value  $\theta^*$  where few edges are congested. Results are in Fig. 2(a).

Since for OT the network manager does not intervene,  $J$  is constant for all  $\theta$ , and it is the origin-destination shortest length. Its profile changes when the network manager influences passengers' routes by tuning the weights. Specifically, for PSGD  $J$  drops when reducing  $\theta$ , making it cheaper for the passengers to move. On the contrary, lower  $\theta$  corresponds to a larger  $J$  for BROTON. This behavior seemingly favors an uninformed network manager (PSGD) over an informed one (BROTON). However, the profile of  $\Omega$  shows that, even though the traveling cost of PSGD is cheaper, all transport networks at convergence are highly congested (large  $\Omega$ ). BROTON successfully trades off the cost of traveling against traffic, outputting low values of  $\Omega$  for all  $\theta$ , with only a mild increase as  $\theta$  approaches zero. This is clarified in Fig. 2(c), where BROTON generates ramified loopy networks.

The dropout parameter  $q$  allows us to explore the minimization landscape of Eqs. (6) and (7). By decreasing  $q$ , i.e., setting more gradients to zero, BROTON returns lower  $J$ s at all  $\theta$ , whereas PSGD gives higher ones, conversely for  $\Omega$ . This impacts the network topologies, which are less ramified and akin to OT trees [48], when  $q$  is lower and for the same  $\theta$  [39]. The trade-off between  $J$  and  $\Omega$  is further laid out in Fig. 2(b) where we show  $J - J_{\text{OT}}$  against  $\Omega - \Omega_0$ ,  $\Omega_0 = 0$ . We highlight in red the nondominated points (also referred to as maximal points) at four values of  $\theta$ , computed over all  $q$  [as in Fig. 2(a)] and 25 random

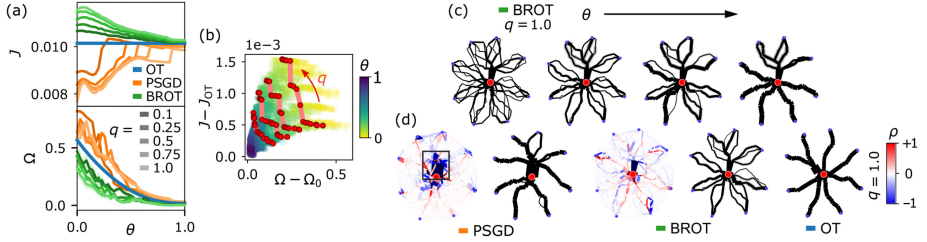


FIG. 2. Overview of the routing schemes. (a)  $J$  and  $\Omega$  against  $\theta$ . (b) Trade-off  $J - J_{\text{OT}}$  vs  $\Omega - \Omega_0$  with varying  $(\theta, q, \xi)$ . Nondominated points for  $\theta/\theta^* \simeq \{0.06, 0.2, 0.3, 0.4\}$  are in red. (c) BROT's networks at different  $\theta$ . Edge widths are proportional to the average fluxes in 50 runs of the algorithm. Gray edge contours are fluxes' standard deviations. (d) Cost (left) and flux (right) networks for all methods and  $\theta/\theta^* = 0.4$ . Flux networks are as in (c), whereas edges in the cost networks are colored with  $\rho$  and their widths are proportional to the fluxes. The black rectangle frames a region where the network manager triggers high congestion. We conveniently normalize  $\theta^*$  and  $\rho$ .

initializations of BROT. Such points are the best  $J$ - $\Omega$  trade-off attained by the experimental runs [39].

For all  $q$  and sufficiently low  $\theta$ , the price of anarchy (PoA) [49] is greater for PSGD than for OT, i.e., the network manager's intervention increases traffic congestion, having the opposite effect to that intended. We illustrate exemplary networks at convergence in Fig. 2(c). The parameter  $\rho = w_X^* - \ell$  ( $X = \text{BROT, PSGD}$ ), expressing the variation of cost, indicates that the uninformed network manager naively—and significantly—decreases the cost of a small fraction of edges [square in Fig. 2(d)]. This encourages fluxes to largely concentrate on them, thus creating congestion.

To further discern the nature of congestion, we propose two additional metrics. First, the Gini coefficient of the fluxes,  $\text{Gini} = \sum_{mn} |x_m - x_n|/2|E|^2 \bar{x}$ , where  $\bar{x} = \sum_e x_e/|E|$  and  $x_e = \sum_i |F_e^i|$ .  $\text{Gini} = 0$  corresponds to uniformly distributed fluxes and larger Gini corresponds to high congestion. Second, the total travel time  $T_\theta(s) = \sum_{ei} t_{\theta,e}(s) |F_e^i|$ , computed with an affine latency function for overtrafficked edges [38,50], namely,  $t_{\theta,e}(s) = \ell_e(1 + s\Delta_e/\theta)/v_\infty$  if  $\sum_i |F_e^i| \geq \theta$ , and  $t_{\theta,e}(s) = \ell_e/v_\infty$  otherwise. Here  $v_\infty = 1$  is a (conventionally fixed) free-flow velocity, and  $s$  is a sensitivity coefficient to penalize traffic. Results are in Fig. 3.

The Gini coefficient of PSGD fluctuates slightly around the high values attained by the congested shortest path network of OT. For BROT, as  $\theta$  decreases—more flux gets penalized—Gini sharply drops, yielding progressively distributed networks. The total travel time reveals once again that the uncoordinated action of passengers and the network manager may be detrimental compared to having no tuning of  $w$ . In fact, times for PSGD are higher than those for OT. BROT keeps  $T_\theta(s)$  small for any value of  $\theta$  and for both low and high sensitivity. Finally, as  $\theta$  increases, traffic gradually mitigates, with  $\lim_{\theta \rightarrow +\infty} T_\theta(s) = T_\infty$  ( $T_\infty = J_{\text{OT}}$ ) being the travel time for infinite capacities, when all passengers flow freely.

*The E-road network.*—We study the methods on a graph extracted from the international European highways (E-road) [51,52], of size  $|V| = 541$  and  $|E| = 712$ . Entry inflows of passengers are populations of 15 large cities. We assume that all passengers travel from one city to another. Thus, we set for  $O^i$  and  $v \in D^i$  (being also origin nodes  $O^j$ ) the exiting number of passengers  $\tilde{S}_v$  to be proportional to the product  $r_v = \tilde{S}_{O^i} \tilde{S}_{O^j}$ , properly normalized to ensure conservation of mass. In this way, cities with high inflows have large outflows, and vice versa for small ones. The total number of passengers to be routed is  $\sum_i \tilde{S}_{O^i} \simeq 3 \times 10^7$ . We fix  $\hat{\theta}$  (dimensionalized by  $S_c$ ) so that 43% of the passengers reroute from their congested shortest path, found with Dijkstra's and  $w = \ell$ .

Results are in Fig. 4. We observe that, in the shortest path configuration of OT, a large volume of passengers travels between the two most populous cities, Madrid and Berlin, on the southernmost region of the network. The uninformed network (PSGD) heavily increases the price of the connections to Milan [39]. This causes a heavy rerouting from Madrid to the north and congests the roads connecting Madrid to Paris and then from Paris to Berlin. In contrast, BROT distributes traffic over a ramified road network.

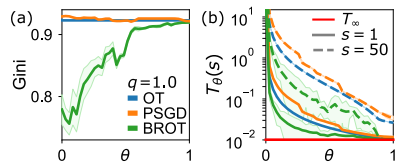


FIG. 3. Measuring traffic congestion,  $D = 8$ . (a) Gini coefficient against  $\theta$ . (b)  $T_\theta(s)$  against  $\theta$ . Solid lines correspond to low sensitivity  $s = 1$  and dashed ones to  $s = 50$ ; in red we draw  $T_\infty$  (free flow). Shades are standard deviations over 50 realizations of the algorithms.

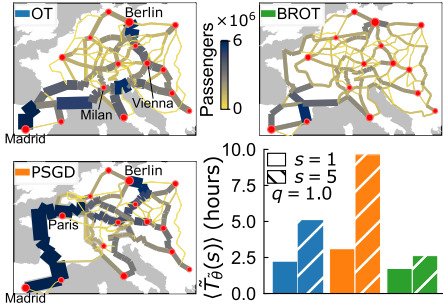


FIG. 4. E-road transport networks. Nodes in red are 15 main cities taken as passenger inflows, their size is proportional to the entry inflows. Edge widths are the total number of passengers  $\sum_i |\bar{F}_e^i|$ ; gray shades are standard deviations over 50 realizations of the algorithms.

We study the average travel time for all routing schemes. This is  $\langle \bar{T}_{\bar{\theta}}(s) \rangle = \sum_{ei} \bar{\tau}_{e,\bar{\theta}}(s) |\bar{F}_e^i| / \sum_{ei} |\bar{F}_e^i|$ , where  $\bar{\tau}_{\bar{\theta}}$  is a dimensionalized latency function computed using  $\bar{\ell}$ , the Euclidean distance between cities, and  $v_{\infty} = 100$  (km/h).

Results for  $s = 1, 5$  in Fig. 4 show that the average travel time of BROT is substantially lower than that of OT and PSGD. Particularly, for low sensitivity BROT's  $\langle \bar{T}_{\bar{\theta}}(s) \rangle$  is approximately 1.7 h), while OT's and PSGD's are 2.3 and 3.1 h. Here, BROT leads to a reduction in traveled time of approximately 26% and 45% compared to OT and PSGD. This result becomes starker if the sensitivity increases, here BROT reduces  $\langle \bar{T}_{\bar{\theta}}(s) \rangle$  of 48% compared to OT—from 5 to 2.6 h—and of 74% compared to PSGD—whose heavy congestion gives  $\langle \bar{T}_{\bar{\theta}}(s) \rangle \simeq 10$  h. Once again, the PoA (the travel time) is higher if the network manager's intervention is uncoordinated with the passengers (PSGD), as opposed to when there is no intervention (OT).

Experiments on the E-road network for  $q = 0.25, 0.5, \text{ and } 0.75$  are in the Supplemental Material [39].

**Conclusion.**—BROT relies on theoretical assumptions that can be challenging to meet in real-world traffic control [53], e.g., passengers rerouting more unpredictably than expected by theoretical models. Nevertheless, our analysis on the E-road network demonstrates how an informed tuning of road tolls—where the network manager factors in passengers' rerouting—can be beneficial for reducing the carbon footprint of roads, since traffic jams, and hence longer travels, critically impact greenhouse gas emissions of vehicles [54–56].

To facilitate practitioners using our algorithms, we open source our code [47].

The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting A. L.

\*alessandro.lonardi@tuebingen.mpg.de

†caterina.debacco@tuebingen.mpg.de

- [1] L. Koçillari *et al.*, The widened pipe model of plant hydraulic evolution, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2100314118 (2021).
- [2] E. Katifori, G. J. Szöllösi, and M. O. Magnasco, Damage and fluctuations induce loops in optimal transport networks, *Phys. Rev. Lett.* **104**, 048704 (2010).
- [3] H. Ronellenfitch and E. Katifori, Global optimization, local adaptation, and the role of growth in distribution networks, *Phys. Rev. Lett.* **117**, 138301 (2016).
- [4] H. Ronellenfitch and E. Katifori, Phenotypes of vascular flow networks, *Phys. Rev. Lett.* **123**, 248101 (2019).
- [5] A. Tero, S. Takagi, T. Saigusa, K. Ito, D. P. Bebber, M. D. Fricker, K. Yumiki, R. Kobayashi, and T. Nakagaki, Rules for biologically inspired adaptive network design, *Science* **327**, 439 (2010).
- [6] A. Lonardi, E. Facca, M. Putti, and C. De Bacco, Designing optimal networks for multicommodity transport problem, *Phys. Rev. Res.* **3**, 043010 (2021).
- [7] A. Lonardi, M. Putti, and C. De Bacco, Multicommodity routing optimization for engineering networks, *Sci. Rep.* **12**, 7474 (2022).
- [8] A. Lonardi, E. Facca, M. Putti, and C. De Bacco, Infrastructure adaptation and emergence of loops in network routing with time-dependent loads, *Phys. Rev. E* **107**, 024302 (2023).
- [9] A. A. Ibrahim, A. Lonardi, and C. De Bacco, Optimal transport in multilayer networks for traffic flow optimization, *Algorithms Mol. Biol.* **14**, 189 (2021).
- [10] A. A. Ibrahim, D. Leite, and C. De Bacco, Sustainable optimal transport in multilayer networks, *Phys. Rev. E* **105**, 064302 (2022).
- [11] V. Bonifaci, E. Facca, F. Folz, A. Karrenbauer, P. Kolev, K. Mehlhorn, G. Morigi, G. Shahkarami, and Q. Vermande, Physarum-inspired multi-commodity flow dynamics, *Theor. Comput. Sci.* **920**, 1 (2022).
- [12] H. Youn, M. T. Gastner, and H. Jeong, Price of anarchy in transportation networks: Efficiency and optimality control, *Phys. Rev. Lett.* **101**, 128701 (2008).
- [13] M. T. Gastner and M. E. J. Newman, Optimal design of spatial distribution networks, *Phys. Rev. E* **74**, 016117 (2006).
- [14] R. Selten, T. Chmura, T. Pitz, S. Kube, and M. Schreckenberg, Commuters route choice behaviour, *Games Econ. Behav.* **58**, 394 (2007).
- [15] M. Beckmann, C. B. McGuire, and C. B. Winsten, *Studies in the Economics of Transportation* (Cowles Commission for Research in Economics, 1956).
- [16] V. Bonifaci, K. Mehlhorn, and G. Varma, Physarum can compute shortest paths, *J. Theor. Biol.* **309**, 121 (2012).
- [17] V. Bonifaci, Physarum can compute shortest paths: A short proof, *Inf. Proc. Lett.* **113**, 4 (2013).

- [18] A. Tero, R. Kobayashi, and T. Nakagaki, A mathematical model for adaptive transport network in path finding by true slime mold, *J. Theor. Biol.* **244**, 553 (2007).
- [19] F. Folz, K. Mehlhorn, and G. Morigi, Noise-induced network topologies, *Phys. Rev. Lett.* **130**, 267401 (2023).
- [20] D. Hu and D. Cai, Adaptation and optimization of biological transport networks, *Phys. Rev. Lett.* **111**, 138701 (2013).
- [21] J. B. Kirkegaard and K. Sneppen, Optimal transport flows for distributed production networks, *Phys. Rev. Lett.* **124**, 208101 (2020).
- [22] D. Baptista, D. Leite, E. Facca, M. Putti, and C. De Bacco, Network extraction by routing optimization, *Sci. Rep.* **10**, 088702 (2020).
- [23] E. Facca, F. Cardin, and M. Putti, Towards a stationary Monge-Kantorovich dynamics: The physarum polycyphalum experience, *SIAM J. Appl. Math.* **78**, 651 (2016).
- [24] E. Facca, S. Daneri, F. Cardin, and M. Putti, Numerical solution of Monge-Kantorovich equations via a dynamic formulation, *J. Sci. Comput.* **82**, 68 (2020).
- [25] E. Facca, F. Cardin, and M. Putti, Branching structures emerging from a continuous optimal transport model, *J. Comput. Phys.* **447**, 110700 (2021).
- [26] D. Leite and C. D. Bacco, Revealing the similarity between urban transportation networks and optimal transport-based infrastructures, [arXiv:2209.06751](https://arxiv.org/abs/2209.06751).
- [27] L. J. LeBlanc, E. K. Morlok, and W. P. Pierskalla, An efficient approach to solving the road network equilibrium traffic assignment problem, *Transp. Res.* **9**, 309 (1975).
- [28] Y. Bao, Z. Gao, M. Xu, H. Sun, and H. Yang, Travel mental budgeting under road toll: An investigation based on user equilibrium, *Transp. Res. Part A* **73**, 1 (2015).
- [29] A. Lonardi, D. Baptista, and C. De Bacco, Immiscible color flows in optimal transport networks for image classification, *Front. Phys.* **11**, 1089114 (2023).
- [30] H. F. Po, C. H. Yeung, and D. Saad, Futility of being selfish in optimized traffic, *Phys. Rev. E* **103**, 022306 (2021).
- [31] Y.-Z. Xu, H. F. Po, C. H. Yeung, and D. Saad, Scalable node-disjoint and edge-disjoint multiwavelength routing, *Phys. Rev. E* **105**, 044316 (2022).
- [32] C. H. Yeung, Coordinating dynamical routes with statistical physics on space-time networks, *Phys. Rev. E* **99**, 042123 (2019).
- [33] J. Jiang, X. Wang, and Y.-C. Lai, Optimizing biologically inspired transport networks by control, *Phys. Rev. E* **100**, 032309 (2019).
- [34] V. Colizza, J. R. Banavar, A. Maritan, and A. Rinaldo, Network structures from selection principles, *Phys. Rev. Lett.* **92**, 198701 (2004).
- [35] T. S. Tai and C. H. Yeung, Global benefit of randomness in individual routing on transportation networks, *Phys. Rev. E* **100**, 012311 (2019).
- [36] T. S. Tai and C. H. Yeung, Adaptive strategies for route selection en-route in transportation networks, *Chin. J. Phys. (Taipei)* **77**, 712 (2022).
- [37] G. Yan, T. Zhou, B. Hu, Z.-Q. Fu, and B.-H. Wang, Efficient routing on complex networks, *Phys. Rev. E* **73**, 046108 (2006).
- [38] B. Li, D. Saad, and C. H. Yeung, Bilevel optimization in flow networks: A message-passing approach, *Phys. Rev. E* **106**, L042301 (2022).
- [39] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.131.267401> for more information on the problem formulation, closed-form expressions of the gradients, and additional experiments, which includes [4,40–45].
- [40] A. A. Ibrahim, M. Muehlebach, and C. D. Bacco, Optimal transport with constraints: From mirror descent to classical mechanics, [arXiv:2309.04727](https://arxiv.org/abs/2309.04727).
- [41] M. Muehlebach and M. I. Jordan, On constraints in first-order optimization: A view from non-smooth dynamical systems, *J. Mach. Learn. Res.* **23**, 1 (2022).
- [42] G. H. Golub and V. Pereyra, The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate, *SIAM J. Numer. Anal.* **10**, 413 (1973).
- [43] L. Kantorovich, Mathematical methods of organizing and planning production, *Manage. Sci.* **6**, 366 (1960).
- [44] C. Villani, *Optimal Transport: Old and New* (Springer, Berlin, Heidelberg, 2009), Vol. 338.
- [45] A.-G. Rolland-Lagan and P. Prusinkiewicz, Reviewing models of auxin canalization in the context of leaf vein pattern formation in Arabidopsis, *Plant J.* **44**, 854 (2005).
- [46] J. W. Rocks, H. Ronellenfitsch, A. J. Liu, S. R. Nagel, and E. Katifori, Limits of multifunctionality in tunable networks, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 2506 (2019).
- [47] Bilevel routing on network with optimal transport: BROT, <https://github.com/aleable/BROT>.
- [48] J. R. Banavar, F. Colaiori, A. Flammini, A. Maritan, and A. Rinaldo, Topology of the fittest transportation network, *Phys. Rev. Lett.* **84**, 4745 (2000).
- [49] The price of anarchy measures how much a system degrades due to the greedy behavior of its agents. In our case, the global efficiency of the system is measured by  $\Omega$ , which increases as the network gets more congested by greedy passagers.
- [50] T. Roughgarden and E. Tardos, How bad is selfish routing?, *J. ACM* **49**, 236 (2002).
- [51] J. Kunegis, KONECT: The Koblenz network collection, in *Proceedings of the 22nd International Conference on World Wide Web, WWW '13 Companion* (Association for Computing Machinery, New York, 2013), pp. 1343–1350.
- [52] L. Šubelj and M. Bajec, Robust network community detection using balanced propagation, *Eur. Phys. J. B* **81**, 353 (2011).
- [53] M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos, and Y. Wang, Review of road traffic control strategies, *Proc. IEEE* **91**, 2043 (2003).
- [54] T. Van Woensel, R. Creten, and N. Vandaele, Managing the environmental externalities of traffic logistics: The issue of emissions, *Prod. Oper. Manag.* **10**, 207 (2001).
- [55] F. Kellner, Exploring the impact of traffic congestion on CO<sub>2</sub> emissions in freight distribution networks, *Logist. Res.* **9**, 21 (2016).
- [56] M. Barth and K. Boriboonsomsin, Traffic congestion and greenhouse gases, *ACCESS Mag.* **1**, 2 (2009).

# Bibliography

- [1] A. A. Ibrahim, A. Lonardi, and C. De Bacco, *Optimal Transport in Multilayer Networks for Traffic Flow Optimization*, *Algorithms* **14** (2021), 10.3390/a14070189.
- [2] A. Lonardi, E. Facca, M. Putti, and C. De Bacco, *Designing optimal networks for multicommodity transport problem*, *Phys. Rev. Research* **3**, 043010 (2021).
- [3] A. Lonardi, M. Putti, and C. De Bacco, *Multicommodity routing optimization for engineering networks*, *Scientific Reports* **12**, 7474 (2022).
- [4] A. Lonardi, E. Facca, M. Putti, and C. De Bacco, *Infrastructure adaptation and emergence of loops in network routing with time-dependent loads*, *Phys. Rev. E* **107**, 024302 (2023).
- [5] A. Lonardi, D. Baptista, and C. De Bacco, *Immiscible color flows in optimal transport networks for image classification*, *Frontiers in Physics* **11** (2023), 10.3389/fphy.2023.1089114.
- [6] A. Lonardi and C. De Bacco, *Bilevel optimization for traffic mitigation in optimal transport networks*, *Phys. Rev. Lett.* **131**, 267401 (2023).
- [7] N. Ruggeri, A. Lonardi, and C. De Bacco, *Message-passing on hypergraphs: detectability, phase transitions and higher-order information*, *Journal of Statistical Mechanics: Theory and Experiment* **2024**, 043403 (2024).
- [8] A. Lonardi, M. Szell, and C. De Bacco, *Optimal transport routing in multilayer networks for a cohesive design of bicycle infrastructures*, (2023), in preparation.

- [9] J. R. Banavar, F. Colaiori, A. Flammini, A. Maritan, and A. Rinaldo, *Topology of the Fittest Transportation Network*, *Phys. Rev. Lett.* **84**, 4745 (2000).
- [10] E. Katifori, G. J. Szöllösi, and M. O. Magnasco, *Damage and Fluctuations Induce Loops in Optimal Transport Networks*, *Phys. Rev. Lett.* **104**, 048704 (2010).
- [11] A. Nardini, G. Pedà, and N. L. Rocca, *Trade-offs between leaf hydraulic capacity and drought vulnerability: morpho-anatomical bases, carbon costs and ecological consequences*, *New Phytologist* **196**, 788 (2012).
- [12] T. Nelson and N. Dengler, *Leaf Vascular Pattern Formation*, *The Plant Cell* **9**, 1121 (1997).
- [13] L. Koçillari, M. E. Olson, S. Suweis, R. P. Rocha, A. Lovison, F. Cardin, T. E. Dawson, A. Echeverría, A. Fajardo, S. Lechthaler, C. Martínez-Pérez, C. R. Marcati, K.-F. Chung, J. A. Rosell, A. Segovia-Rivas, C. B. Williams, E. Petrone-Mendoza, A. Rinaldo, T. Anfodillo, J. R. Banavar, and A. Maritan, *The Widened Pipe Model of plant hydraulic evolution*, *Proceedings of the National Academy of Sciences* **118**, e2100314118 (2021).
- [14] I. Rodríguez-Iturbe and A. Rinaldo, *Fractal River Basins: Chance and Self-Organization*, *Fractal River Basins: Chance and Self-organization* (Cambridge University Press, 1997).
- [15] A. Rinaldo, I. Rodriguez-Iturbe, R. Rigon, E. Ijjasz-Vasquez, and R. L. Bras, *Self-organized fractal river networks*, *Phys. Rev. Lett.* **70**, 822 (1993).
- [16] A. Konkol, J. Schwenk, E. Katifori, and J. B. Shaw, *Interplay of River and Tidal Forcings Promotes Loops in Coastal Channel Networks*, *Geophysical Research Letters* **49**, e2022GL098284 (2022).
- [17] F. Santambrogio, *Optimal channel networks, landscape function and branched transport*, *Interfaces and Free Boundaries* **9**, 149 (2007).

- [18] T. Nakagaki, H. Yamada, and Á. Tóth, *Maze-solving by an amoeboid organism*, *Nature* **407**, 470 (2000).
- [19] A. Tero, R. Kobayashi, and T. Nakagaki, *Physarum solver: A biologically inspired method of road-network navigation*, *Physica A: Statistical Mechanics and its Applications* **363**, 115 (2006), information and Material Flows in Complex Networks.
- [20] A. Tero, S. Takagi, T. Saigusa, K. Ito, D. P. Bebbler, M. D. Fricker, K. Yumiki, R. Kobayashi, and T. Nakagaki, *Rules for Biologically Inspired Adaptive Network Design*, *Science* **327**, 439 (2010).
- [21] B. Chazelle, Natural Algorithms, in *Proceedings of the 2009 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)* (Springer, Berlin, Heidelberg, 2009) pp. 422–431.
- [22] V. Bonifaci, K. Mehlhorn, and G. Varma, *Physarum can compute shortest paths*, *Journal of Theoretical Biology* **309**, 121 (2012).
- [23] V. Bonifaci, *Physarum can compute shortest paths: A short proof*, *Information Processing Letters* **113**, 4 (2013).
- [24] F. Kaiser, H. Ronellenfitsch, and D. Witthaut, *Discontinuous transition to loop formation in optimal supply networks*, *Nature communications* **11**, 1 (2020).
- [25] P. M. Pour, J. Standop, and S. K. Batra, *Are islet cells the gatekeepers of the pancreas?* *Pancreatology* **2**, 440 (2002).
- [26] H. Hong, J. Jo, and S.-J. Sin, *Stable and flexible system for glucose homeostasis*, *Phys. Rev. E* **88**, 032711 (2013).
- [27] J. B. Kirkegaard and K. Sneppen, *Optimal Transport Flows for Distributed Production Networks*, *Phys. Rev. Lett.* **124**, 208101 (2020).
- [28] C. Villani, *Optimal transport: Old and New*, Vol. 338 (Springer, 2009).

- [29] E. Facca, F. Cardin, and M. Putti, *Towards a Stationary Monge-Kantorovich Dynamics: The Physarum Polycephalum Experience*, *SIAM Journal on Applied Mathematics* **78**, 651 (2018).
- [30] E. Facca, F. Cardin, and M. Putti, *Physarum Dynamics and Optimal Transport for Basis Pursuit*, (2018).
- [31] E. Facca, S. Daneri, F. Cardin, and M. Putti, *Numerical Solution of Monge-Kantorovich Equations via a Dynamic Formulation*, *Journal of Scientific Computing* **82**, 68 (2020).
- [32] E. Facca, F. Cardin, and M. Putti, *Branching structures emerging from a continuous optimal transport model*, *Journal of Computational Physics* **447**, 110700 (2021).
- [33] E. Facca and M. Benzi, *Fast Iterative Solution of the Optimal Transport Problem on Graphs*, *SIAM Journal on Scientific Computing* **43**, A2295 (2021).
- [34] V. Bonifaci, E. Facca, F. Folz, A. Karrenbauer, P. Kolev, K. Mehlhorn, G. Morigi, G. Shahkarami, and Q. Vermande, *Physarum-inspired multi-commodity flow dynamics*, *Theoretical Computer Science* (2022), <https://doi.org/10.1016/j.tcs.2022.02.001>.
- [35] D. Baptista, D. Leite, E. Facca, M. Putti, and C. De Bacco, *Network extraction by routing optimization*, *Scientific Reports* **10**, 088702 (2020).
- [36] D. Baptista and C. De Bacco, in *Complex Networks & Their Applications X*, edited by R. M. Benito, C. Cherifi, H. Cherifi, E. Moro, L. M. Rocha, and M. Sales-Pardo (Springer International Publishing, Cham, 2022) pp. 578–592.
- [37] V. Bonifaci, *A Laplacian approach to  $\ell_1$ -norm minimization*, *Computational Optimization and Applications* **79**, 441 (2021).



- [38] G. Peyré and M. Cuturi, *Computational Optimal Transport: With Applications to Data Science*, *Foundations and Trends® in Machine Learning* **11**, 355 (2019).
- [39] M. Cuturi, in *Advances in Neural Information Processing Systems*, Vol. 26 (Curran Associates, Inc., 2013) pp. 2292–2300.
- [40] E. W. Dijkstra, *A note on two problems in connexion with graphs*, *Numerische Mathematik* **1**, 269 (1959).
- [41] R. Kay, A. Mattacchione, C. Katrycz, and B. D. Hatton, *Stepwise slime mould growth as a template for urban design*, *Scientific Reports* **12**, 1322 (2022).
- [42] A. I. Adamatzky, *Route 20, Autobahn 7, and Slime Mold: Approximating the Longest Roads in USA and Germany With Slime Mold on 3-D Terrains*, *IEEE Transactions on Cybernetics* **44**, 126 (2014).
- [43] A. Adamatzky and J. Jones, *Road planning with slime mould: If physarum built motorways it would route M6/M74 through Newcastle*, *International Journal of Bifurcation and Chaos* **20**, 3065 (2010).
- [44] K. Li, C. E. Torres, K. Thomas, L. F. Rossi, and C.-C. Shen, *Slime mold inspired routing protocols for wireless sensor networks*, *Swarm Intelligence* **5**, 183 (2011).
- [45] A. Tero, R. Kobayashi, and T. Nakagaki, *A mathematical model for adaptive transport network in path finding by true slime mold*, *Journal of Theoretical Biology* **244**, 553 (2007).
- [46] A. Tero, K. Yumiki, R. Kobayashi, T. Saigusa, and T. Nakagaki, *Flow-network adaptation in Physarum amoebae*, *Theory in Biosciences* **127**, 89 (2008).
- [47] D. Hu, D. Cai, and A. V. Rangan, *Blood Vessel Adaptation with Fluctuations in Capillary Flow Distribution*, *PLoS ONE* **7**, 1 (2012).

- [48] D. Hu and D. Cai, *Adaptation and Optimization of Biological Transport Networks*, *Phys. Rev. Lett.* **111**, 138701 (2013).
- [49] F. Corson, *Fluctuations and Redundancy in Optimal Transport Networks*, *Phys. Rev. Lett.* **104**, 048703 (2010).
- [50] H. Ronellenfitsch and E. Katifori, *Phenotypes of vascular flow networks*, *Phys. Rev. Lett.* **123**, 248101 (2019).
- [51] H. Ronellenfitsch and E. Katifori, *Global Optimization, Local Adaptation, and the Role of Growth in Distribution Networks*, *Phys. Rev. Lett.* **117**, 138301 (2016).
- [52] A. A. Ibrahim, D. Leite, and C. De Bacco, *Sustainable optimal transport in multilayer networks*, *Physical Review E* **105**, 064302 (2022).
- [53] D. Leite and C. De Bacco, *Revealing the similarity between urban transportation networks and optimal transport-based infrastructures*, (2022), [arXiv:2209.06751](https://arxiv.org/abs/2209.06751) [physics.soc-ph].
- [54] D. Leite, D. Baptista, A. A. Ibrahim, E. Facca, and C. De Bacco, *Community detection in networks by dynamical optimal transport formulation*, *Scientific Reports* **12**, 16811 (2022).
- [55] F. Folz, K. Mehlhorn, and G. Morigi, *Interplay of periodic dynamics and noise: Insights from a simple adaptive system*, *Phys. Rev. E* **104**, 054215 (2021).
- [56] F. Folz, K. Mehlhorn, and G. Morigi, *Noise-Induced Network Topologies*, *Phys. Rev. Lett.* **130**, 267401 (2023).
- [57] D. Baptista and C. De Bacco, *Principled network extraction from images*, *R. Soc. open sci.* **8**, 210025 (2021).
- [58] G. Monge, *Mémoire sur la théorie des déblais et des remblais*, (1781), *histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année.*

- [59] A. Tolstoi, *Metody nakhozheniya naimen'shego summovogo kilometrazha pri planirovanii perezozok v prostranstve*, Planirovanie Perekozok, Sbornik pervyi, Transpechat' NKPS , 23 (1930).
- [60] F. L. Hitchcock, *The Distribution of a Product from Several Sources to Numerous Localities*, *Journal of Mathematics and Physics* **20**, 224 (1941).
- [61] L. Kantorovich, *On the transfer of masses (in Russian)*, (1942).
- [62] M. Cuturi and D. Avis, *Ground Metric Learning*, *Journal of Machine Learning Research* **15**, 533 (2014).
- [63] F. Santambrogio, *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling* (Birkhäuser Cham, 2015).
- [64] P. Demetci, R. Santorella, B. Sandstede, W. S. Noble, and R. Singh, *Gromov-Wasserstein optimal transport to align single-cell multiomics data*, *bioRxiv* (2020), 10.1101/2020.04.28.066787.
- [65] C. Bunne, L. Papaxanthos, A. Krause, and M. Cuturi, in *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, *Proceedings of Machine Learning Research*, Vol. 151, edited by G. Camps-Valls, F. J. R. Ruiz, and I. Valera (PMLR, 2022) pp. 6511–6528.
- [66] C. Bunne, S. G. Stark, G. Gut, J. S. del Castillo, M. Levesque, K.-V. Lehmann, L. Pelkmans, A. Krause, and G. Rätsch, *Learning single-cell perturbation responses using neural optimal transport*, *Nature Methods* (2023), 10.1038/s41592-023-01969-x.
- [67] G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, J. Gould, S. Liu, S. Lin, P. Berube, L. Lee, J. Chen, J. Brumbaugh, P. Rigollet, K. Hochedlinger, R. Jaenisch, A. Regev, and E. S. Lander, *Optimal-Transport Analysis of Single-cell Gene Expression Identifies Developmental Trajectories in Reprogramming*, *Cell* **176**, 928 (2019).

- [68] O. Pele and M. Werman, in *Computer Vision – ECCV 2008* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008) pp. 495–508.
- [69] O. Pele and M. Werman, in *2009 IEEE 12th International Conference on Computer Vision* (2009) pp. 460–467.
- [70] S. Peleg, M. Werman, and H. Rom, *A unified approach to the change of resolution: space and gray-level*, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**, 739 (1989).
- [71] M. Werman, S. Peleg, and A. Rosenfeld, *A Distance Metric for Multidimensional Histograms*, *Computer Vision, Graphics, and Image Processing* **32**, 328 (1985).
- [72] Y. Rubner, C. Tomasi, and L. J. Guibas, in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)* (1998) pp. 59–66.
- [73] Y. Rubner, C. Tomasi, and L. J. Guibas, *The Earth Mover’s Distance as a Metric for Image Retrieval*, *International Journal of Computer Vision* **40**, 99 (2000).
- [74] E. Aurell, C. Mejía-Monasterio, and P. Muratore-Ginanneschi, *Optimal Protocols and Optimal Transport in Stochastic Thermodynamics*, *Phys. Rev. Lett.* **106**, 250601 (2011).
- [75] A. Genevay, M. Cuturi, G. Peyré, and F. Bach, in *Advances in Neural Information Processing Systems*, Vol. 29, edited by D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (Curran Associates, Inc., 2016).
- [76] J. Altschuler, J. Niles-Weed, and P. Rigollet, in *Advances in Neural Information Processing Systems*, Vol. 30, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., 2017).

- [77] T. Lin, N. Ho, and M. Jordan, in *Proceedings of the 36th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 97 (PMLR, 2019) pp. 3982–3991.
- [78] P. Dvurechensky, A. Gasnikov, and A. Kroshnin, in *Proceedings of the 35th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 80 (PMLR, 2018) pp. 1367–1376.
- [79] M. Eisenberger, A. Toker, L. Leal-Taixé, F. Bernard, and D. Cremers, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022) pp. 509–518.
- [80] R. Flamary, N. Courty, A. Gramfort, M. Z. Alaya, A. Boisbunon, S. Chambon, L. Chapel, A. Corenflos, K. Fatras, N. Fournier, L. Gautheron, N. T. Gayraud, H. Janati, A. Rakotomamonjy, I. Redko, A. Rolet, A. Schutz, V. Seguy, D. J. Sutherland, R. Tavenard, A. Tong, and T. Vayer, *POT: Python Optimal Transport*, *Journal of Machine Learning Research* **22**, 1 (2021).
- [81] M. Cuturi, L. Meng-Papaxanthos, Y. Tian, C. Bunne, G. Davis, and O. Teboul, Optimal Transport Tools (OTT): A JAX Toolbox for all things Wasserstein, (2022), [2201.12324](https://arxiv.org/abs/2201.12324).
- [82] D. Klein, G. Palla, M. Lange, M. Klein, Z. Piran, M. Gander, L. Meng-Papaxanthos, M. Sterr, A. Bastidas-Ponce, M. Tarquis-Medina, H. Lickert, M. Bakhti, M. Nitzan, M. Cuturi, and F. J. Theis, *Mapping cells through time and space with moscot*, *bioRxiv* (2023), [10.1101/2023.05.11.540374](https://doi.org/10.1101/2023.05.11.540374).
- [83] M. Arjovsky, S. Chintala, and L. Bottou, in *Proceedings of the 34th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 70, edited by D. Precup and Y. W. Teh (PMLR, 2017) pp. 214–223.
- [84] A. Genevay, G. Peyre, and M. Cuturi, in *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*,

- Proceedings of Machine Learning Research, Vol. 84, edited by A. Storkey and F. Perez-Cruz (PMLR, 2018) pp. 1608–1617.
- [85] T. Salimans, H. Zhang, A. Radford, and D. N. Metaxas, in *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings* (2018).
- [86] K. Sinclair and R. C. Ball, *Mechanism for Global Optimization of River Networks from Local Erosion Rules*, *Phys. Rev. Lett.* **76**, 3360 (1996).
- [87] S. Bohn and M. O. Magnasco, *Structure, Scaling, and Phase Transition in the Optimal Transport Network*, *Phys. Rev. Lett.* **98**, 088702 (2007).
- [88] A. Rinaldo, I. Rodriguez-Iturbe, R. Rigon, R. L. Bras, E. Ijjasz-Vasquez, and A. Marani, *Minimum energy and fractal structures of drainage networks*, *Water Resources Research* **28**, 2183 (1992).
- [89] Q. Xia, *Optimal paths related to transport problems*, *Communications in Contemporary Mathematics* **5**, 251 (2003).
- [90] Q. Xia, *The formation of a tree leaf*, *ESAIM: Control, Optimisation and Calculus of Variations* **13**, 359 (2007).
- [91] <https://github.com/aleable/McOpt>.
- [92] <https://github.com/cdebacco/MultiOT>.
- [93] <https://github.com/aleable/MODI>.
- [94] <https://github.com/aleable/N-STARK>.
- [95] <https://github.com/aleable/BROT>.
- [96] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard, *Scaling algorithms for unbalanced optimal transport problems*, *Mathematics of Computation* **87**, 2563 (2018).

- [97] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard, *Unbalanced optimal transport: Dynamic and Kantorovich formulations*, *Journal of Functional Analysis* **274**, 3090 (2018).
- [98] M. Newman, *Networks* (Oxford university press, 2010).
- [99] G. Bianconi, *Multilayer networks: structure and function* (Oxford university press, 2018).
- [100] S. P. Suter and R. Skalak, *The History of Poiseuille’s Law*, *Annual Review of Fluid Mechanics* **25**, 1 (1993).
- [101] D. J. Klein and M. Randić, *Resistance distance*, *Journal of Mathematical Chemistry* **12**, 81 (1993).
- [102] K. Salimifard and S. Bigharaz, *The multicommodity network flow problem: state of the art classification, applications, and solution methods*, *Operational Research* , 1 (2020).
- [103] Y.-Z. Xu, H. F. Po, C. H. Yeung, and D. Saad, *Scalable node-disjoint and edge-disjoint multiwavelength routing*, *Phys. Rev. E* **105**, 044316 (2022).
- [104] T. S. Tai and C. H. Yeung, *Optimally coordinated traffic diversion by statistical physics*, *Phys. Rev. E* **104**, 024311 (2021).
- [105] C. H. Yeung and D. Saad, *Competition for shortest paths on sparse graphs*, *Phys. Rev. Lett.* **108**, 208701 (2012).
- [106] C. H. Yeung, D. Saad, and K. Y. M. Wong, *From the physics of interacting polymers to optimizing routes on the London Underground*, *Proceedings of the National Academy of Sciences* **110**, 13717 (2013).
- [107] F. Altarelli, A. Braunstein, L. Dall’Asta, C. De Bacco, and S. Franz, *The Edge-Disjoint Path Problem on Random Graphs by Message-Passing*, *PLOS ONE* **10**, 1 (2016).

- [108] C. De Bacco, S. Franz, D. Saad, and C. H. Yeung, *Shortest node-disjoint paths on random graphs*, *Journal of Statistical Mechanics: Theory and Experiment* **2014**, P07009 (2014).
- [109] P. M. Dixon, J. Weiner, T. Mitchell-Olds, and R. Woodley, *Bootstrapping the Gini Coefficient of Inequality*, *Ecology* **68**, 1548 (1987).
- [110] T. A. Davis, *Algorithm 832: UMFPACK V4.3—an Unsymmetric-Pattern Multifrontal Method*, *ACM Trans. Math. Softw.* **30**, 196–199 (2004).
- [111] U. Trottenberg, C. W. Oosterlee, and A. Schuller, *Multigrid* (Elsevier Science, 2001).
- [112] B. Folkow, O. Lundgren, and I. Wallentin, *Studies on the relationship between flow resistance, capillary filtration coefficient and regional blood volume in the intestine of the cat*, *Acta physiologica scandinavica* **57**, 270 (1963).
- [113] D. N. Granger, P. R. Kvietyts, and M. A. Perry, *Role of exchange vessels in the regulation of intestinal oxygenation*, *American Journal of Physiology-Gastrointestinal and Liver Physiology* **242**, G570 (1982).
- [114] R. J. Widmer, R. H. Stewart, M. F. Young, J. E. Laurinec, G. A. Laine, and C. M. Quick, *Application of local heat induces capillary recruitment in the pallid bat wing*, *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **292**, R2312 (2007).
- [115] “Transport for London open data”, accessed: 2021-09-20 (2021).
- [116] H. Youn, M. T. Gastner, and H. Jeong, *Price of Anarchy in Transportation Networks: Efficiency and Optimality Control*, *Phys. Rev. Lett.* **101**, 128701 (2008).
- [117] M. T. Gastner and M. E. J. Newman, *Optimal design of spatial distribution networks*, *Phys. Rev. E* **74**, 016117 (2006).



- [118] R. Selten, T. Chmura, T. Pitz, S. Kube, and M. Schreckenberg, *Commuters route choice behaviour*, *Games and Economic Behavior* **58**, 394 (2007).
- [119] R. J. Smeed, *Studies in the Economics of Transportation*, *The Economic Journal* **67**, 116 (1957).
- [120] B. Li, D. Saad, and C. H. Yeung, *Bilevel optimization in flow networks: A message-passing approach*, *Phys. Rev. E* **106**, L042301 (2022).
- [121] H. F. Po, C. H. Yeung, and D. Saad, *Futility of being selfish in optimized traffic*, *Phys. Rev. E* **103**, 022306 (2021).
- [122] C. H. Yeung, *Coordinating dynamical routes with statistical physics on space-time networks*, *Phys. Rev. E* **99**, 042123 (2019).
- [123] V. Colizza, J. R. Banavar, A. Maritan, and A. Rinaldo, *Network structures from selection principles*, *Phys. Rev. Lett.* **92**, 198701 (2004).
- [124] T. S. Tai and C. H. Yeung, *Global benefit of randomness in individual routing on transportation networks*, *Phys. Rev. E* **100**, 012311 (2019).
- [125] T. S. Tai and C. H. Yeung, *Adaptive strategies for route selection en-route in transportation networks*, *Chinese Journal of Physics* **77**, 712 (2022).
- [126] G. Yan, T. Zhou, B. Hu, Z.-Q. Fu, and B.-H. Wang, *Efficient routing on complex networks*, *Phys. Rev. E* **73**, 046108 (2006).
- [127] E. Koutsoupias and C. Papadimitriou, *Worst-case equilibria*, *Computer Science Review* **3**, 65 (2009).
- [128] J. W. Rocks, H. Ronellenfitsch, A. J. Liu, S. R. Nagel, and E. Katifori, *Limits of multifunctionality in tunable networks*, *Proceedings of the National Academy of Sciences* **116**, 2506 (2019).

- [129] G. H. Golub and V. Pereyra, *The Differentiation of Pseudo-Inverses and Nonlinear Least Squares Problems Whose Variables Separate*, *SIAM Journal on Numerical Analysis* **10**, 413 (1973).
- [130] M. Muehlebach and M. I. Jordan, *On Constraints in First-Order Optimization: A View from Non-Smooth Dynamical Systems*, *Journal of Machine Learning Research* **23**, 1 (2022).
- [131] A. A. Ibrahim, M. Muehlebach, and C. D. Bacco, *Optimal transport with constraints: from mirror descent to classical mechanics*, (2023), [arXiv:2309.04727](https://arxiv.org/abs/2309.04727) [physics.soc-ph].
- [132] T. Roughgarden and E. Tardos, *How Bad is Selfish Routing?* *J. ACM* **49**, 236–259 (2002).
- [133] D. Braess, *Über ein Paradoxon aus der Verkehrsplanung*, *Unternehmensforschung* **12**, 258 (1968).
- [134] J. E. Cohen and P. Horowitz, *Paradoxical behaviour of mechanical and electrical networks*, *Nature* **352**, 699 (1991).
- [135] B. Schäfer, T. Pesch, D. Manik, J. Gollenstede, G. Lin, H.-P. Beck, D. Witthaut, and M. Timme, *Understanding Braess’ Paradox in power grids*, *Nature Communications* **13**, 5396 (2022).
- [136] S. Bittihn and A. Schadschneider, *Braess’ paradox in the age of traffic information*, *Journal of Statistical Mechanics: Theory and Experiment* **2021**, 033401 (2021).
- [137] G. Kolata, *What if They Closed 42d Street and Nobody Noticed?* *New York Times* (1990).
- [138] R. Kujala, C. Weckström, R. K. Darst, M. N. Mladenović, and J. Saramäki, *A collection of public transport network data sets for 25 cities*, *Scientific data* **5**, 180089 (2018).
- [139] “Trafic annuel entrant par station du réseau ferré 2019”, accessed: 2020-08-28 (2019).

- [140] M. Seeland, M. Rzanny, N. Alaqraa, J. Wäldchen, and P. Mäder, *Jena Flowers 30 Dataset*, (2017).
- [141] P. A. Macanhã, D. M. Eler, R. E. Garcia, and W. E. M. Junior, in *Information Technology - New Generations* (Springer International Publishing, Cham, 2018) pp. 699–705.
- [142] R. G. Morris and M. Barthelemy, *Transport on Coupled Spatial Networks*, *Phys. Rev. Lett.* **109**, 128703 (2012).
- [143] K. Grauman and T. Darrell, in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004*, Vol. 1 (2004) pp. I–I.
- [144] P. Koehl, M. Delarue, and H. Orland, *Optimal transport at finite temperature*, *Phys. Rev. E* **100**, 013310 (2019).
- [145] M. Thorpe, S. Park, S. Kolouri, G. K. Rohde, and D. Slepčev, *A transportation  $L^p$  Distance for Signal Analysis*, *Journal of mathematical imaging and vision* **59**, 187 (2017).
- [146] R. N. Shepard, *Toward a Universal Law of Generalization for Psychological Science*, *Science* **237**, 1317 (1987).
- [147] B. Schmitzer, *Stabilized Sparse Scaling Algorithms for Entropy Regularized Transport Problems*, *SIAM Journal on Scientific Computing* **41**, A1443 (2019).