Routing Optimization for Transport and Sustainability

Dissertation

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

vorgelegt von
Abdullahi Adinoyi Ibrahim, M. Sc.
aus Okene, Nigerian

Tübingen
2024

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

Tag der mündlichen Qualifikation:	19.07.2024
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatterin:	Dr. Caterina De Bacco
2. Berichterstatter:	Prof. Dr. Robert Bamler

Disclaimer: this thesis adopts the template designed by Felix Dangel, which draws its structure from Federico Marotta's `kaobook` template.

Acknowledgments

I would like to express my deepest gratitude to my esteemed Ph.D. supervisor, Caterina De Bacco, for her unwavering patience, continuous support, and invaluable guidance throughout my doctoral journey. Your mentorship has been instrumental in shaping my academic and personal growth. Your dedication, tireless effort and belief in my potential has been a constant source of motivation, inspiring me to overcome challenges and strive for excellence. I am truly fortunate to have had such an exceptional mentor who not only imparted knowledge but also fostered a nurturing and conducive learning environment.

To my PhD co-advisor, Gabriele Schweikert, my gratitude for her mentorship, dedication and significant contributions during my thesis advisory commitee meetings. I truly appreciate your time and efforts. I would also like to appreciate Robert Bamler for his regular support and his participation in reviewing this thesis and as one of the examination commitee. My extended gratitude to one of my thesis advisory committee; Nicole Ludwig, and to examination commitees; Anna Levina and Georg Martius. I appreciate your commitments and contribution to the academic community.

I would like to take a moment to express my sincere appreciation for the invaluable support and funding provided by the International Max Planck Research School for Intelligent Systems (IMPRS-IS). Your investment in cutting-edge and quality research has been instrumental in enabling me to pursue my research goals and advance in my field.

I would like to express my deepest gratitude to my incredible lab colleagues, whose unwavering support, camaraderie and collaboration have made my journey truly memorable. To each of you, from different corners of the globe, thank you for bringing your unique perspectives, talents and experiences to our shared workspace. Your international backgrounds have enriched our discussions, broadened my horizons and fostered an environment of inclusivity and mutual respect.

To my loving wife Rabiat Ibrahim and our precious son; Imran Ibrahim, your unwavering love, understanding and support have been the cornerstone of my success. Your sacrifices, encouragement and patience have sustained me through the ups and downs of this journey. Your belief in me has been my greatest motivation and I am forever grateful for your unwavering faith and endless encouragement.

Thank you!

Abdullahi Ibrahim
Tübingen, April 10, 2024

Abstract

Traffic congestion is a major challenge in the transport industry, affecting both the economy and environment. Designing efficient and sustainable transport models requires a multifaceted approach. One of these facets is extracting optimal trajectories for each passenger type, a task well-addressed by the principles of optimal transport theory. By leveraging optimal transport principles, we can model passenger flows in networks to reduce congestion. However, recent research based on optimal transport overlooks crucial factors such as environmental impacts, multilayer transport network analysis, and fails to consider practical constraints such as road capacity limitations.

In response to these gaps, this thesis introduces optimal transport-based methods for modeling flows within multilayer transport networks, with a primary focus on addressing congestion and optimizing traffic flow. Additionally, we extend the application of optimal transport theory to tackle community detection problems within networks. This broader scope allows us to not only enhance our understanding of traffic dynamics but also explore diverse applications of optimal transport in networks.

First, we propose efficient methods, based on optimal transport theory, for modeling passenger flows within multilayer transport networks. Our approach generates both distributed and single-trajectory flows for each passenger types, and shows how these trajectories can alleviate traffic congestion and reduce CO2 emissions. Second, to address the limitation of existing methods on realistic constraints in transport network, we delve into a constrained framework. This framework accommodates nonlinear and non-convex constraints within optimal transport problems, providing a computationally efficient tool for minimizing congestion. As an application, we consider real multilayer transport networks where each layer is associated with a different transport mode, and show how the traffic distribution varies with relevant quantities (such as transport regime, origin-destination pairs, imposed constraints, etc.) across layers.

Lastly, we present an optimal transport-based approach for detecting communities in networks. By incorporating the Ollivier-Ricci curvature, our model provides various transport regimes that allow for better control of information flow between node neighborhoods. The algorithm not only exhibits improved accuracy in identifying communities, but also outperforms conventional OT-based methods, providing deeper insights into geometric approaches to analyzing complex networks.

Overall, the methods presented in this thesis enhance our understanding of traffic dynamics within multilayer transport networks, provides valuable insights that contribute to sustainable transport systems. By addressing congestion through optimal transport-based approaches, we pave the way for more efficient and environmentally friendly transport systems. Furthermore, extending the application of optimal transport to community detection problems highlights its versatility in analyzing complex networks beyond transportation networks.

Zusammenfassung

Verkehrsüberlastung ist eine große Herausforderung für die Verkehrsbranche, die sowohl die Wirtschaft als auch die Umwelt beeinträchtigt. Die Entwicklung effizienter und nachhaltiger Verkehrsmodelle erfordert einen vielschichtigen Ansatz. Eine dieser Facetten ist die Extraktion optimaler Trajektorien für jeden Passagiertyp, eine Aufgabe, die durch die Prinzipien der opti- malen Verkehrstheorie gut gelöst wird. Durch die Nutzung der Grundsätze des optimalen Verkehrs können wir die Passagierströme in Netzwerken modellieren, um Staus zu verringern. In der jüngsten Forschung zu den Grundlagen des optimalen Verkehrs werden jedoch entscheidende Faktoren wie die Auswirkungen auf die Umwelt und die Analyse mehrschichtiger Verkehrsnetze außer Acht gelassen, und praktische Einschränkungen wie die begrenzte Straßenkapazität werden nicht berücksichtigt.

Als Antwort auf diese Lücken werden in dieser Arbeit Methoden zur Modellierung von Verkehrsflüssen in mehrschichtigen Verkehrsnetzen vorgestellt. Diese basieren auf optimalem Transport. Dabei liegt der Schwerpunkt auf der Bewältigung von Verkehrsüberlastungen und der Optimierung des Verkehrsflusses. Darüber hinaus erweitern wir die Anwendung der Theorie des optimalen Transports, um Probleme der Erkennung von Gemeinschaften in Netzwerken anzugehen. Durch diese Erweiterung können wir nicht nur unser Verständnis der Verkehrsdynamik verbessern, sondern auch neue Anwendungen des optimalen Transports in Netzwerken erforschen.

Zunächst schlagen wir effiziente, auf der Theorie des optimalen Transports basierende Metho- den zur Modellierung von Passagierströmen in mehrschichtigen Verkehrsnetzen vor. Unser Ansatz generiert sowohl verteilte als auch einzelne Trajektorien für jeden Passagiertyp und zeigt, wie diese Trajektorien Verkehrsstaus entschärfen und CO2-Emissionen reduzieren können. Zweitens: Um die Beschränkung bestehender Methoden auf realistische Beschränkungen in Verkehrsnetzen zu beheben, haben wir einen Rahmen mit Beschränkungen entwickelt. Dieser Rahmen berücksichtigt nichtlineare und nichtkonvexe Beschränkungen in optimalen Verkehrsproblemen und bietet ein rechnerisch effizientes Instrument zur Minimierung von Staus. Als Anwendung betrachten wir reale mehrschichtige Transportnetze, bei denen jede Schicht mit einem anderen Transportmodus verbunden ist, und zeigen, wie die Verkehrsverteilung mit einigen Größen (wie z.B. Transportregime, Start-Ziel-Paare, auferlegte Beschränkungen usw.) über die Schichten hinweg variiert.

Schließlich stellen wir einen optimalen Transport-basierten Ansatz zur Erkennung

von Gemein- schaften in Netzen vor. Durch die Einbeziehung der Ollivier-Ricci-Krümmung bietet unser Modell verschiedene Transportregime, die eine bessere Kontrolle des Informationsflusses zwischen den Nach- barschaften der Knoten ermöglichen. Der Algorithmus weist nicht nur eine verbesserte Genauigkeit bei der Identifizierung von Gemeinschaften auf, sondern übertrifft auch herkömmliche OT-basierte Methoden und bietet tiefere Einblicke in geometrische Ansätze zur Analyse komplexer Netzwerke.

Insgesamt verbessern die in dieser Arbeit vorgestellten Methoden unser Verständnis der Verkehrs- dynamik in mehrschichtigen Verkehrsnetzen und liefern wertvolle Erkenntnisse, die zu nachhaltigen Verkehrssystemen beitragen. Indem wir Staus mit Hilfe von auf optimalem Verkehr basierenden Ansätzen angehen, ebnen wir den Weg für effizientere und umweltfreundlichere Verkehrssysteme. Darüber hinaus wird durch die Ausweitung der Anwendung des optimalen Verkehrs auf Probleme der Gemeinschaftserkennung seine Vielseitigkeit bei der Analyse komplexer Netzwerke über das Verkehrsnetz hinaus deutlich.

Table of Contents

Acknowledgments **v**

Abstract **viii**

Zusammenfassung **x**

Table of Contents **xiv**

1 Introduction **1**
1.1 Overview and motivation 1
1.2 Outline 3

2 Background **4**
2.1 Transport networks as Graphs 4
2.2 Flux and Conductivity 5
2.3 Adaptative Dynamics 7
2.4 Transport cost 8

3 Challenges **10**
3.1 Multilayer transport network 10
3.2 Paths distribution 11
3.3 Constrained dynamics 12

4 Published Work **14**
4.1 Optimal transport in multilayer networks for traffic flow optimization 15
4.2 Sustainable optimal transport in multilayer networks 16
4.3 Optimal transport with constraints: from mirror descent to classical mechanics 17
4.4 Community detection in networks by dynamical optimal transport formulation 18

5 Discussion and Conclusion **20**
5.1 Optimal transport in multilayer networks 20
5.2 Sustainable optimal transport 21
5.3 Constrained optimal transport 22
5.4 Application on community detection problem 26
5.5 Conclusion 26

Bibliography **28**

Appendix A: Complete Publications 29

Introduction 1

In this cumulative thesis, I present the summary of my Ph.D. study through the presentation of four papers. Of these, three are first-author contributions (cited as [36–38]) and one is a collaborative paper [48], all published during my study with the *Physics for Inference and Optimization* group at the Max Planck Institute for Intelligent Systems, Tübingen.

1.1 Overview and motivation

Traffic congestion is a growing problem around the world due to its impacts on the economy (such as cost and less productivity) and the environment (such as carbon emission). Traffic congestion is estimated to cost around €3.9 billion in Germany, £9.5 billion in the UK, and $81 billion in the US [58], taking into account both additional fuel consumption and hours lost due to congestion. Recent research still lacks some important factors, such as i) solving this problem without considering the environmental impact, ii) without using a real transport network, and iii) without considering realistic constraints such as the use of roads with a limited capacity of vehicles traveling at the same time. Given the challenges of congestion, it is important to explore realistic strategies to alleviate this problem. The relevant question remains: how can we simultaneously decrease costs, enhance efficiency, and alleviate congestion within transport systems?

Designing reliable and efficient transport systems requires a multi-faceted approach. One of these facets is to provide multimodal transport systems to reduce dependence on one mode of transport. This multimodal system, depending on the city infrastructure, can either have individual infrastructure (for example bus and tram system with individual paths), or shared infrastructure (the bus and tram networks having shared infrastructure at several points). Expanding infrastructure can alleviate congestion but this also incurs it own monetary and environmental costs [59]. Another facet is to implements coordinated routing systems for individual passengers [51], considering that passengers may not always choose the shortest route but rather the most convenient one [61, 79]. While the implementation of coordinated routing systems reduces congestion and environmental costs [22, 54, 77], it also leads to a corresponding increase in the average path length traveled by passengers [76].

Several approaches have tackled the congestion problem by using various methods to generate routes for each passenger. These methods include the shortest path

minimization [6, 47, 69, 70], diffusion processes [17, 18, 33, 34], assignment strategies [32], adaptive dynamics [35, 63, 73] and other physics of complex systems approaches [46, 64, 68, 78].

However, these approaches did not (sufficiently) describe some important scenarios, such as i) the impact of multilayer networks on traffic and their potential for optimizing congestion, ii) the potential for reducing environmental costs, and iii) instances where transport flows on network edges (i.e., roads) are restricted by capacity constraints, resembling real-world scenarios where roads have limited capacity. To address these scenarios, we develop algorithm relying on the optimal transport theory [41]. Optimal Transport (OT) has been used to model and optimize various transport networks, including the network design [5, 63, 73] and traffic flow [9, 10, 49]. These methods, suitable for both single-layer [50] and multilayer transport networks [37], provide a principled and computationally efficient approach to solving transport problems within network structures.

While previous algorithms aimed at alleviating congestion have made strides, there remains a significant gap in understanding how an efficient tool such as optimal transport, can be effectively employed to reduce congestion. The current body of research lacks a comprehensive exploration of the interplay between multilayer transport networks, optimal transport with realistic constraints, and environmental factors. Bridging this gap is the main motivation behind this thesis, seeking to unravel and showcase the connections that can minimize traffic congestion. More concretely, this thesis addresses the following: i) we develop methods for extracting flows in a multilayer transport network [36, 37] and how these methods ease congestions and environmental costs, ii) we develop a principled physics-based approach to impose constraints flexibly in optimal transport problems [38], and iii) beyond transportation networks, we address community detection problem by developing a dynamical OT method to group nodes based on the encoded information [48].

1.2 Outline

The remainder of the thesis is structured in four chapters as follows.

Chapter 2 lay the groundwork, providing the essential background and preliminaries. Section 2.1 introduces graphs and how transport networks can be represented. Section 2.2 describes the two main quantitues associated with network edges and relevant for the proposed OT models. Section 2.3 describes the general adaptive dynamics for multilayer networks, and the multilayer transport cost function is describe in Section 2.4.

Chapter 3 describes the real challenges and motivates why we address them. Section 3.1 describes the complexities of multilayer transport network. Section 3.2 and Section 3.3 focuses on optimal transport-based approaches. Section 3.2 describes the challenges of having a distributed trajectory for each passengers, and Section 3.3 discusses the lack of realistic constraints on the classical OT methods.

Chapter 4 is a central hub that summarizes the published papers, suggested reading order and contribution. This section includes the following papers:

- A. A. Ibrahim, A. Lonardi, and C. D. Bacco. "Optimal transport in multilayer networks for traffic flow optimization". *Algorithms* 14.7 (2021). [37]
- A. A. Ibrahim, D. Leite, and C. De Bacco. "Sustainable optimal transport in multilayer networks". *Physical Review E* 105.6 (2022). [36]
- A. A. Ibrahim, M. Muehlebach, and C. De Bacco. "Optimal transport with constraints: from mirror descent to classical mechanics". *Physical Review Letters* 133 (2024). [38]
- 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 (2022). [48]

Chapter 5 discusses the findings of this thesis, based on the published articles, as well as their impacts on the development of optimal transport.

2 Background

This chapter introduces the relevant concepts to provide the reader with the general knowledge to understand the context, relevance and importance of the contributions of each paper.

The following notations will be used throughout this chapter. The set of traffic demands or masses is denoted as $S = \{(o_a, t_a)\}$, and $|S| = M$ is the number of masses. The Euclidean length of an edge e is represented with L_e.

2.1 Transport networks as Graphs

A single-layer network is a transport network that represents a single mode of transport. It can represent a bus or metro network in a city. Single-layer networks can be used to understand traffic dynamics, for example, [16, 49, 50, 59] used a single-layer network to study traffic flow. A single-layer network can be defined as a graph $\mathcal{G}(\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ and $\mathcal{E}$ represent the set of nodes and edges in graph $\mathcal{G}$. A single-layer graph can be either directed or undirected. The information in graph $\mathcal{G}$ can be encoded in an adjacency matrix σ of $\mathcal{N} \times \mathcal{N}$ dimensional, with entries:

$$\sigma_{uv} = \begin{cases} 1 & \text{if there is edge between nodes } u \text{ and } v \\ 0 & \text{otherwise} \quad . \end{cases} \tag{2.1}$$

If $\mathcal{G}$ is an undirected graph, the adjacency matrix can be defined as $\sigma_{uv} = 1$ if and only if $\sigma_{vu} = 1$, that is, $\sigma_{uv} = \sigma_{vu} = 1$.

A multilayer graph on the other hand represents multiple modes of transport that is combined into one framework. Unlike single-layer, representing a multilayer network as a graph requires more parameters. A multilayer graph can be defined as graph $\bar{\mathcal{G}}(\{\mathcal{N}_\alpha\}_\alpha, \{\mathcal{E}_\alpha\}_\alpha, \{\mathcal{E}_{\alpha\gamma}\}_{\alpha\gamma})$, where $\mathcal{N}_\alpha$ and $\mathcal{E}_\alpha$ represent the set of nodes and edges in layer α, and $\mathcal{E}_{\alpha\gamma}$ is the set of edges connecting shared nodes between layer α and γ. There are two types of edges in a multilayer network; intralayer and interlayer edges[1]. One can represent a multilayer network with two tensors; the intralayer tensor, denoted with σ^α_{uv} is the adjacency for intralayer edges. $\sigma^\alpha_{uv} = 1$ indicates there is a connection between nodes u and v in layer α, and $\sigma^\alpha_{uv} = 0$

[1] Intralayer are edges within a layer. Interlayer edges are edges connecting different layers. For example, an inter-layer edge can be thought of as the stairs that connect the subway entrance to the train station entrance.

otherwise. Then the interlayer tensor, written as $\sigma_{uv}^{\alpha\gamma}$ is the adjacency between layers, where $\sigma_{uv}^{\alpha\gamma} = 1$ indicates connection between nodes u and v in layers α and γ, and $\sigma_{uv}^{\alpha\gamma} = 0$ otherwise. Fig. 2.1 depicts an example of a multilayer network where the adjacency matrix of each layer is combined into a large framework referred to as the Supra-adjacency matrix.

To determine the number of nodes and edges, the total set of nodes and edges in a single-layer network are $N = |\mathcal{N}|$ and $E = |\mathcal{E}|$, respectively, whereas in a multilayer graph, there are $N = \bigcup_\alpha |\mathcal{N}_\alpha|$ nodes and $E = (\bigcup_\alpha \mathcal{E}_\alpha) \bigcup (\bigcup_{\alpha\gamma} \mathcal{E}_{\alpha\gamma})$ edges.

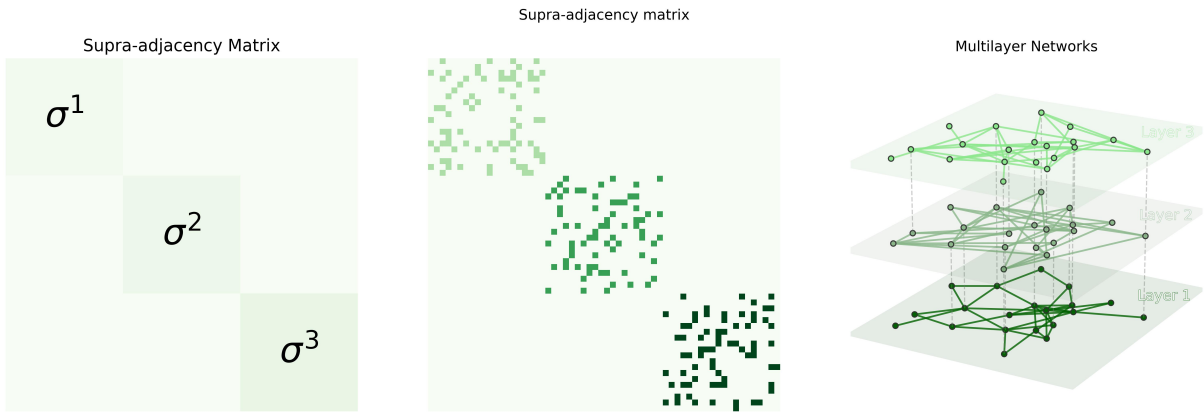


Figure 2.1: Example of multilayer graphs. (Left-Middle) represents the supra-adjacency for each layer. (Right) A multilayer graph with $\mathcal{N}_\alpha = 20$ and $\sum_\alpha \mathcal{E}_\alpha = 119$ edges.

Multilayer networks play a crucial role in traffic congestion by taking into account the interactions between different layers, these networks offer the possibility of fine-tuning traffic flows. This multilayer configuration not only improves the overall efficiency of transport systems, but also contributes to a more balanced distribution of traffic. Furthermore, a multilayer network approach goes beyond traditional methods and provides a framework that is realistic and adaptable. This approach will be instrumental in gaining a comprehensive understanding of the dynamics of traffic congestion and, consequently, in formulating reliable methods to reduce it. The adaptability of the multilayer framework ensures that interventions and policies can be tailored to specific challenges, making it an invaluable tool in the quest for sustainable and efficient transport systems.

2.2 Flux and Conductivity

We now describe two main quantities associated with network edges that would be relevant for modeling with optimal transport:

i) the passenger flux denoted as Q_e^a, to represent the volume of passenger type a traversing a given edge e, and
ii) conductivities denoted as D_e, these are proportional to the size of an edge e. The conductivity plays a crucial role in regulating the passenger flux passing through an edge e.

To express this mathematically, the flow of mass type a traveling along an edge (u, v) can be defined as:

$$Q_{u,v}^a := \frac{D_{u,v}}{L_{u,v}} \left(p_u^a - p_v^a\right) \quad a \in S, (u, v) \in E \quad , \tag{2.2}$$

where $D_{u,v}$ and $L_{u,v}$ denote the conductivity and length of edge $e = (u, v)$, respectively. The quantities p_u^a and p_v^a are the pressure potentials of mass of type a at nodes u and v, respectively. The flow Q in Eq. (2.2) can be written in matrix notation as:

$$Q = D \cdot L^{-1} \cdot \sigma \cdot P \quad , \tag{2.3}$$

with D and L denoting the diagonal matrices of conductivity and length, respectively. Potential P is a $\mathbb{R}^{N \times M}$ matrix and $Q \in \mathbb{R}^{E \times M}$; $\sigma \in \mathbb{R}^{N \times E}$ are the entries of the signed network incidence matrix of undirected $\bar{\mathcal{G}}$:

$$\sigma_{ve} = \begin{cases} 1 & \text{if node } v \text{ is the start of } e \quad , \\ -1 & \text{if node } v \text{ is the end of } e \quad , \\ 0 & \text{otherwise} \quad . \end{cases} \tag{2.4}$$

The quantities L and matrix σ are given by the input multilayer graph $\bar{\mathcal{G}}$. The main goal is to determine the quantities D and P, that in turn determine the fluxes Q, as we show below.
Kirchhoff's law can then be defined as follows:

$$\sigma^T \cdot Q = S \quad , \tag{2.5}$$

where S is a source matrix such that $\sum_v S_v^a = 0$. By substituting Eq. 2.3 into Eq. 2.5, the relationship between P and S are established as follows:

$$\sigma^T \cdot \left(D \cdot L^{-1} \cdot \sigma \cdot P\right) = S \quad , \tag{2.6}$$

and to determine the P, the solution is obtained as

$$P = \left(\sigma^T \cdot D \cdot L^{-1} \cdot \sigma\right)^{\dagger} \cdot S \quad , \tag{2.7}$$

with † denoting the Moore-Penrose pseudoinverse. When the value of D is given and P is obtained using Eq. (2.7), one can now compute the flux Q with P and D as follows:

$$Q = D \cdot L^{-1} \cdot \sigma \cdot \left(\sigma^T \cdot D \cdot L^{-1} \cdot \sigma\right)^{\dagger} \cdot S \quad . \tag{2.8}$$

Eq. (2.8) shows the relationship between Q and D, indicating that one quantity can influence another: large conductivity leads to higher flux of an edge, and vice versa [73]. This mechanism has been widely used in various networks [4, 42, 63, 73]. An example of a transport network with two traffic demands is depicted in Fig. 2.2.

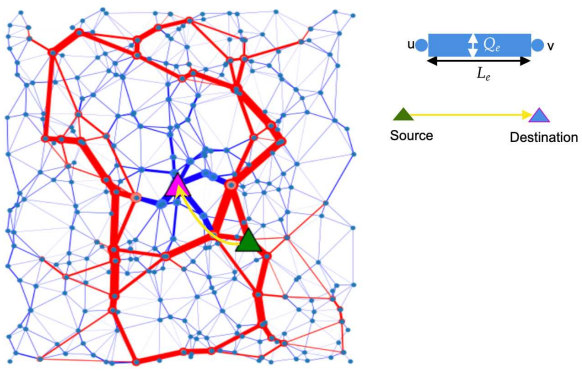


Figure 2.2: An example of a multilayer transport network. Layer 1 with 300 nodes is highlighted in blue and layer 2 with 60 nodes is highlighted in red. One passenger type is highlighted $S = \{(o_1, t_1)\}$. The trajectory of passenger type; (o_1, t_1) is indicated with arrow from the green triangle-shaped (source) node to the triangle-shaped magenta (destination) node. The trajectory of (o_1, t_1) uses edges from both layers.

2.3 Adaptative Dynamics

Assuming that D_e relies on Q_e, the adaptation equation for conductivity evolution D_e(t) is defined as:

$$\frac{d(D_e)}{dt} = h(|Q_e|) - D_e \quad , \tag{2.9}$$

with $|\cdot|$ denoting absolute value. The term $h(|Q_e|)$ is a positive feedback mechanism between flux and conductivity and describes a relationship where the behavior of D_e influences the flux Q_e. The original model of Eq. (2.9), described in [73], considered two functional forms $h(|Q_e|)$; $h(|Q_e|) = (|Q|)^x$ and $h(|Q_e|) = |Q_e|^x/(1 + \zeta|Q_e|^x)$, with $x > 0$ and ζ is a constant. Instead, we consider a regularized $h(|Q_e|)$ to capture different transport regimes. Specifically, we define $h(|Q_e|) = ||Q_e||_2^2$. We

then re-define Eq. (2.9) more concretely, the adaptive dynamics is as follows:

$$\frac{d(D_e)}{dt} = D_e^{\beta_{q_e}-2}||Q_e||_2^2 - D_e \equiv D_e^{\beta_{q_e}} \frac{\sum_a (p_u^a - p_v^a)^2}{L_e^2} - D_e \ , \quad e \in E_\alpha, a \in S \quad , \tag{2.10}$$

where $\beta_{q_e} \in (0,2)$ is a regularization parameter to study different transportation scenarios, with $q_e \equiv q_e(\alpha)$ takes value $q_e = \alpha$ for every $e \in E_\alpha$ and $\alpha \in F$. We have $\beta_{q_e} \in (0,2)$ which plays a critical role in governing the desired optimal transport dynamics within the system. When $\beta_{q_e} < 1$, it imposes a penalty on traffic congestion by favoring the distribution of paths across a greater number of edges. Conversely, for $\beta_{q_e} > 1$, the system encourages the consolidation of paths onto fewer highways. In this context, when $\beta_{q_e} = 1$, the transportation regime resembles that of the shortest path, emphasizing the minimization of travel distance. Constant β_{q_e} therefore serves as a tuning parameter, allowing for the adjustment of the transportation behavior in the network model, with different values influencing the trade-off between path distribution, consolidation, and adherence to the shortest path paradigm.

The dynamics described in Eq. (2.10) can be demonstrated to accommodate a Lyapunov function:

$$\mathcal{L}_{\beta_{q_e}} := \frac{1}{2} \sum_a \sum_v p_v^a(D) S_v^a + \frac{1}{2\kappa} \sum_e L_e D_e^\kappa \ , \quad a \in M, v \in N, e \in E \quad , \tag{2.11}$$

that combines the cost of operating the network (first term of Eq. (2.11)) and the cost of building the infrastructure (second term of Eq. (2.11)) [38, 49]. Here we define $\kappa = 2 - \beta_{q_e}$.

While Eq. (2.10) is designed for a multilayer network, one can equally adapt this to a single layer network by replacing β_{q_e} with a constant β, see [49, 50]. In general, the adaptive dynamics Eq. (2.9) has been used in various areas [35, 39, 62, 63]. Specifically, [39] modified Eq. (2.9) by adding a control term to a fraction of edges in the network, to improve the solution of the dynamics. Ref. [31] added a constrained additive noise to the dynamics and showed that stochastic dynamics can boost transport on a nonlinear network. Also, [38] generalizes Eq. (2.10) by imposing arbitrary constraints on the dynamics.

2.4 Transport cost

Minimizing transport costs within our model is crucial as it increases efficiency, promotes competitive advantage, reduces environmental impact and ultimately contributes to a sustainable and realistic transport system. One efficient way is

to adopt the regularized Monge–Kantorovich optimal transport [41], which have proved to be efficient [3, 10, 27, 30, 40, 43, 44, 65].

Generalizing [15, 35, 39, 43, 50, 63], the total transport cost is defined as:

$$\min_{D\in\mathbb{R}^{E\times M},Q\in\mathbb{R}^{E\times M}} \left\{ \frac{1}{2}\sum_{e\in E} \frac{w_e L_e}{D_e} ||Q_e||_2^2 \right\} \quad , \tag{2.12}$$

and the constraints:

$$\sum_{e\in E} L_e D_e^{\kappa} = \rho^{\kappa} \quad , \quad \text{(Material Cost)} \tag{2.13}$$

$$\sum_{e\in E} \sigma_{ve} Q_e^a = S_v^a \quad . \quad \text{(Local Kirchhoff)} \tag{2.14}$$

Here we have $\kappa = 2 - \beta_{q_e}$, $\rho > 0$ is a fixed constant and the parameter $w_e \in [0, 1]$, $w_e L_e$ is proportional to time needed to travel along edge e. The quantity Q_e is defined per traffic demand $a \in M$, and requires in the optimization setup to adopt 2-norm $||Q_e||_2$ [49]. One can equally define the flux Q for one traffic demand [3, 9]. The description of β_{q_e} is the same as in Section 2.3. The cost constraint Eq. (2.13) is the amount of material available to build the multilayer transport network, and Eq. (2.14) ensures mass conservation.

The constrained optimization setup in Eq. (2.12) to (2.14), is equivalent to the minimization problem:

$$J_{\Gamma(\beta_{q_e})} = \sum_{\alpha\in F}\sum_{e\in E} w_e L_e ||Q_e||_2^{\Gamma(\beta_{q_e})} \quad , \tag{2.15}$$

where $\Gamma(\beta_{q_e}) = 2\kappa/(1+\kappa)$, and $||\cdot||_2$ is computed over $M-$ entries of each Q_e. The overall transport cost in Eq. (2.15) is defined for a multilayer network. In analogous to a single-layer network, the settings: $w_e = 1$ and β_{q_e} as β in Eq. (2.15) gives the same cost function as in [49].

3 Challenges

This chapter describes in detail the challenges this thesis tackles, grouped under three themes; i) multilayer transport network, ii) paths distribution and iii) constrained adaption dynamics.

3.1 Multilayer transport network

A multilayer network is a type of complex network in which multiple graphs or layers are combined to form a single entity or framework. Multilayer networks have been widely used to study different complex systems including; communication networks [60, 81], ecological networks [57], community detection problems [14] transport networks [1, 18, 24, 70, 71], epidemic spreading [12, 20, 66] and climate science [23]. Multilayer networks, as studied in various works [2, 7, 8, 45], provide a valuable approach to studying complex transport systems involving multiple modes.

When analyzing a multilayer transport network, one of the challenges is the dynamic interactions between the various transport layers and the unpredictable behavior of passengers. In a real scenario, passengers switch between transport modes based on personal preferences, traffic conditions, or unforeseen events like accidents or road closures. The inherent complexity of passengers' decision-making and mode-switching behavior introduces an additional layer of complexity to the analysis of traffic congestion in multimodal systems.

Developing an algorithm to generate optimal routes and addressing the challenges of multilayer transport networks requires a robust approach. Some existing literature [18, 32, 52, 70, 80] on multilayer networks described methods to extract passengers trajectories. De Domenico *et al* [18] addresses the assessment of navigability in multilayer networks, focusing on the impact of random failures. Using random walks, the study provided insights into the design of effective search and navigability strategies in a multilayer transport network. Morris and Barthelemy [52] introduced a utility measure, named coupling[1], to capture essential features in a multilayer network. The study considers not only the topological aspects of a multilayer network but also the distribution of sources and sinks and the method of route allocation, using shortest-path minimization. Overall, the study shows that coupled

[1] Coupling in a multi-layer transport network emphasizes the interconnected nature of different transport modes and their mutual influence on each other.

multilayer systems exhibit behavior that is highly dependent on the interplay between coupling and randomness in the source-sink distribution. Solé-Ribalta *et al* [70] addresses the problem of congestion in multilayer networks, focusing on the complex connectivity of multilayer transport networks. The research analytically proves that the structure of multilayer networks can induce congestion in transport flows. Gao *et al* [32] introduced a flow assignment strategy tailored to multilayer networks and the contribution of [80], which develops a recurrent algorithm designed for communication networks.

Beyond these methods, efforts have been made to design efficient and principled approaches, relying on optimal transport theory principles, to extract optimal routes for multiple passengers on both single [4, 5, 10, 49] and multilayer [37] transport networks.

In Section 4.1, we propose *MultiOT*; an efficient optimal transport-based method for modeling optimal flows in multilayer networks. *MultiOT* addresses a broader scope of transport cost minimization and involves a regularized formulation derived from the Monge-Kantorovich optimal transport problems [41].

Section 3.2 and Section 3.3 will now focus on the challenges of optimal transport-based approaches.

3.2 Paths distribution

Methods relying on the principles of optimal transport (OT) have been used to model various aspects of transport networks, including the network design [5, 63, 73] and traffic flow [9, 10, 37, 49]. These researches have shown that OT-based methods guarantee a principled and computationally efficient way of solving traffic congestion problems on multilayer transport networks (also valid on single-layer).

In the standard OT methods, the flows can be generated either per passenger types (traffic demand[2]) or for all traffic demands at once without distinguishing the passenger types. Empirical results have shown that routes generated from the OT method tend to generate distributed trajectories for passengers of the same type. This approach achieves optimal distribution of passengers across the network ensuring reduced traffic congestion, and may not pose a problem in a single transport mode. However, in a multilayer scenario, distributing the trajectories of passengers of the same type to different layers may no longer be optimal. To address the challenges

[2] a traffic demand means one origin-destination pair or one passenger type. Passengers of the same type are passengers with the same origin and destination. We have multiple passengers traveling on the network.

of path distribution in multilayer settings, there's a need to develop an alternative algorithm that mimics the *Dijkstra* approach [21] but relies on OT principles.

Section 4.2 of this thesis propose *MultiOTsp* that interpolates between classical OT and shortest-path minimization approaches. Similar to *Dijkstra*, *MultiOTsp* generates a single path for every passenger of the same type. The optimal trajectories of *MultiOTsp* are significantly different from those obtained by *Dijkstra*, which are independent of the surrounding environment. Additionally, three relevant properties, using a multilayer transport network, were discussed to critically investigate the routes generated by *MultiOTsp* in comparison with *MultiOT* and *Dijkstra*. These properties include: the length of the generated paths, the amount of carbon generated following these paths, the traffic encountered on these paths, and the overall transport cost to navigate these routes.

3.3 Constrained dynamics

Over the last few decades, a wide range of transport systems have been successfully modeled using optimal transport, from biological networks such as leaf veins and blood vessels [63, 72] to engineering networks [50] such as urban transport or communication networks. In this context, adaptive dynamics are used to describe how conductivities, flow and pressure potentials evolve interdependently to form an optimal network structure (see Section 2.3), and have been used to study a various transport scenarios [15, 35, 43, 44, 63], and have been shown to explain with a high degree of similarity to observed real networks [73]. Additionally this approach has shown to be principled [26, 75] and computationally efficient [28, 56].

However, current approaches based on adaptative dynamics fall short on realistic constraints, beyond the standard constraints such as conservation of mass and positivity, as part of the general framework. As a result, the network generated by these models can be unrealistic[3] in practice without the inclusion of relevant constraints, such as constraints on the global network infrastructure or imposing edge capacity constraints. The main challenge preventing the natural incorporation of constraints into the current OT framework is that first-order methods for constrained optimization (e.g. projected gradient descent) are only effective when the feasible set has a very simple structure, such as a low-dimensional hyperplane or Euclidean norm ball, where projections can be evaluated in closed form.

[3] Tendency for traffic to concentrate on a few edges, potentially deviating from a structurally feasible distribution. For example, one can consider a monocentric destination for all passenger types. This concentration could lead to suboptimal utilization of the transport network.

Section 4.3 of this thesis addresses this flaw, by introducing a constrained dynamics that imposes arbitrary constraints on the adaptive dynamics. In addition, we show three examples where analytical solutions can be used to update the dynamics.

4 Published Work

This chapter summarizes the three first-author peer-reviewed articles and one collaborative peer-reviewed article, I (co-)authored during my studies. The complete articles are attached in the appendix section.

The articles in the appendix can be read in the order given below:

1.) A. A. Ibrahim, A. Lonardi, and C. D. Bacco. "Optimal transport in multilayer networks for traffic flow optimization". *Algorithms* 14.7 (2021). [37]
2.) A. A. Ibrahim, D. Leite, and C. De Bacco. "Sustainable optimal transport in multilayer networks". *Physical Review E* 105.6 (2022). [36]
3.) A. A. Ibrahim, M. Muehlebach, and C. De Bacco. "Optimal transport with constraints: from mirror descent to classical mechanics". *Physical Review Letters* 133 (2024). [38]
4.) 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 (2022). [48]

The first article introduced how to model and generate optimal routes in multilayer networks. Some of these routes can be distributed for passengers of the same type. The second paper proposed a single trajectory for each passenger type and also established three concrete properties of these methods. The third article shows how to impose arbitrary constraints on optimal transport problems in a principled and flexible way. While the first three articles focus on solving transport problems, the fourth article shows another application of optimal transport formulation, by solving community detection problems.

Authors Contribution statement

To determine and evaluate my contribution in each published paper, I described my role and summarized it with a descriptive terms, and in a tabular form. Specifically, I categorize my contribution as either significant, major, medium or minor. To clarify, **significant** denotes a contribution of paramount importance, **major** denote a crucial involvement without which the submission might not have been possible, **medium** denotes an important contribution that enhances the quality of the paper and potentially affects its acceptance, and **minor** highlights an improvement made

to the paper. This framework serves as a guide to express the varying degrees of impact each co-author had on the published papers.

4.1 Optimal transport in multilayer networks for traffic flow optimization

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 are distributed 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 using 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 with 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.

Research contribution: The main takeaway from this study is that optimal transport can be used to model passenger flows in a multilayer network. While optimal transport is efficient for single-layer networks, this study adapts these principles to multilayer scenarios. The model introduces a parameter that allows flexible tuning of the sensitivity to traffic congestion across layers.

Author contribution: I led this project as the first author. Together with my PhD advisor, we conceived this project's idea. I participated in the writing of the codes, and carried out extensive analysis of the results. I conducted all experimental work and most of the scientific plots. In addition, I schedule regular meeting with the co-authors. I contributed to the writing of the paper and responses to the reviewer's comments. Finally, I was responsible for submitting the paper to the journal and preparing the arXiv version. Finally, I have prepared and uploaded the code to GitHub, making it accessible to readers.

Remark 4.1 My contributions are summarized as follows:

	Ideas	Implementation	Experiment	Analysis	Writing
A.A.I.	significant	significant	significant	significant	significant

Venue: *Algorithms* is a peer-reviewed journal published by MDPI since 2008. Publications in this venue are open access and have high visibility. The journal focuses on research related to algorithms and their applications. This interdisciplinary journal welcomes new and innovative research ideas as well as special issues on specific topics or areas.

4.2 Sustainable optimal transport in multilayer networks

Abstract: Traffic congestion is one of the major challenges faced by the transportation industry. While this problem carries a high economic and environmental cost, the need for an efficient design of optimal paths for passengers in multilayer network infrastructures is imperative. We consider an approach based on optimal transport theory to route passengers preferably along layers that are more carbon-efficient than the road, e.g., rails. By analyzing the impact of this choice on performance, we find that this approach reduces carbon emissions considerably compared to shortest-path minimization. Similarly, we find that this approach distributes traffic more homogeneously, thus alleviating the risk of traffic congestion. Our results shed light on the impact of distributing traffic flexibly across layers guided by optimal transport theory.

Research contribution: The main contributions of this study include an optimal transport-based approach for extracting a single trajectory for each passenger in multilayer transport networks; an interpolation method between optimal transport and the shortest path minimization, and the demonstration that these optimal routes lead to reduced traffic and CO2 emissions, resulting in reduced environmental costs.

Author contribution: I led this project as the first author. Together with my PhD advisor, we conceived this project's idea. I participated in the writing of the codes, and carried out analysis of the experimental results. I conducted all experimental work and most of the scientific plots. In addition, I schedule regular meeting with the co-authors. I contributed to the writing of the paper and responses to the reviewer's comments. Finally, I was responsible for submitting the paper to the journal and preparing the arXiv version. Finally, I have prepared and uploaded the code to GitHub, making it accessible to readers.

Remark 4.2 My contributions are summarized as follows:

	Ideas	**Implementation**	**Experiment**	**Analysis**	**Writing**
A.A.I.	significant	significant	significant	significant	significant

Venue: *Physical Review E* (PRE) is a peer-reviewed and open access journal published by the American Physical Society. Established in 1993, this physics journal accepts research that make a significant contribution to a specific research area. It is one of the leading journal in networks and (physics-based) algorithms. PRE is known for high-quality publications both in traditional and emerging research areas.

4.3 Optimal transport with constraints: from mirror descent to classical mechanics

Abstract: Finding optimal trajectories for multiple traffic demands in a congested network is a challenging task. Optimal transport theory is a principled approach that has been used successfully to study various transportation problems. Its usage is limited by the lack of principled and flexible ways to incorporate realistic constraints. We propose a principled physics-based approach to impose constraints flexibly in such optimal transport problems. Constraints are included in mirror descent dynamics using the principle of D'Alembert-Lagrange from classical mechanics. This leads to a sparse, local and linear approximation of the feasible set leading in many cases to closed-form updates.

Research contribution: Existing approaches based on adaptive dynamics are limited by the fact that they do not incorporate constraints beyond standard constraints such as conservation of mass and positivity constraints, as part of the general framework. To address this flaw, we proposed a framework powerful enough to handle nonlinear and non-convex constraints in optimal transport problems.

To the best of my knowledge, this is the first theoretical formulation that imposes constraints on adaptation equations and follows a physics-based perspective on including constraints by leveraging the principle of d'Alembert-Lagrange from classical mechanics.

Author contribution: I led this project as the first author and was the only student author involved in this project. Together with the co-authors, we conceived this project's idea. I participated in the writing of the codes, and carried out analysis of the experimental results. I conducted all experimental work and most of the scientific plots. In addition, I participated in regular meeting with the co-authors to discuss the progress of the project. I contributed to the writing of the paper and

responses to the reviewer's comments. Finally, I was responsible for submitting the paper to the journal and preparing the arXiv version. Finally, I have prepared and uploaded the code to GitHub, making it accessible to readers.

Remark 4.3 My contributions are summarized as follows:

	Ideas	**Implementation**	**Experiment**	**Analysis**	**Writing**
A.A.I.	significant	significant	significant	significant	significant

Venue: *Physical Review Letters* (PRL) is a peer-reviewed, open-access journal published by the American Physical Society. Founded in 1958, PRL is recognized worldwide as the premier journal for physics letters. With rigorous acceptance criteria, it selectively publishes papers that significantly advance research areas. PRL publications are characterized by their conciseness, high quality and wide visibility.

4.4 Community detection in networks by dynamical optimal transport formulation

Abstract: Detecting communities in networks is important in various domains of applications. While a variety of methods exist to perform this task, recent efforts propose Optimal Transport (OT) principles combined with the geometric notion of Ollivier–Ricci curvature to classify nodes into groups by comparing the information encoded into nodes' neighborhoods. We present an OT-based approach that exploits recent advances in OT theory to allow tuning between different transportation regimes. This allows for better control of the information shared between nodes' neighborhoods. As a result, our model can flexibly capture different types of network structures and thus increase performance accuracy in recovering communities, compared to standard OT-based formulations. We test the performance of our algorithm on both synthetic and real networks, achieving a comparable or better performance than other OT-based methods in the former case, while finding communities that better represent node metadata in real data. This pushes further our understanding of geometric approaches in their ability to capture patterns in complex networks.

Research contribution: This study presents an optimal transport-based approach for detecting communities in networks. By incorporating the Ollivier-Ricci curvature, the model offers various transport regimes that provide better control over information flow between node neighborhoods. The algorithm demonstrates improved accuracy in recovering communities, outperforms standard OT-based methods in

synthetic networks, and better represents node metadata in real data, advancing our understanding of geometric approaches to analyzing complex networks.

Author contribution: I contributed to the idea formulation, participated in the coding and implementation, obtained the metadata of real networks, and took part in the results discussion. In addition, I participated in regular project meetings. I contributed to the writing of the paper and responses to the reviewer's comments.

Remark 4.4 My contributions are summarized as follows:

	Ideas	**Implementation**	**Experiment**	**Analysis**	**Writing**
A.A.I.	medium	significant	medium	major	major

Venue: *Scientific Reports*, a prestigious publication within the Nature portfolio, has been a pioneering force in the dissemination of cutting-edge research since its launch in 2011. With a commitment to open access and strict peer review, the journal provides a platform for scientists worldwide to present their original research across a range of disciplines including the natural sciences. As a testament to its impact, Scientific Reports is ranked the 5th most cited journal in the world.

5 Discussion and Conclusion

This chapter discusses the key challenges and how the published papers addresses them. For specific and detailed results, readers can visit the published articles. I summarized the published papers according to common themes.

Throughout this chapter, I denote with α and γ the bus and tram layers, respectively.

5.1 Optimal transport in multilayer networks

One effective strategy for alleviating congestion involves leveraging a multilayer transport network [13, 74], which provides users with the flexibility to switch between different modes of transportation during their travels. Our approach focuses on extracting optimal trajectories for various types of passengers. While this approach reduces congestion and environmental costs, the concept of passengers selfish routing cannot be ignored [59]. As demonstrated in *MultiOT* [37], we adopted optimal transport principles to extract optimal trajectory for each passenger type in a multilayer network. We consider various traffic regimes using the parameter $0 < \beta_{q_e} < 2$, and how different settings of β_{q_e} influence relevant outputs such as the optimal routes for each passenger and network topology. In a multilayer network with two layers α and γ, we use $\beta_{q_e} \in \{\beta_\alpha, \beta_\gamma\}$. This means one needs to set layer-wise values to determine how the passenger should be routed in the transport network. We study the behaviour of *MultiOT* on the bus and tram network of the city of Bordeaux. Using the settings $\beta_\alpha < 0.5$ and $\beta_\gamma > 1.5$, we observe that about 17% of the total passenger flows are routed on the γ (tram) layer in order to reduce congestion on the α (bus) layer. These results also show that each traffic regime has a different network topology. In particular, when paths are consolidated along a few edges, the topology is significantly different from when paths are more distributed along many edges.

The optimal trajectories generated from *MultiOT* are more distributed along many edges for some passenger types. We then introduced *MultiOTsp* [36], a variant of *MultiOT* that generates a single trajectory from source to destination for each passenger type. We observe that the average path length traveled by passengers following the trajectories generated by the OT-based methods (i.e., *MultiOT* and *MultiOTsp*) is increased. This is because the OT-based methods are tuned to encourage higher usage of the tram layer to decongest roads. In fact, we obtain that OT-based

methods record up to 53% more rail usage than that obtained via shortest-path optimization.
In addition, we observe how the parameter β_{q_e} also influences the overall transport cost and thus the resulting trajectories. The settings $\beta_\alpha, \beta_\gamma < 1$ discourage traffic congestion on both layers. In comparison to the *Dijkstra* algorithm, setting $\beta_\alpha = \beta_\gamma = 0.5$ in the optimal transport model leads to a higher transport cost, up to 50% more than the cost obtained from *Dijkstra*. However, the same experimental results (see [36]) with $\beta_\alpha = 0.5$ and $\beta_\gamma = 1.9$, i.e. when we encourage path consolidation in the tram network, shows that the cost of navigating routes generated by the OT-based model is 30% lower than that of *Dijkstra*.

Overall, combining optimal transport and multilayer transport networks can be used to alleviate traffic congestion. Depending on the desired goals of a network manager, our model allows to arbitrarily tune the main model parameters to simulate the desired transport regime and quantify how traffic congestion can be decreased on roads and rail usage encouraged.

5.2 Sustainable optimal transport

To assess the sustainability of transport networks, the provision of alternative transport modes is a relevant approach. As demonstrated in [36, 37], a multilayer network can be used to alleviate traffic from a congested network. To measure the environmental impacts, we assess the amount of carbon emission per passenger per unit of length (pkm) on the routes generated by *MultiOT*, *MultiOTsp* and *Dijkstra*. Considering a two-layer network consisting of bus and tram layers, where a bus emits on average $101.87 g/pkm$ [19] and a train emits $28.39 g/pkm$ [25], *MultiOTsp* recorded the lowest carbon emissions, with 25% less emissions than a shortest-path approach. *MultiOTsp* achieved its lowest carbon emission value when path consolidation was not overly enforced, specifically at $\beta_\gamma = 1.3$. This highlights the trade-off between minimizing the average path length of passengers and routing more passengers on the tram layer.

Accounting for traffic on roads is another important sustainability metric, as it directly reflects the efficiency and functionality of transport networks, affecting factors such as fuel consumption and passenger satisfaction. We demonstrate in [36–38] by measuring the Gini coefficient[1], how traffic loads are distributed for OT and shortest-path algorithms. $\text{Gini}(T_e) \in [0, 1]$, where T_e is the edge-wise traffic loads. A Gini close to 1 indicates a high degree of inequality in flow allocation

[1] A coefficient that measures the inequality of a certain quantity, in this case we consider traffic distribution on a transport network.

along the edges, and a Gini close to 0 indicates a balanced flow along the edges. For the various transport regimes considered, the exact value of Gini relies on the distribution of traffic demands. From the experimental result, we observed that *MultiOT* shows a more evenly distributed traffic pattern compared to the other algorithms. Even when congestion levels increase with increasing values of β_γ, our results show that *MultiOT* maintains lower Gini coefficients than *Dijkstra* for all values of β_γ considered. This phenomenon can be attributed to the tendency of paths to concentrate on fewer edges on the bus layer that facilitate connections to the tram layer. As a result, the Gini coefficient of *MultiOTsp* exceeds that of *Dijkstra* as the central edges of the bus layer become congested, especially as passengers move from the tram network to city center destinations. Furthermore, we observed differences in passenger distribution between the two OT-based variants in terms of their impact on tram network utilization. While both variants show a tendency to distribute passengers more along the tram network, thereby reducing road traffic, the *Dijkstra* uses the tram predominantly in the vicinity of central nodes. This means the OT methods recorded higher tram usage compared to *Dijkstra*. Also, *MultiOT* shows a higher intensity of road use compared to MultiOTsp, but with less congestion on the road edges. Conversely, MultiOTsp shows a distinct pattern where road edges experience increased traffic near tram stations. Comparing the OT models, the *MultiOT* balances traffic on edges better whereas the *MultiOTsp* achieved lower value of carbon emission.

From the results of our OT models and their comparison with the traditional shortest path algorithm, OT offers a promising way to design dynamic and sustainable transport networks within multilayer infrastructures. Our analysis underlines the crucial role of path configurations - whether single or multiple paths - and the parameter β_{q_e} in shaping the optimal sustainability metric. These findings pave the way for tailored strategies aimed at maximizing efficiency while minimizing environmental impact in transport planning.

5.3 Constrained optimal transport

The number of passengers passing through an edge in these OT methods described in [36, 37] is unconstrained, as passenger flows tend to concentrate on a few edges, especially when passenger paths are consolidated (i.e. $\beta_{q_e} > 1$) and feature a monocentric destination. This characteristic tends to limit the usage of OT methods in real transport networks. We addressed this crucial flaw in [38] and showed that, in addition to the standard[2] constraints imposed by classical OT methods, one can

[2] Classical optimal transport has standard constraints such as positivity constraint and conservation of mass.

also impose other realistic constraints on optimal transport problems in a principled and flexible way. In particular, we show how arbitrary constraints can be introduced into optimal transport problems in a computationally efficient way. Our constrained approach is simple and intuitive as it involves adding an auxiliary force to the dynamics. This allows to seamlessly incorporate constraints of various types, e.g. nonlinear or non-convex, into the mirror descent dynamic. It yields a standard quadratic program which can often be solved in closed form. The formalism can be applied to different constraint functions, provided one can derive the gradient of the constraint function. In our analysis, we considered different constraint functions in [38] to mimic real scenarios.

The first constraint involves structural limit imposing edge capacity constraints on the number of passengers traveling along any road (or edge) in the transport network. In [38], we considered various sizes of edge capacities; small, medium and large edge capacities. The experimental results indicated that imposing edge capacity constraints on the transport network balances traffic more than the classical OT method, such as the *MultiOT*. Setting the right edge capacity c plays a role in the overall network structure. From the various sizes of edge capacities considered, we observed that a lower edge capacity imposes a more restrictive constraint on the flow of traffic, leading to a more balanced distribution along the edges of the network. This equitable distribution is reflected in a lower Gini coefficient. We considered homogeneous edge capacities for a single layer network in [38], but one can equally consider heterogeneous capacities for a multilayer network, see Fig. 5.1 for example. For a multilayer network, we define the edge capacity as $c_{q_e} \in \{c_e^\alpha, c_e^\gamma\}$ where c_e^α, c_e^γ are the edge capacities for the layer α, layer γ, and $c_e^\gamma > c_e^\alpha$ always, as depicted in Fig. 5.1. Each layer has different edge capacities to mimic a real scenario where a tram has more capacity than a bus. While a lower edge capacity value balances traffic more, a high value of c_{q_e} reduces the path length of passengers (see Fig. 5.1(C-D)). In particular, lower c_{q_e} values result in longer average total path lengths for passengers traveling on the network. The relationship between edge capacity c_{q_e} and parameter β_{q_e} is integral to understanding and optimizing the performance of the network.

Next, we introduce a budget constraint that includes all the edges simultaneously, reflecting a scenario where a network manager is faced with a fixed, limited investment resources. Specifically, we reduced the initial network budget by 50% (i.e., half of *MultiOT*) and observe that the resulting routes with enforced path consolidation exhibit a more balance traffic without a significant increase in the total path length of passengers, compared to those generated by *MultiOT* [38]. In the budget constraint scenario, we do not impose capacities on individual edges, but consider a global constraint of an overall network budget.

Overall, imposing additional constraints on classical OT models improves the performance and gives a more realistic results. While we highlight two scenarios

of real-world constraints that can be applied to adaptive dynamics, non-linear constraint can be used as well (see [38] for another scenario).

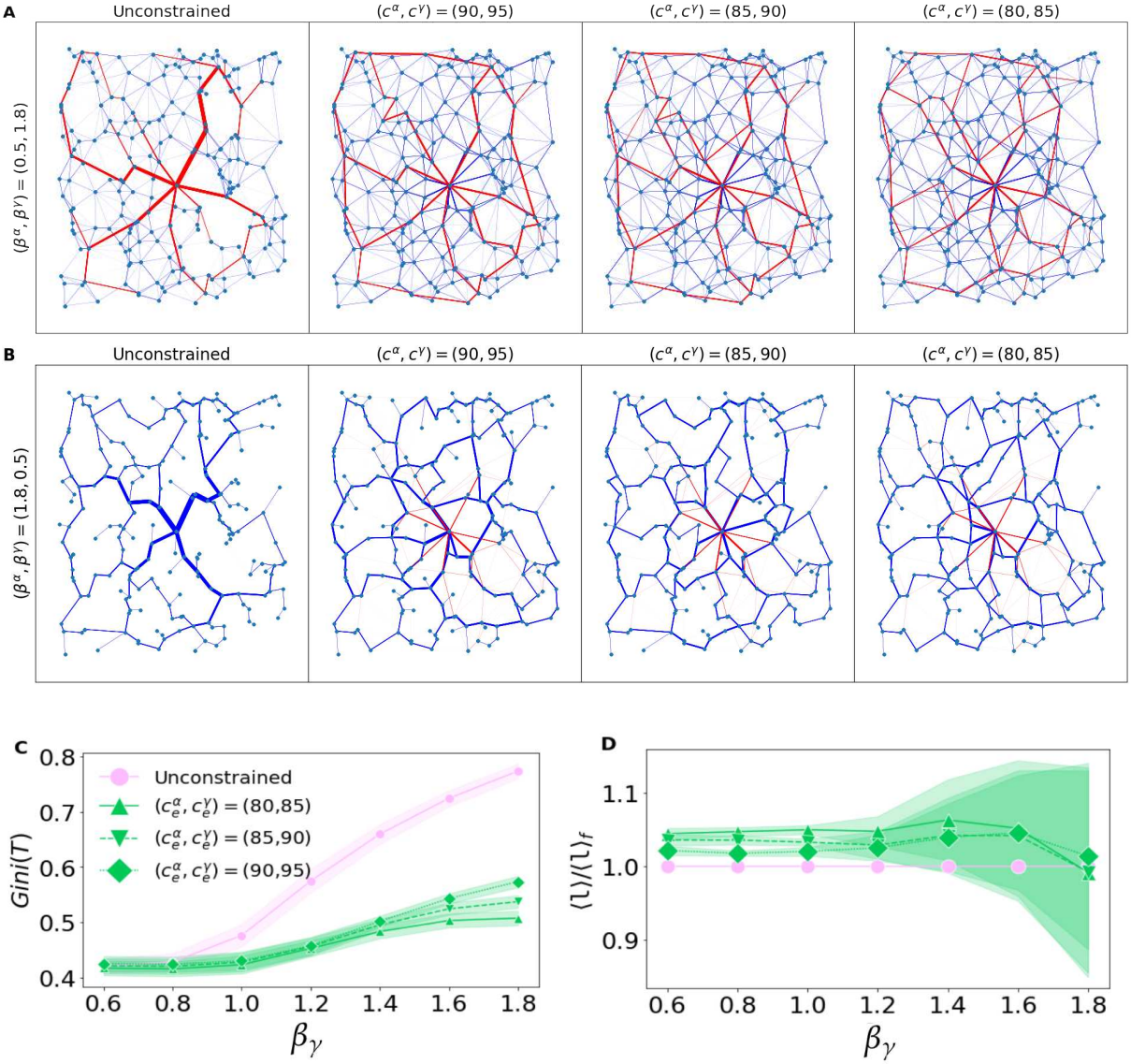


Figure 5.1: Example of a synthetic multilayer network. The edge capacity for each layer is selected as the percentiles of the distribution of μ over edges obtained in the unconstrained case (Unconstrained). Panel (A) favors the usage of the tram layer (i.e., $\beta_\gamma > 1$ and (B) favors the usage of the bus layer (i.e., $\beta_\alpha > 1$). The edges belonging to the bus and tram layers are indicated in blue and red colors, respectively. Panels (c) and (D) show the Gini coefficient of the traffic distribution on edges and the ratio of averaged path length to that of *Unconstrained*, respectively. Markers and shadows are averages and standard deviations over 10 network realizations, with 50 randomly selected origins and destinations. Edge widths are proportional to the amount of passengers traveling through an edge. Other settings: $N_\alpha = 200$, $N_\gamma = 50$ Nodes.

5.4 Application on community detection problem

While the OT models [36–38] and described in Sections 5.1 to 5.3 are developed for transportation networks, we developed OT model named *ORC-Nextrout* [48] that focus on the detection of communities in large networks. Community detection in networks is another area of network science where optimal transport models can be applied. In the context of community detection, *ORC-Nextrout* uses geometric algorithms similar to the Ollivier-Ricci (ORC) [55] to compute curvature. Specifically, edges with positive curvature are considered to belong to the same community, and vice versa for negative curvature. The integration of optimal transport (OT) and ORC in community detection was initially motivated by works such as [53, 67]. However, *ORC-Nextrout* advances this approach by generalizing it to consider two regimes: branched [29, 65] and congested [11] in optimal transport problems. This technique allows the regulation of information sharing between nodes within the network.

The results obtained on four (4) real data (namely; Les Miserables, Dolphins, American Football and Political books) show that the performance of *ORC-Nextrout* depends on the structure of the ground truth and the transport regime. In some cases (i.e., Les Miserables and Dolphins), the congested regime of *ORC-Nextrout* achieved the highest accuracy, effectively recovering the communities, while in other cases (i.e., American Football) the branched regime of *ORC-Nextrout* showed superior performance. Overall, *ORC-Nextrout* outperformed other OT-based methods as it provides the flexibility to tune between different regimes. In addition, we observed that OT models, including *ORC-Nextrout*, tend to extract a larger number of communities compared to those identified from node metadata, with many of the additional communities consisting of a single node. This phenomenon suggests that OT methods tends to identify densely connected community structures. In summary, by adjusting between different regimes, *ORC-Nextrout* demonstrates superior or comparable performance in recovering community structures consistent with node metadata.

5.5 Conclusion

The methods presented in this thesis advances the fields of optimal transport theory and network science. Firstly, we developed computationally efficient methods based on optimal transport to extract passenger routes across multilayer transport networks. This methods presented enhance our understanding of traffic dynamics within multilayer transport networks, provides valuable insights that contribute to sustainability efforts in the transport sector. In addition, we extended classical optimal

transport methods to accommodate additional and realistic constraints in optimal transport problems. These extensions are particularly relevant to reducing congestion and minimizing the environmental impact of transport systems. Furthermore, our research demonstrates the versatility of optimal transport formulations in solving community detection problems within networks. By exploiting optimal transport principles, we offer a dynamic approach to community detection that provides valuable insights into network structures and functionalities.

Bibliography

[1] A. Aleta, S. Meloni, and Y. Moreno. "A multilayer perspective for the analysis of urban transportation systems". *Scientific reports* 7.1 (2017), pages 1–9.

[2] A. Aleta and Y. Moreno. "Multilayer networks in a nutshell". *Annual Review of Condensed Matter Physics* 10 (2019), pages 45–62.

[3] J. R. Banavar, A. Maritan, and A. Rinaldo. "Size and form in efficient transportation networks". *Nature* 399.6732 (1999), pages 130–132.

[4] D. Baptista and C. De Bacco. "Principled network extraction from images". *arXiv preprint arXiv:2012.12758* (2020). arXiv: 2012.12758 [cs.CV].

[5] D. Baptista, D. Leite, E. Facca, M. Putti, and C. De Bacco. "Network extraction by routing optimization". *Scientific Reports* 10 (1 2020), page 088702.

[6] M. Barthélemy. "Spatial networks". *Physics Reports* 499.1-3 (2011), pages 1–101.

[7] G. Bianconi. "Multilayer networks: structure and function". *Oxford university press*, 2018.

[8] S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin. "The structure and dynamics of multilayer networks". *Physics reports* 544.1 (2014), pages 1–122.

[9] S. Bohn and M. O. Magnasco. "Structure, Scaling, and Phase Transition in the Optimal Transport Network". *Phys. Rev. Lett.* 98 (8 2007), page 088702.

[10] V. Bonifaci, K. Mehlhorn, and G. Varma. "Physarum can compute shortest paths". *Journal of Theoretical Biology* 309 (2012), pages 121 –133.

[11] L. Brasco, G. Carlier, and F. Santambrogio. "Congested traffic dynamics, weak flows and very degenerate elliptic equations". *Journal de mathématiques pures et appliquées* 93.6 (2010), pages 652–671.

[12] J. Chen, M.-B. Hu, and M. Li. "Traffic-driven epidemic spreading in multiplex networks". *Physical Review E* 101.1 (2020), page 012301.

[13] J. Chen, C.-Y. Wu, M. Li, and M.-B. Hu. "Hybrid traffic dynamics on coupled networks". *Physica A: Statistical Mechanics and its Applications* 516 (2019), pages 98–104.

[14] M. Contisciani, A Power Eleanor, and C. De Bacco. "Community detection with node attributes in multilayer networks". *Scientific Reports* 10 (2020), page 15736.

[15] F. Corson. "Fluctuations and redundancy in optimal transport networks". *Physical Review Letters* 104.4 (2010), page 048703.

[16] 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* 2014.7 (2014), P07009.

[17] M. De Domenico, C. Granell, M. A. Porter, and A. Arenas. "The physics of spreading processes in multilayer networks". *Nature Physics* 12.10 (2016), pages 901–906.

[18] M. De Domenico, A. Solé-Ribalta, S. Gómez, and A. Arenas. "Navigability of interconnected networks under random failures". *Proceedings of the National Academy of Sciences* 111.23 (2014), pages 8351–8356.

[19] DEFRA. *Department for Environment, Food and Rural Affairs. Greenhouse Gas Reporting: Conversion Factors 2017*. 2017. URL: https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2017.

[20] M. Dickison, S. Havlin, and H. E. Stanley. "Epidemics on interconnected networks". *Physical Review E* 85.6 (2012), page 066109.

[21] E. W. Dijkstra et al. "A note on two problems in connexion with graphs". *Numerische mathematik* 1.1 (1959), pages 269–271.

[22] R Dobrin and P. Duxbury. "Minimum spanning trees on random networks". *Physical Review Letters* 86.22 (2001), page 5076.

[23] J. F. Donges, H. C. Schultz, N. Marwan, Y. Zou, and J. Kurths. "Investigating the topology of interacting networks". *The European Physical Journal B* 84.4 (2011), pages 635–651.

[24] W.-B. Du, X.-L. Zhou, O. Lordan, Z. Wang, C. Zhao, and Y.-B. Zhu. "Analysis of the Chinese Airline Network as multi-layer networks". *Transportation Research Part E: Logistics and Transportation Review* 89 (2016), pages 108–116.

[25] EEA. *Energy efficiency and specific CO2 emissions*. 2017. URL: https://www.eea.europa.eu/data-and-maps/indicators/energy-efficiency-and-specific-co2-emissions/energy-efficiency-and-specific-co2-9.

[26] E. Facca, F. Cardin, and M. Putti. "Physarum Dynamics and Optimal Transport for Basis Pursuit". arXiv: 1812.11782.

[27] E. Facca, F. Cardin, and M. Putti. "Towards a Stationary Monge-Kantorovich Dynamics: The Physarum Polycephalum Experience". *SIAM Journal on Applied Mathematics* 78.2 (2016), pages 651–676.

[28] E. Facca, F. Cardin, and M. Putti. "Branching structures emerging from a continuous optimal transport model". *Journal of Computational Physics* 447 (2021), page 110700.

[29] E. Facca, F. Cardin, and M. Putti. "Branching structures emerging from a continuous optimal transport model". *Journal of Computational Physics* 447 (2021), page 110700.

[30] E. Facca, S. Daneri, F. Cardin, and M. Putti. "Numerical Solution of Monge-Kantorovich Equations via a Dynamic Formulation". *Journal of Scientific Computing* 82.3 (2020), page 68.

[31] F. Folz, K. Mehlhorn, and G. Morigi. "Noise-induced network topologies". *Physical Review Letters* 130.26 (2023), page 267401.

[32] L. Gao, P. Shu, M. Tang, W. Wang, and H. Gao. "Effective traffic-flow assignment strategy on multilayer networks". *Physical Review E* 100.1 (2019), page 012310.

[33] S. Gomez, A. Diaz-Guilera, J. Gomez-Gardenes, C. J. Perez-Vicente, Y. Moreno, and A. Arenas. "Diffusion dynamics on multiplex networks". *Physical review letters* 110.2 (2013), page 028701.

[34] J. Gómez-Gardenes and V. Latora. "Entropy rate of diffusion processes on complex networks". *Physical Review E* 78.6 (2008), page 065102.

[35] D. Hu and D. Cai. "Adaptation and optimization of biological transport networks". *Physical review letters* 111.13 (2013), page 138701.

[36] A. A. Ibrahim, D. Leite, and C. De Bacco. "Sustainable optimal transport in multilayer networks". *Physical Review E* 105.6 (2022), page 064302.

[37] A. A. Ibrahim, A. Lonardi, and C. D. Bacco. "Optimal transport in multilayer networks for traffic flow optimization". *Algorithms* 14.7 (2021), page 189.

[38] A. A. Ibrahim, M. Muehlebach, and C. De Bacco. "Optimal transport with constraints: from mirror descent to classical mechanics". *Physical Review Letters* 133 (2024), page 057401.

[39] J. Jiang, X. Wang, and Y.-C. Lai. "Optimizing biologically inspired transport networks by control". *Physical Review E* 100.3 (2019), page 032309.

[40] F. Kaiser, H. Ronellenfitsch, and D. Witthaut. "Discontinuous transition to loop formation in optimal supply networks". *Nature communications* 11.1 (2020), pages 1–11.

[41] L. V. Kantorovich. "On the translocation of masses". *Journal of mathematical sciences* 133.4 (2006), pages 1381–1382.

[42] E. Katifori. "The transport network of a leaf". *Comptes Rendus Physique* 19.4 (2018), pages 244–252.

[43] E. Katifori, G. J. Szöllősi, and M. O. Magnasco. "Damage and Fluctuations Induce Loops in Optimal Transport Networks". *Phys. Rev. Lett.* 104 (4 2010), page 048704.

[44] J. B. Kirkegaard and K. Sneppen. "Optimal Transport Flows for Distributed Production Networks". *Phys. Rev. Lett.* 124 (20 2020), page 208101.

[45] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter. "Multilayer networks". *Journal of complex networks* 2.3 (2014), pages 203–271.

[46] L. Lacasa, M. Cea, and M. Zanin. "Jamming transition in air transportation networks". *Physica A: Statistical Mechanics and its Applications* 388.18 (2009), pages 3948–3954.

[47] A. Lampo, J. Borge-Holthoefer, S. Gómez, and A. Solé-Ribalta. "Multiple abrupt phase transitions in urban transport congestion". *Physical Review Research* 3.1 (2021), page 013267.

[48] 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 (2022), page 16811.

[49] A. Lonardi, E. Facca, M. Putti, and C. De Bacco. "Designing optimal networks for multicommodity transport problem". *Physical Review Research* 3.4 (2021), page 043010.

[50] A. Lonardi, M. Putti, and C. De Bacco. "Multicommodity routing optimization for engineering networks". *Scientific reports* 12.1 (2022), pages 1–11.

[51] M. J. Mogridge. "The self-defeating nature of urban road capacity policy: A review of theories, disputes and available evidence". *Transport policy* 4.1 (1997), pages 5–23.

[52] R. G. Morris and M. Barthelemy. "Transport on coupled spatial networks". *Physical review letters* 109.12 (2012), page 128703.

[53] C.-C. Ni, Y.-Y. Lin, F. Luo, and J. Gao. "Community detection on networks with Ricci flow". *Scientific reports* 9.1 (2019), page 9984.

[54] J. D. Noh and H. Rieger. "Stability of shortest paths in complex networks with random edge weights". *Physical Review E* 66.6 (2002), page 066127.

[55] Y. Ollivier. "Ricci curvature of Markov chains on metric spaces". *Journal of Functional Analysis* 256.3 (2009), pages 810–864.

[56] G. Peyré, M. Cuturi, et al. "Computational optimal transport: With applications to data science". *Foundations and Trends® in Machine Learning* 11.5-6 (2019), pages 355–607.

[57] S. Pilosof, M. A. Porter, M. Pascual, and S. Kéfi. "The multilayer nature of ecological networks". *Nature Ecology & Evolution* 1.4 (2017), page 0101.

[58] B. Pishue. "INRIX Global Traffic Scorecard" (2022).

[59] H. F. Po, C. H. Yeung, and D. Saad. "Futility of being selfish in optimized traffic". *Physical Review E* 103.2 (2021), page 022306.

[60] C. Pu, S. Li, X. Yang, J. Yang, and K. Wang. "Information transport in multiplex networks". *Physica A: Statistical Mechanics and its Applications* 447 (2016), pages 261–269.

[61] D. Quercia, R. Schifanella, and L. M. Aiello. "The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city". *Proceedings of the 25th ACM conference on Hypertext and social media*. 2014, pages 116–125.

[62] A.-G. Rolland-Lagan and P. Prusinkiewicz. "Reviewing models of auxin canalization in the context of leaf vein pattern formation in Arabidopsis". *The Plant Journal* 44.5 (2005), pages 854–865.

[63] H. Ronellenfitsch and E. Katifori. "Global optimization, local adaptation, and the role of growth in distribution networks". *Physical review letters* 117.13 (2016), page 138301.

[64] M. Rosvall, A. Grönlund, P. Minnhagen, and K. Sneppen. "Searchability of networks". *Physical Review E* 72.4 (2005), page 046117.

[65] F. Santambrogio. "Optimal channel networks, landscape function and branched transport". *Interfaces and Free Boundaries* 9 (2007), pages 149–169.

[66] A. Saumell-Mendiola, M. Á. Serrano, and M. Boguná. "Epidemic spreading on interconnected networks". *Physical Review E* 86.2 (2012), page 026106.

[67] J. Sia, E. Jonckheere, and P. Bogdan. "Ollivier-ricci curvature-based method to community detection in complex networks". *Scientific reports* 9.1 (2019), page 9800.

[68] K. Sneppen, A. Trusina, and M. Rosvall. "Hide-and-seek on complex networks". *Europhysics Letters* 69.5 (2005), page 853.

[69] A. Solé-Ribalta, A. Arenas, and S. Gómez. "Effect of shortest path multiplicity on congestion of multiplex networks". *New Journal of Physics* 21.3 (2019), page 035003.

[70] A. Solé-Ribalta, S. Gómez, and A. Arenas. "Congestion induced by the structure of multiplex networks". *Physical review letters* 116.10 (2016), page 108701.

[71] E. Strano, S. Shai, S. Dobson, and M. Barthelemy. "Multiplex networks in metropolitan areas: generic features and local effects". *Journal of The Royal Society Interface* 12.111 (2015), page 20150651.

[72] A. Takamatsu, T. Gomi, T. Endo, T. Hirai, and T. Sasaki. "Energy-saving with low dimensional network in Physarum plasmodium". *Journal of Physics D: Applied Physics* 50.15 (2017), page 154003.

[73] 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* 327.5964 (2010), pages 439–442.

[74] J. Wu, C. Pu, L. Li, and G. Cao. "Traffic dynamics on multilayer networks". *Digital Communications and Networks* 6.1 (2020), pages 58–63.

[75] Q. Xia. "Optimal paths related to transport problems". *Communications in Contemporary Mathematics* 5.02 (2003), pages 251–279.

[76] C. H. Yeung. "Coordinating dynamical routes with statistical physics on space-time networks". *Physical Review E* 99.4 (2019), page 042123.

[77] C. H. Yeung and D. Saad. "Competition for shortest paths on sparse graphs". *Physical Review Letters* 108.20 (2012), page 208701.

[78] L. Zhao, Y.-C. Lai, K. Park, and N. Ye. "Onset of traffic congestion in complex networks". *Physical Review E* 71.2 (2005), page 026125.

[79] Z. Zheng, Z. Huang, F. Zhang, and P. Wang. "Understanding coupling dynamics of public transportation networks". *EPJ Data Science* 7 (2018), pages 1–16.

[80] J. Zhou, G. Yan, and C.-H. Lai. "Efficient routing on multilayered communication networks". *EPL (Europhysics Letters)* 102.2 (2013), page 28002.

[81] Y. Zhuo, Y. Peng, C. Liu, Y. Liu, and K. Long. "Traffic dynamics on layered complex networks". *Physica A: Statistical Mechanics and its Applications* 390.12 (2011), pages 2401–2407.

Appendix A: Complete Publications

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

Supplementary Information:

1.) Optimal transport with constraints: from mirror descent to classical mechanics Supporting Material. [38]
2.) Supplementary information for Community Detection in networks by Dynamical Optimal Transport Formulation [48]

algorithms

MDPI

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 (A.A.I.); alessandro.lonardi@tuebingen.mpg.de (A.L.); 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

check for updates

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 β 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 β 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_2 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 α, respectively, and $\mathcal{E}_{\alpha\gamma}$ is the set of edges between nodes in layer α and nodes in layer γ. Here, $\alpha = 1, \ldots, L$, where L is the number of layers. We denote with $N_\alpha = |\mathcal{V}_\alpha|$ the number of nodes in layer α, and with $E_\alpha = |\mathcal{E}_\alpha|$ the number of edges in layer α, $E_{\alpha\gamma} = |\mathcal{E}_{\alpha\gamma}|$ is the number of edges between nodes in layer α and γ. 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 $A^{\alpha}_{uv} = 1$ if there is an edge between nodes u and v in layer α, 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 $\hat{A}^{\alpha\gamma}_{uv} = 1$ if there is an edge between node u in layer α and node v in layer γ, and 0 otherwise. Without loss of generality, in our applications, we have $\hat{A}^{\alpha\gamma}_{uv} = 0$ 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_α exists for more than one value of α, we may not be able to tell whether the passengers entering u entered from u_α, u_γ 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_α, u_γ) and replace them with auxiliary *inter-super* edges (u_α, u), connecting all the instances u_α of node u with the super node u, as in a star graph, so that the original edge (u_α, u_γ) 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_α in layer α 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 $\alpha = 1, \ldots, L + 1$. 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_α as the inverse of a velocity, then ℓ_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 α than one in

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

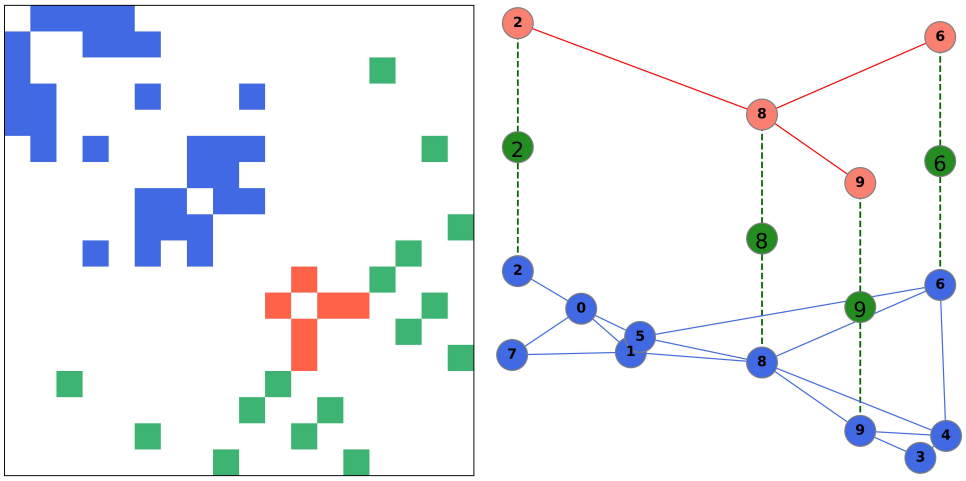


Figure 1. Example of multilayer structure. We show an example of a 2-layer network with $N = 18$ ($N_1 = 10$, $N_2 = 4$ and $N_{super} = 4$). (**Left**) adjacency matrix A, colors denote the layer type: blue is layer 1, red is layer 2 and green is the super layer. (**Right**) the 2-layer network with 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 μ_e, which are quantities determining the flux passing through an edge e. Intuitively, the conductivity μ_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 $a \in \mathcal{S}$, where $\mathcal{S} \subseteq \mathcal{V}$ is the set of stations where passengers enter, and we denote with $M = |\mathcal{S}|$ 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 μ_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:

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

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 $a \in \mathcal{S}$ 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:

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

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

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

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:

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

where $\Gamma(\beta_\alpha) = 2(2 - \beta_\alpha)/(3 - \beta_\alpha)$ for all α 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 β_{q_e} (taking value β_α on layer α) 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 α, 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 β_α 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_α and the β_α. 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 μ_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_β 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_β. 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 $\mathcal{O}(M\,N^2)$.

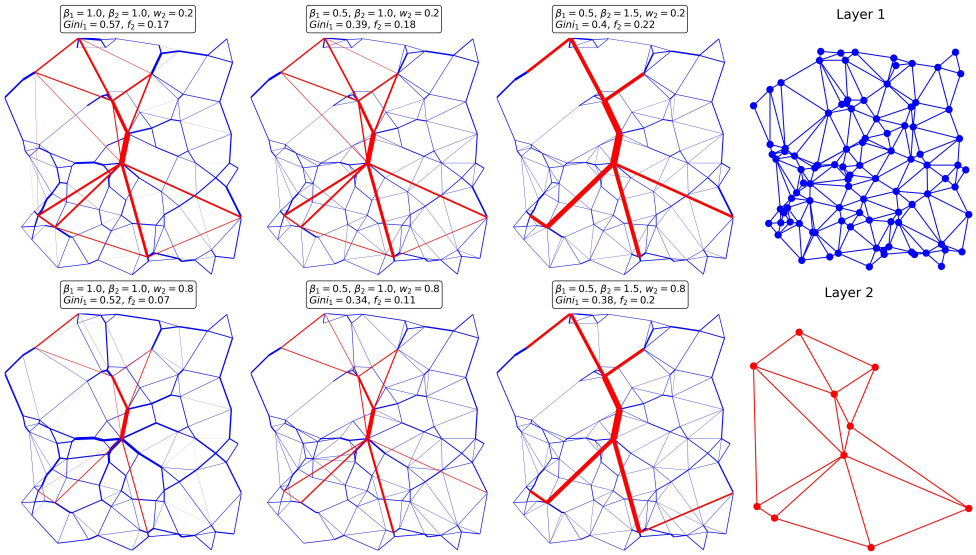


Figure 2. Example of optimal paths. We show an example of optimal paths obtained with: $p = 0.2$ and (**top**) $w_1 = 0.2$, (**bottom**) $w_1 = 0.8$. Values of β_1, β_2 are those reported on top of each network. The statistics $Gini_1$ and f_2 are those defined in Section 3.1. The width of edges is proportional 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.

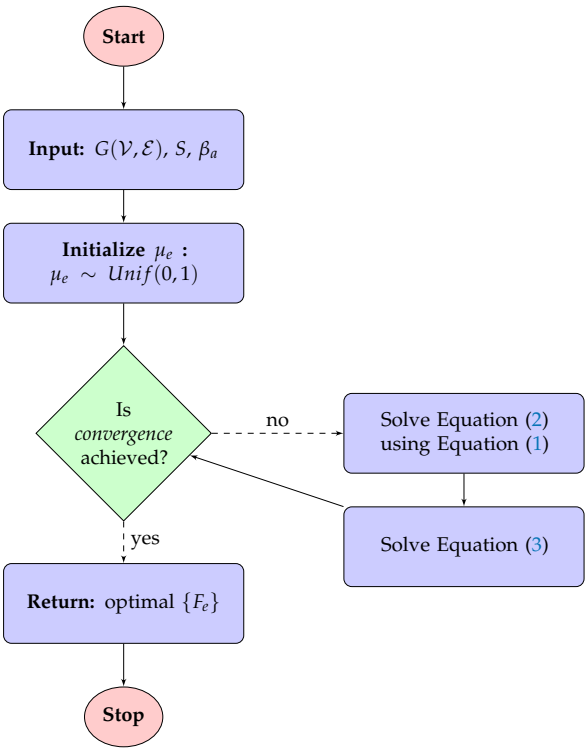


Figure 3. Block diagram of Algorithm 1. We give a pictorial representation of the pseudocode in Algorithm 1. Here, rectangular blocks are *action* blocks, corresponding to the update of a variable, to an input initialization, or to the output of the fluxes at convergence. *Conditional* blocks are diamond-shaped;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

Algorithm 1 Multilayer optimal transport.

Input: multilayer network $G(\mathcal{V}, \mathcal{E})$, source matrix S, β_α
Initialize: $\{\mu_e\}$ (e.g., sampling as i.i.d. $\mu_e \sim Unif(0,1)$)
while convergence not achieved **do**
    use Equation (1) to solve Kirchhoff's law as in Equation (2) $\rightarrow \{p_u^a\}$
    solve the dynamics in Equation (3): $\{\mu_e^t\} \rightarrow \{\mu_e^{t+1}\}$
end while
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 β. 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]:

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

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 β_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 β_1.

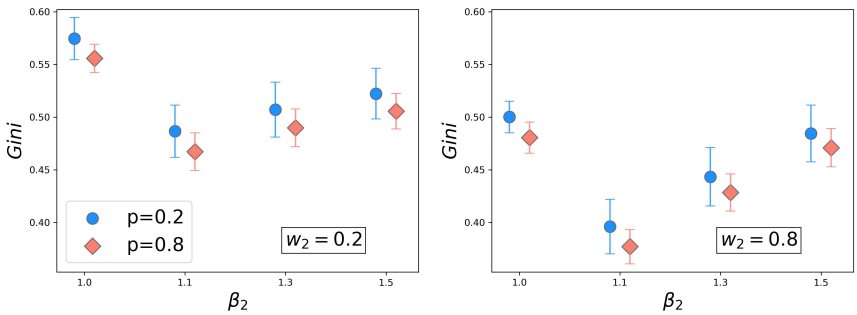


Figure 4. Results on synthetic data. We show the *Gini* w.r.t. the optimal $||F_e||_2$ (y axis) vs. β_2 (x axis) for synthetic 2-layer networks generated as in Section 3.1. Blue and red markers denote $p = 0.2, 0.8$, respectively, $w_1 = 1$ in all cases, while $w_2 = 0.2$ (**left**) and $w_2 = 0.8$ (**right**); $\beta_1 = 0.5$ 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 heterogenous 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 asses 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 $f_2 = \sum_{e\in\mathcal{E}_2} ||F_e||_2 / (\sum_{e\in\mathcal{E}_1} ||F_e||_1 + \sum_{e\in\mathcal{E}_2} ||F_e||_2)$ 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.

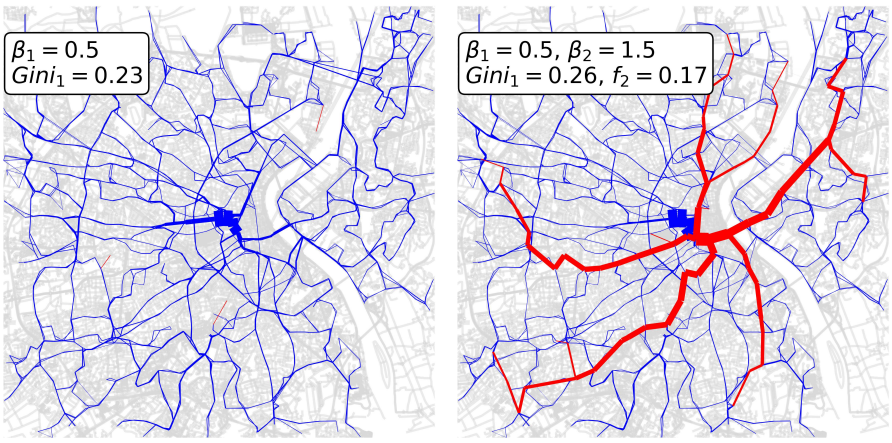


Figure 5. Example of optimal paths in the city of Bordeaux for a bus and tram network. The paths are obtained with (**left**) and without (**right**) the tram layer. Here, $\beta_1 = 0.5$ in both cases, while $\beta_2 = 1.5$ in the second case. The width of the edges is proportional to the optimal $||F_e||_2$.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

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 β_α 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 β_α. 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

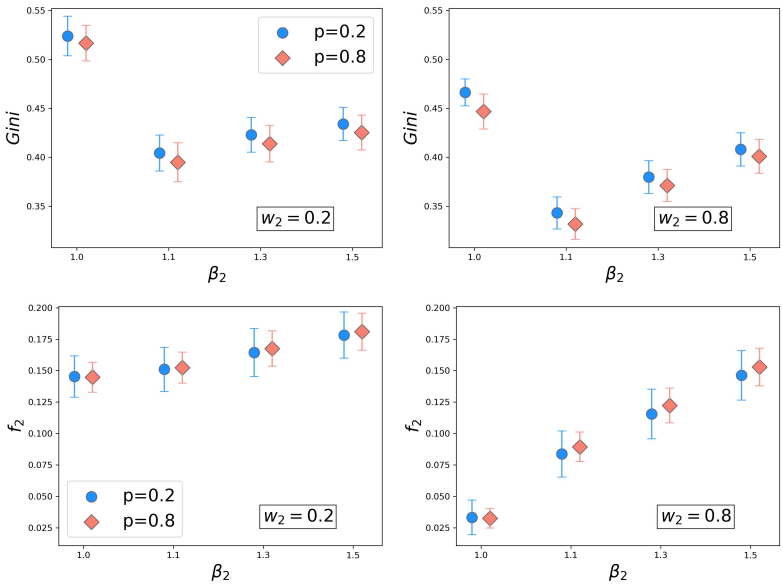


Figure A1. Additional results on synthetic data. We show the *Gini* w.r.t. the optimal *FBC* (**top**) and the total percentage flux f_2 on layer 2 (**bottom**) vs. β_2 (x axis), for synthetic 2-layer networks generated as in Section 3.1; $w_2 = 0.2, 0.8$ (**left**,**right**), $\beta_1 = 0.5$ in all cases, except for the case where $\beta_2 = 1$, for which $\beta_1 = 1$. This cases is a shortest path-like baseline. Markers are averages over 20 network samples and 50 source matrix samples.

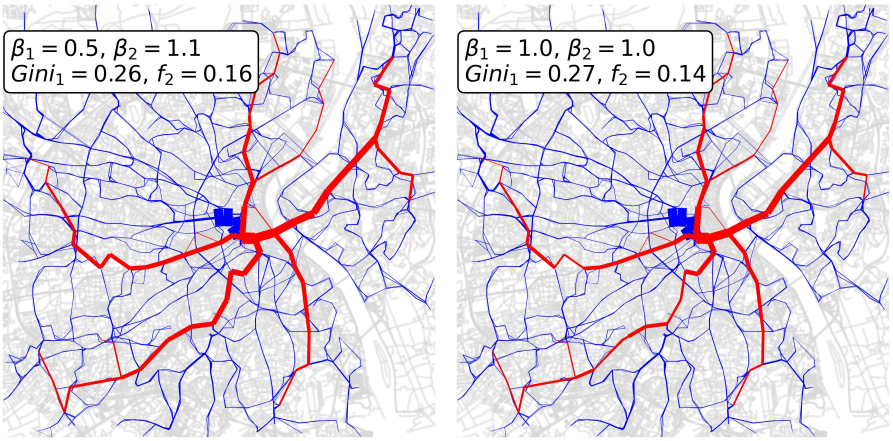


Figure A2. Additional example of optimal paths in the city of Bordeaux for a bus and tram network. Here $p = 0.0, w_2 = 0.2, \beta_1, \beta_2 = (0.5, 1.1), (1.0, 1.0)$ (**left**,**right**). The width of the edges is proportional to the optimal $||F_e||_2$. $Gini_1$ is calculated w.r.t. to the flux on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

References

1. Boccaletti, S.; Bianconi, G.; Criado, R.; Del Genio, C.I.; Gómez-Gardenes, J.; Romance, M.; Sendina-Nadal, I.; Wang, Z.; Zanin, M. The structure and dynamics of multilayer networks. *Phys. Rep.* **2014**, *544*, 1–122. [CrossRef]
2. Gómez-Gardenes, J.; Reinares, I.; Arenas, A.; Floría, L.M. Evolution of cooperation in multiplex networks. *Sci. Rep.* **2012**, *2*, 620. [CrossRef] [PubMed]
3. 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]
4. Saumell-Mendiola, A.; Serrano, M.Á.; Boguná, M. Epidemic spreading on interconnected networks. *Phys. Rev. E* **2012**, *86*, 026106. [CrossRef]
5. Dickison, M.; Havlin, S.; Stanley, H.E. Epidemics on interconnected networks. *Phys. Rev. E* **2012**, *85*, 066109. [CrossRef] [PubMed]
6. Chen, J.; Hu, M.B.; Li, M. Traffic-driven epidemic spreading in multiplex networks. *Phys. Rev. E* **2020**, *101*, 012301. [CrossRef] [PubMed]
7. Kurant, M.; Thiran, P. Layered complex networks. *Phys. Rev. Lett.* **2006**, *96*, 138701. [CrossRef] [PubMed]
8. Wu, J.; Pu, C.; Li, L.; Cao, G. Traffic dynamics on multilayer networks. *Digit. Commun. Netw.* **2020**, *6*, 58–63. [CrossRef]
9. 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]
10. Aleta, A.; Moreno, Y. Multilayer networks in a nutshell. *Annu. Rev. Condens. Matter Phys.* **2019**, *10*, 45–62. [CrossRef]
11. Bianconi, G. *Multilayer Networks: Structure and Function*; Oxford University Press: Oxford, UK, 2018.
12. Kivelä, M.; Arenas, A.; Barthelemy, M.; Gleeson, J.P.; Moreno, Y.; Porter, M.A. Multilayer networks. *J. Complex Netw.* **2014**, *2*, 203–271. [CrossRef]
13. Strano, E.; Shai, S.; Dobson, S.; Barthelemy, M. Multiplex networks in metropolitan areas: generic features and local effects. *J. R. Soc. Interface* **2015**, *12*, 20150651. [CrossRef]
14. Solé-Ribalta, A.; Gómez, S.; Arenas, A. Congestion induced by the structure of multiplex networks. *Phys. Rev. Lett.* **2016**, *116*, 108701. [CrossRef]
15. Aleta, A.; Meloni, S.; Moreno, Y. A multilayer perspective for the analysis of urban transportation systems. *Sci. Rep.* **2017**, *7*, 44359. [CrossRef] [PubMed]
16. 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]
17. 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]
18. 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]
19. Barthélemy, M. Spatial networks. *Phys. Rep.* **2011**, *499*, 1–101. [CrossRef]
20. Solé-Ribalta, A.; Arenas, A.; Gómez, S. Effect of shortest path multiplicity on congestion of multiplex networks. *New J. Phys.* **2019**, *21*, 035003. [CrossRef]
21. 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. Ronellenfitsch, 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.; Szöllősi, G.J.; Magnasco, M.O. Damage and Fluctuations Induce Loops in Optimal Transport Networks. *Phys. Rev. Lett.* **2010**, *104*, 048704. [CrossRef]
37. Ronellenfitsch, 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.; Ronellenfitsch, 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.; Kivelä, 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.; Bebber, 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]

Sustainable optimal transport in multilayer networks

Abdullahi Adinoyi Ibrahim,* Daniela Leite,† and Caterina De Bacco‡
Max Planck Institute for Intelligent Systems, Cyber Valley, Tübingen 72076, Germany

(Received 15 February 2022; accepted 18 May 2022; published 8 June 2022)

Traffic congestion is one of the major challenges faced by the transportation industry. While this problem carries a high economic and environmental cost, the need for an efficient design of optimal paths for passengers in multilayer network infrastructures is imperative. We consider an approach based on optimal transport theory to route passengers preferably along layers that are more carbon-efficient than the road, e.g., rails. By analyzing the impact of this choice on performance, we find that this approach reduces carbon emissions considerably compared to shortest-path minimization. Similarly, we find that this approach distributes traffic more homogeneously, thus alleviating the risk of traffic congestion. Our results shed light on the impact of distributing traffic flexibly across layers guided by optimal transport theory.

DOI: 10.1103/PhysRevE.105.064302

I. INTRODUCTION

Traffic congestion is a major problem in the transportation industry, with significant economic and environmental repercussions. The impacts of the environmental cost, such as carbon emissions and other air pollutants, on public health can be sizable and need to be properly studied [1]. Combining different transportation modalities, as in multilayer networks, can mitigate congestion and thus improve urban sustainability [2]. Modeling traffic congestion on multilayer networks is crucial to investigate the efficiency and cost of operating such infrastructures [3]. Addressing this problem requires extracting what paths passengers take from source to destination, information that can then be used to analyze traffic patterns. Many route extraction methods are based on shortest-path minimization [4–7] or assignment strategy [8]. However, the shortest path (i.e., selfish routing) might not always be the optimal path in a congested network [9–11], hence the need for coordinated traffic congestion. In addition, empirical results have shown that passengers may not always consider the shortest route [12,13]. While efforts have been made to go beyond shortest-path minimization using the cavity method or message-passing algorithms [10,14–17], these approaches are only valid in single-layer networks. In multilayer networks, several works focus more on analyzing the properties of passenger flows rather than proposing models to extract trajectories. They consider random walks [4] or shortest-path optimization [5–7] to extract flows, thus necessarily influencing the results of subsequent analysis based on these strategies. Fewer models have been targeting transport optimization in multilayer networks [3]. For instance, Ref. [8] proposed a flow-assignment strategy on multilayer networks, while Ref. [18] developed a recurrent algorithm for communication networks.

A principled and efficient approach for extracting optimal paths of passengers in networks is optimal transport (OT) theory [19–22]. This approach has been applied recently to multilayer networks [23], where the key idea is to flexibly tune between different cost functions in each of the different layers, thus capturing the specificity of each type of infrastructure. For instance, a road network is more sensitive to traffic congestion than a rail one, while the infrastructure of a rail network may be more costly to build. Our work builds from these ideas by adapting this model to study and evaluate optimal paths on multilayer networks under different scenarios. The goal of our work is to study the trajectories of optimal paths and compare them with those extracted from standard approaches relying on shortest-path minimization to identify key properties that are better optimized if one considers the multilayer character of the network. Our main contribution is threefold: First, we consider an optimal transport-based approach to extract optimal paths for passengers in multilayer networks, contrary to standard approaches based on shortest-path minimization. Second, we propose a variant of this OT-based method that interpolates between OT and shortest-path minimization. While the extracted paths of the two OT-based models are longer than those obtained by shortest-path minimization, the rail layer is used by more passengers. Finally, we show that by using the optimal routes extracted by OT-based algorithms, passengers are more likely to encounter little or no traffic while emitting less CO_2, leading to a reduced environmental cost. Our empirical results on synthetic and real data show the need for approaches that exploit the multilayer nature of multimodal transportation networks.

*abdullahi.ibrahim@tuebingen.mpg.de
†daniela.leite@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.

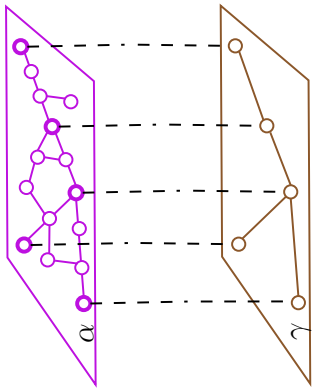


FIG. 1. Multilayer structure with $N_\alpha = 15$ and $N_\gamma = 4$. The network edges are represented by continuous lines (magenta and brown) and the two-edge path by dashed lines. The thicker magenta nodes represent stations belonging to both layers.

II. OPTIMAL TRANSPORT FOR TRAFFIC DISTRIBUTION IN MULTILAYER NETWORKS

We denote a multilayer network as a graph denoted as $G(\{\mathcal{V}_\alpha\}_\alpha, \{\mathcal{E}_\alpha\}_\alpha, \{\mathcal{E}_{\alpha\gamma}\}_{\alpha\gamma})$, where $\mathcal{V}_\alpha$, $\mathcal{E}_\alpha$, and $\mathcal{E}_{\alpha\gamma}$ denote the set of nodes, edges in layer α, and interlayer edges between layers α and γ; $\alpha = 1, \ldots, L$, where L is the number of layers. We denote the number of nodes and edges as N and E, and we assume that edges have length $l_e > 0$, which determines the cost of traveling through them. We consider the case of a two-layer network, but all results are valid for a higher number of layers. We denote the two layers as α, γ and consider a road network for α and a rail network for γ, as explained in detail in Sec. III. We show an example of this structure in Fig. 1.

We consider passengers traveling through the networks and distinguish them by their origin and destination (traffic demands) stations (o_i, t_i), where $o_i, t_i \in \mathcal{V} = \cup_\alpha V_\alpha$. We denote as $\mathcal{S} = \{(o_i, t_i)\}$ the set of all origin-destination pairs, and $|\mathcal{S}| = M$ denotes their number.

We briefly describe the model of [23] to find optimal paths in multilayer networks using optimal transport theory. It considers two main quantities on network edges: fluxes F_e of passengers traveling through edge e, and conductivities μ_e determining F_e passing through an edge e. To keep track of the various routes that passengers have, a multicommodity approach is considered [22,24] in which passengers are distinguished based on their entry station $i \in \mathcal{S}$. With this approach, the flux F_e is an M-dimensional vector, where entries F_e^i denote the flux of passengers of type i traveling on edge e. We assume the fluxes are determined by pressure potentials p_u^i and p_v^i defined on nodes as follows:

$$F_e^i := \frac{\mu_e}{l_e}\left(p_u^i - p_v^i\right), \quad e = (u, v), \tag{1}$$

where l_e is the length of edge e. Kirchhoff's law is imposed on network nodes to properly enforce mass conservation. Finally, the dynamics assumes that the conductivity μ_e depends on flux F_e as follows:

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

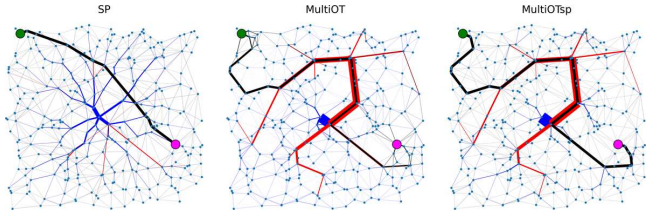


FIG. 2. Example trajectories. We show the trajectory of one type of passenger (black edges) whose origin and destination stations are the green and magenta nodes, respectively. We also highlight the total fluxes on edges, solutions of the OT problem including all other passengers, for a total of $M = 300$. Blue and red edges denote road (α) and rail (γ) layers, respectively. Edge widths are proportional to the amount of passengers traveling through an edge. The exact width of the black edge has been either increased (for SP) or reduced (for OT-based methods) in order to distinguish the flux of this type of passengers from the overall trajectories. Origin-destination pairs have been selected so that 80% of the passengers are directed towards a central node; $\beta_\alpha = 0.5$ and $\beta_\gamma = 1.9$.

where q_e encodes the layer to which the edge e belongs. The parameter $0 < \beta_{q_e} < 2$ determines the type of optimal transport problem one aims to solve: $0 < \beta_{q_e} < 1$ discourage traffic congestion, $1 < \beta_{q_e} < 2$ encourage path consolidation into few highways, while $\beta_{q_e} = 1$ is shortest-path-like. 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, and it decreases otherwise.

It can be shown [22,23] that the stationary solutions of Eq. (2) minimize the multilayer transport cost function:

$$J_\beta = \sum_{\alpha=1}^{L} \sum_{e\in\mathcal{E}_\alpha} l_e ||F_e||_2^{\Gamma(\beta_\alpha)}, \tag{3}$$

where $\Gamma(\beta_\alpha) = 2(2 - \beta_\alpha)/(3 - \beta_\alpha)$ for all α, and the 2-norm is calculated over the M entries of each F_e. Intuitively, solving the system of Eqs. (1) and (2) and Kirchhoff's law is equivalent to finding the optimal trajectories of passengers in a multilayer network, where optimality is given with respect to the transport cost in Eq. (3). We refer to this OT-based algorithm as MultiOT.

A. MultiOTsp: Interpolating between OT and shortest-path minimization

The paths extracted by MultiOT will encourage path consolidation along with the rail network and traffic minimization on the road network. Empirically, we observe that this model tends to distribute passengers of the same type (i.e., the same origin and destination) along various routes, as shown in Fig. 2. While most of these passengers take the shortest among these routes, some distribute on longer ones to prevent traffic congestion. This suggests an alternative algorithm that interpolates between MultiOT and shortest-path minimization to select only the main relevant routes for each origin-destination pair among those extracted by MultiOT. This can be done by inputting the solution of MultiOT for each passenger type i into a weighted shortest-path algorithm with edge weights

Algorithm 1 MultiOTsp

Input Graph $G(V, E, W)$, set $\mathcal{S}$ of origin-destination pairs, $\beta = (\beta_1, \ldots, \beta_L)$

Output Fluxes $\{F_e\}_e$

```
function MultiOTsp(G, S, β)
    {F_e}_e ← MultiOT(G, S, β)
    for i = 1, ..., M do
        {F_e^i}_e ← weighted Dijkstra (G, S, w) with w_e = l_e/|F_e^i|
    end for
    F_e = (F_e^1, ..., F_e^M), ∀ e
end function
```

defined as

$$w_e = \frac{l_e}{|F_e^i|}, \tag{4}$$

where the fluxes F_e are those extracted from MultiOT. All the passengers of type i are then routed along the output path. We call this algorithm MultiOTsp and show a pseudocode in Algorithm 1. The advantage of this weight function w_e is that weakly used edges are consolidated on others that are on the optimal path (according to OT) of many passengers. These edges thus become more desirable when designing an individual "consensus" OT-based path that takes into account both path length *and* optimal fluxes. The paths selected by MultiOTsp rely strongly on how the fluxes are selected in the first place to determine the weights w_e. As the F_e are calculated by considering all the passengers simultaneously (using MultiOT), the final optimal trajectories of MultiOTsp are significantly distinct from those obtained by shortest-path minimization, which are independent from the surrounding environment. We show an example of this in Fig. 2.

In the following, we study the trajectories of optimal paths extracted by the three approaches: MultiOT, MultiOTsp, and shortest-path minimization (SP). We use the implementation in [25] for MultiOT, while for SP we use the Dijkstra algorithm [26].

III. EMPIRICAL RESULTS

To investigate the relevant properties of the optimal paths extracted by the different algorithms, we simulate a variety of realistic traffic scenarios. Specifically, we generate a dataset of synthetic two-layer planar networks, where α simulates a road network and γ simulates a rail network (e.g., a tram). The layer α is constructed first by randomly placing N_α nodes in $[0, 1] \times [0, 1]$, and extracting its Delaunay triangulation [27]. We then select among them a subset of N_γ nodes to build the layer γ with an analogous procedure, thus ensuring that the two layers are connected. In total, in this construction the multilayer network has $N = N_\alpha$ nodes and resembles the situation in which all the stations in the second layer also have access to the road network. Notice that other constructions are possible, but this choice does not impact the validity of our model. In our simulations, we set $N_\alpha = 300$ and $N_\gamma = 60$. We extract 20 different networks and 100 random samples of origin-destination pairs for each of them, for a total of 2000 realizations for each parameters' configuration. With

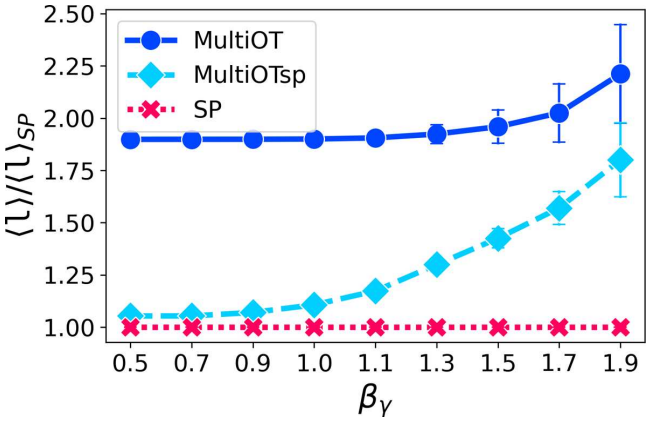


FIG. 3. Average total path-length ratio. We show the ratio of the average total path length to the one extracted from SP. We set $p = 0.5$, $\beta_\alpha = 0.5$, and vary $0 < \beta_\gamma < 2$. The results are averaged over 20 different network realizations with 100 randomly selected origin-destination pairs for each network realization. The markers and error bars are averages and standard deviations.

this, we aim at capturing different transportation scenarios in the two layers, as rail networks are less subject to traffic congestion but more costly to build, while we can state the opposite for road networks. MultiOT (and thus MultiOTsp) can capture these differences by suitably tuning β in each layer: to discourage traffic congestion in the road layer, we set $\beta_\alpha = 0.5$ and vary β_γ in $0 < \beta_\gamma < 2$ to study various scenarios. In realistic scenarios, passengers have different origins and destinations; see Fig. 2 for an example. As we may expect in many urban scenarios that the most frequent destination is located in the city center, we assign to each passenger type its destination by default to be a central node. Then, to explore alternative scenarios where destinations are more heterogeneous, we consider a rewiring probability $p = [0, 1]$ to rewire its destination at random. Specifically, for each passenger type, we rewire its destination to a random node with probability p. Hence, $p = 0.0$ corresponds to having a monocentric destination where all passengers move towards a central node and $p = 1.0$ corresponds to selecting all passen-

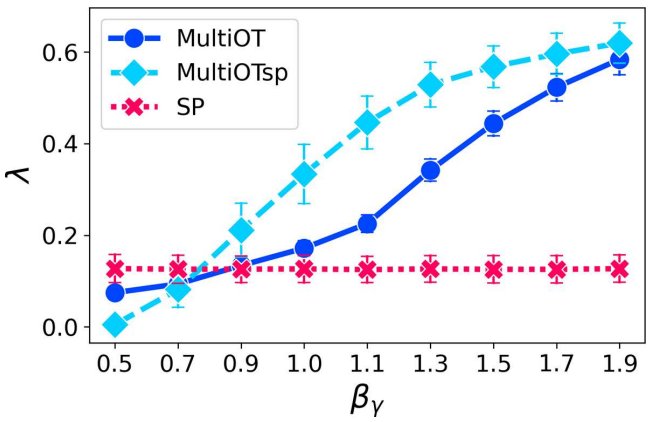


FIG. 4. Coupling between layers. We show the coupling coefficient as defined in Eq. (5). All other settings remain the same as in Fig. 3.

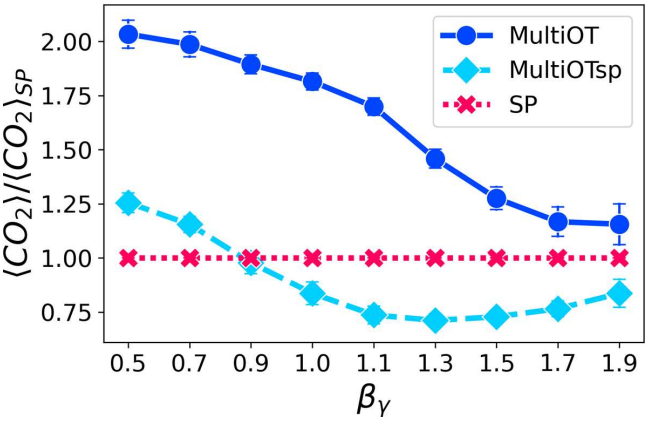


FIG. 5. Carbon emission ratio. We show the ratio of the average carbon emissions as defined in Eq. (6) to that obtained by SP. All other settings remain the same as in Fig. 3.

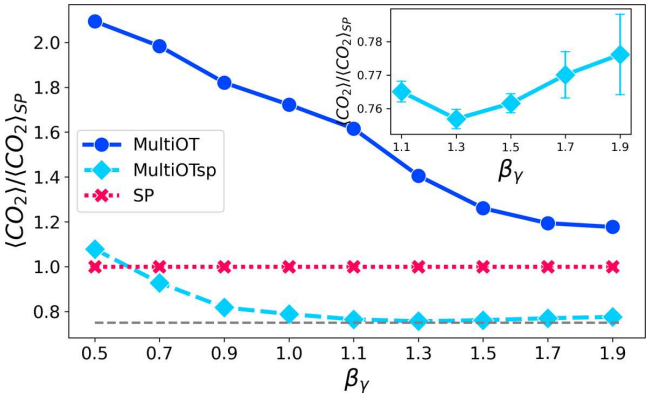


FIG. 7. Carbon emissions on the Bordeaux network. Ratio of the average carbon emission over that of SP. Here, we set $p = 0.2$, thus favoring monocentric destinations. The gray-dashed line shows the minimum value obtained byMultiOTsp, corresponding to $\beta_\gamma = 1.3$. Inset: zoom-in of MultiOTsp values for $1.1 \leqslant \beta_\gamma \leqslant 1.9$.

gers' destinations at random. We consider $p = \{0.2, 0.5, 0.8\}$, but we show results for $p = 0.5$, as the qualitative behavior is similar to that for the others; see Appendix. These settings exhibit three important properties of the OT-based algorithms.

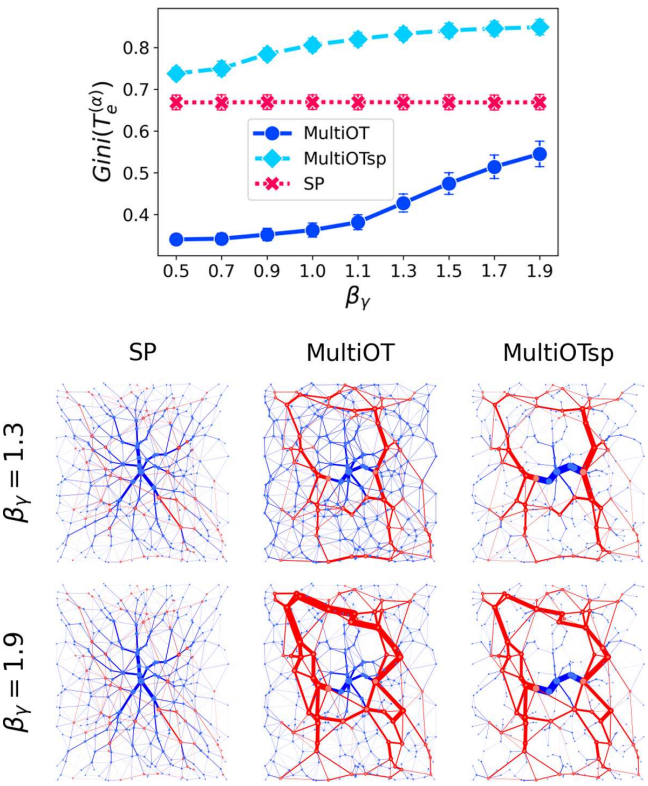


FIG. 6. Traffic distribution. Top: Gini coefficient of the traffic on the road layer α. Bottom: optimal trajectories. Edge widths are proportional to $||F_e||_1$ averaged over 10 samples of origin-destination configurations; $p = 0.5$, $\beta_\alpha = 0.5$. When $\beta_\gamma > 1$, OT-based methods are consolidated into fewer edges in layer γ. The SP, on the other hand, is not affected by this parameter, showing flows of passengers on more edges compared to OT-based methods, which consolidate into fewer edges. Blue and red edges correspond to road (α) and rail (γ) layers, respectively.

IV. LONGER LENGTHS BUT HIGHER RAIL NETWORK USAGE

Shortest-path optimization is utilized to minimize the total path length taken by passengers, hence we expect MultiOT and MultiOTsp to underperform SP on this task. In fact, the performance of OT-based algorithms is expected to decrease as β_γ increases, as shown in Fig. 3 by the average path length $\langle l \rangle = \frac{1}{M}\sum_{e\in\mathcal{E}} l_e||F_e||_1$ over the one obtained from a shortest-path algorithm.

This is expected given that higher β_γ encourages more traffic to be routed towards the rail network at the cost of increased distance to cover, as the rail network has fewer and more distant nodes to reach than a road network. We then measure how passengers are distributed in the two layers by defining a coupling coefficient, a known concept to describe how well two layers are linked [6]. We define

$$\lambda = \frac{1}{M}\sum_{i\in\mathcal{S}}\left(\frac{\sum_{e\in\mathcal{E}_\gamma}|F_e^i|}{\sum_{e\in\mathcal{E}_\alpha\cup\mathcal{E}_\gamma}|F_e^i|}\right), \tag{5}$$

where the numerator inside the parentheses contains only the flux in the rail layer so we can distinguish how many passenger types effectively use that layer in their trajectories. This definition is valid for two-layer networks, such as the empirical networks studied here. However, one can appropriately generalize it for networks with more than two layers. The usage of the rail layer increases monotonically for both OT-based algorithms, as shown in Fig. 4, with MultiOTsp reaching higher usage values. This suggests that the shortest-path routes selected from the possible paths output by MultiOT are composed of a significant amount of rail edges. This also shows that the raw solution output of MultiOT consider paths more distributed across the road layer, as qualitatively observed in Fig. 2.

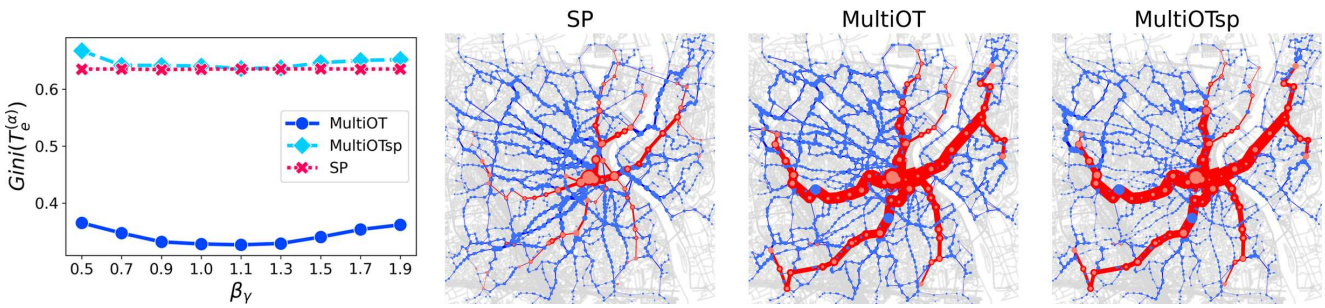


FIG. 8. Traffic distribution on the Bordeaux network. We set $p = 0.2$ and $\beta_\alpha = 0.5$. Left: Gini coefficient calculated using the traffic on the road network. Error bars (standard deviation) are smaller than marker size. Center-right: Traffic distribution for each of the algorithms with $\beta_\gamma = 1.9$. Red and blue edges denote tram and road layers, respectively. Node and edge sizes are proportional to the amount of passengers traveling through them. The results are averaged over 100 samples of origin-destination pairs.

V. MULTILAYER OPTIMIZATION CAN DECREASE CARBON CONSUMPTION

As more passengers take longer paths while being encouraged to use the rail network, they also consume less carbon per unit of length. The question is whether the increased length can be properly compensated by the decrease in carbon consumption. We tested this on the same paths extracted to plot Fig. 3 by measuring the average CO_2 consumption per passenger as

$$\langle CO_2 \rangle = \frac{1}{M} \sum_{e \in \mathcal{E}} r_{q_e} l_e ||F_e||_1, \tag{6}$$

where r_α is the carbon emission rate in layer α. This has a dimension of unit of mass (e.g., g) per passenger per unit of length (e.g., pkm). For instance, a bus on average generates 101.87 g/pkm [28] while a train generates 28.39 g/pkm [29]. Hence, defining r_α as the rate of the road layer and considering buses traveling on it, we can set $r_\gamma = 28.39 r_\alpha / 101.87 = 0.28\, r_\alpha$. These values can be changed accordingly with more specific values if a traffic manager has precise statistics of vehicle types traveling on the network. By leveraging optimal transport with a bias towards shortest paths, MultiOTsp is able to decrease carbon consumption the most compared to SP, measured by the ratio of its $\langle CO_2 \rangle$ over that produced by SP. A minimum is reached for $1.1 \leqslant \beta_\gamma \leqslant 1.5$ where MultiOTsp produces 25% fewer emissions than a shortest-path routing algorithm, as shown in Fig. 5. This important result is a consequence of flexibly tuning the cost to be optimized in each layer, as allowed by β in Eq. (3). In particular, $\beta_\gamma > 1$ encourages paths to consolidate on the rail layer, while $\beta_\alpha = 0.5$ controls for traffic congestion on the road layer. The fact that the minimum consumption of MultiOTsp has not been realized at the highest value $\beta_\gamma = 1.9$, where the paths are consolidated into the fewest rail routes, further suggests that there is a tradeoff between keeping the path lengths short while directing more passengers towards the rail layer. In fact, at $\beta_\gamma = 1.9$, as the number of passengers redirected towards the second layer increases, they also have to take longer routes, thus emitting more carbon. A value of $\beta = 1.3$ results in a nice tradeoff between these two competing behaviors in terms of carbon emission. On the contrary, MultiOT shows a monotonic decreasing behavior with a minimum reached at $\beta = 1.9$, but still higher than that emitted by SP. This is a consequence of the higher number of possible paths that passengers can take as routed by MultiOT, which are by default longer than those obtained by MultiOTsp and use more edges of the road layer. As a consequence, the longer length does not seem to justify the higher usage of the rail layer.

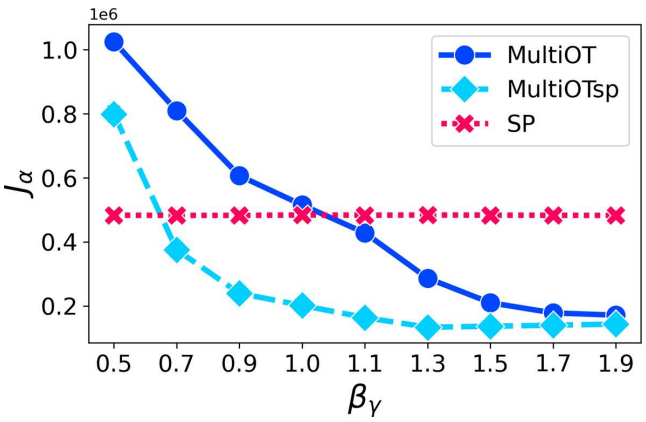


FIG. 9. Transport cost on the road layer α of the Bordeaux network. The cost is defined as in Eq. (8); here $p = 0.2$.

VI. TRAFFIC CONGESTION

All the results of the previous section were interpreted with the assumption that the flow of passengers is regular, even on high-traffic edges. Instead, if we account for traffic slowing down the flow on edges with a high density of travelers, those vehicles emit more carbon while they keep their engines on longer. The routes suggested by MultiOT are less sensitive to this, hence we also expect a lower carbon emission than that shown in Fig. 5 when accounting for traffic. We thus measure traffic load on edges as

$$T_e = \frac{1}{n} ||F_e||_1, \tag{7}$$

where n is the total number of passengers and measures the Gini coefficient Gini $(T_e) \in [0, 1]$ as a global network metric of inequality of how traffic is distributed on the network [30], with a Gini close to 1 meaning high inequality in flow assignment along edges. As the road layer is the one more sensitive to potential traffic bottlenecks, we consider only the

traffic on road edges and denote with Gini ($T_e^{(\alpha)}$) the Gini coefficient calculated using only $e \in \mathcal{E}_\alpha$. As expected, MultiOT has more balanced traffic than the other two algorithms, as shown in Fig. 6. While congestion increases with β_γ, even at the maximum $\beta = 1.9$ the Gini coefficient is lower than that of SP. The reason for this increase is that paths consolidate more on those fewer road edges that allow a connection to the rail layer, as can be seen on the example optimal routes in Fig. 6, a behavior also observed in previous studies [31,32]. This is exacerbated in MultiOTsp, as one can notice that central road edges are overly trafficked when many passengers exit the rail to reach the final destination in the center. This also causes the Gini coefficient of MultiOTsp to be higher than that of SP. In other words, few central edges cause most of the traffic for MultiOTsp. This can be partially alleviated by increasing p towards 1 as fewer destinations are directed towards the network center, although this may become an unrealistic assumption in urban scenarios. Alternatively, one can simply add rail stations in the center, so that passengers do not have to commute one extra mile to reach their final destination, a scenario that we explore below in the case of a real network.

VII. REAL MULTILAYER NETWORK

Next, we examine these properties on a real two-layer network of the city of Bordeaux [33], where α and γ represent the bus and tram networks, respectively. Similar to the synthetic network, we compare the performances of OT-based algorithms with SP on this network. We set $p = 0.2$ to consider the situation in which the majority of passengers are directed towards the city center, a central node coinciding with a tram station, and we extract 100 realizations of origin-destination pairs.

The tram in this city travels through its own reserved lanes, independently from other vehicles. Hence, although the two layers are physically located next to each other, there is no mixing of fluxes from the two layers on edges. This may differ in other real situations. While in principle our model is best suited for independent usage of the space by the various layers (e.g., road and subway or the case studied here), the results shown here may still apply if we assume that the physical presence of the tram only marginally impacts the traffic in the road layer, compared to other vehicles. Specifically, the combination of high enough capacity (number of passengers that can fit into a tram) and lower frequency of trams than other vehicles on the road may allow us to assume independence between the tram and road layer. In fact, while the tram may have many passengers traveling at any given time along an edge, these are all entering inside the same wagons. Thus, the space occupied by the tram is limited by its physical shape. Instead, the same amount of passengers would need to distribute in many different cars, thus occupying much more space, potentially creating congestion. In general, in situations in which trams have reserved lanes that cars cannot enter, our treatment applies without any further assumption.

We find that MultiOTsp produces 25% fewer carbon emissions than SP for $1.1 \leqslant \beta_\gamma \leqslant 1.5$, as shown in Fig. 7, similar to what is observed on synthetic networks. MultiOT has a minimum at $\beta_\gamma = 1.9$, but the emissions are higher than SP. We argue that also in this case this is due to the assumption

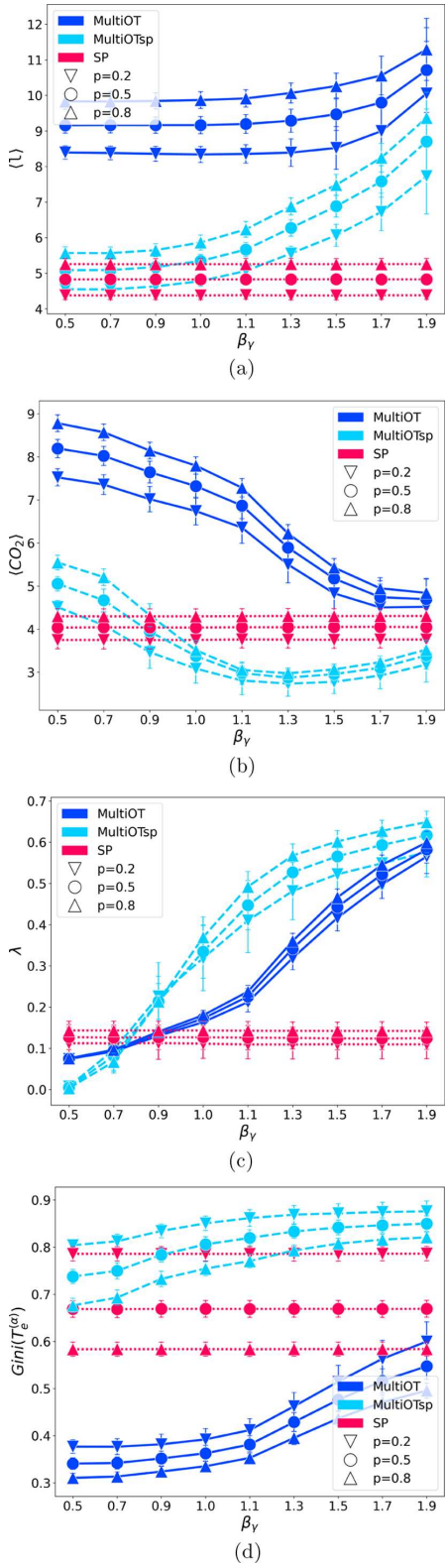


FIG. 10. Additional results on synthetic data for varying p. The results are averages and standard deviations over 20 different network realizations with 100 independent samples of origin-destination pairs on each network realization. Other parameters used here are $\beta_\alpha = 0.5$, $N_\alpha = 300$, $N_\gamma = 60$.

that the flow of vehicles is smoothly moving, with no traffic congestion causing velocity to decrease, and thus causing emissions to increase nearby traffic bottlenecks. To assess this hypothesis, we investigate the distribution of fluxes on the road layer by measuring the traffic T_e on edges in layer α. We find that indeed MultiOT has path trajectories more homogeneously distributed across the road layer as measured by the Gini coefficient plotted in Fig. 8 along with an example solution, potentially lowering the number of traffic jams. As we can see from an example solution in the same figure, the two OT-based variants distribute passengers in higher amounts along the tram network, thus lowering the road's usage, while SP makes use of the tram mainly in the vicinity of the central node. We can further notice how MultiOT uses the road with higher intensity than MultiOTsp, but the road edges have less traffic than those obtained by SP. As for MultiOTsp, the road edges with the most traffic are those near tram stations, like those in the upper left corner in the figure.

To better quantify the potential impact of traffic congestion as a proxy for the potential increase in CO_2 emission, we consider a measure of transport cost used before in similar problems [14,15] and defined as

$$J_\alpha = \sum_{e\in\mathcal{E}_\alpha} l_e ||F_e||_1^2, \tag{8}$$

where the exponent 2 discourage traffic. This should not be confused with the definition of Eq. (3), which is the one used to extract the optimal paths in our model, i.e., to solve the OT problem. Specifically, Eq. (3) uses different β for edges in different layers. In particular, it allows us to encourage both path consolidation in one layer and path distribution in another. Instead, Eq. (8) only discourage traffic, as the exponent is greater than 1. In addition, Eq. (3) considers the norm-2 of the passengers' flows, while Eq. (8) considers norm-1. While the latter is more intuitive, as it is the total number of passengers traveling along an edge, the former admits rigorous theoretical guarantees for OT to converge to an optimal solution. This does not apply to a cost function using norm-1; see Ref. [24] for a detailed discussion.

As seen in Fig. 9, both OT-based algorithms outperform SP as β_γ increases, meaning that passengers traveling on paths generated by the OT-based algorithms will generally record less road traffic congestion compared with the paths extracted by SP. Assuming that velocity decreases along congested edges, we conjecture that this would result in MultiOT having lower carbon emissions than SP.

VIII. DISCUSSION AND CONCLUSION

Designing and extracting optimal passenger flows in a transportation network is crucial for reducing traffic congestion and environmental costs. Methods based on shortest-path optimization are optimal in terms of reducing the average shortest-path length to reach destination, but they may fail in terms of other relevant transportation metrics. In addition, passengers do not always follow the shortest route [12], hence the need for alternative approaches to extract path trajectories and investigate their properties in multilayer networks. We present two models based on optimal transport theory that can flexibly tune the amount of traffic routed in the different layers to encourage usage of rail networks while reducing traffic on the road. As a result, optimal trajectories extracted with these methods significantly decrease the amount of carbon emissions compared to shortest-path minimization, while also being more robust to traffic congestions. In particular, we found that MultiOTsp, by interpolating between optimal transport and shortest-path minimization, can achieve the lowest amount of carbon emissions under the hypothesis of smooth flow of passengers in a network. Instead, MultiOT, based purely on optimal transport, distributes paths more homogeneously, thus being potentially more robust against increased carbon emissions when accounting for passengers' flow slowing down along traffic bottlenecks. This can be tested quantitatively in real scenarios by having access to empirical data of different velocities during traffic congestion, along with detailed velocity limits imposed by regulations in different parts of the network. One could potentially compare the theoretical results with the empirical ones observed from real data as in [34].

In general, we show that models based on optimal transport can be used to design optimal routes for passengers in a multilayer network, and we investigate scenarios beyond those obtained by using standard shortest-path algorithms. In this work, we assumed fixed origin-destination pairs, but one can further generalize this analysis by considering dynamical traffic demands that change in time. This would require suitably adapting the models studied in this work to account for this, for instance borrowing ideas from [35–39]. Similarly, we did not explore here the possibility of traffic diversions due to road blockages or changing conditions in the network structure [40–43]. Studying the robustness of the methods investigated in this work to these scenarios would be an interesting subject for future work. Finally, it would be interesting to investigate more complex scenarios with more than two layers, possibly on a larger scale than that of a unique urban scenario. To facilitate future analysis, we provide an open source implementation of our code at [25].

ACKNOWLEDGMENTS

The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting A.A.I. and D.L.

APPENDIX: ADDITIONAL RESULTS VARYING *p*

We set $p = \{0.2, 0.5, 0.8\}$ to capture different traffic demand scenarios, where $p = 0.2$ and 0.8 correspond to having the majority and minority of the passengers with a monocentric destination. We show in Fig. 10 the performance of the algorithms in terms of the same metrics investigated in the body of the manuscript. All displayed results have the same settings described in Sec. III.

[1] P. J. Landrigan, R. Fuller, N. J. Acosta, O. Adeyi, R. Arnold, A. B. Baldé, R. Bertollini, S. Bose-O'Reilly, J. I. Boufford, P. N. Breysse *et al.*, The Lancet **391**, 462 (2018).
[2] L. G. N. Orozco, L. Alessandretti, M. Saberi, M. Szell, and F. Battiston, arXiv:2111.02152.
[3] J. Wu, C. Pu, L. Li, and G. Cao, Digital Commun. Netw. **6**, 58 (2020).
[4] M. De Domenico, A. Solé-Ribalta, S. Gómez, and A. Arenas, Proc. Natl. Acad. Sci. (USA) **111**, 8351 (2014).
[5] A. Lampo, J. Borge-Holthoefer, S. Gómez, and A. Solé-Ribalta, Phys. Rev. Research **3**, 013267 (2021).
[6] R. G. Morris and M. Barthelemy, Phys. Rev. Lett. **109**, 128703 (2012).
[7] A. Solé-Ribalta, S. Gómez, and A. Arenas, Phys. Rev. Lett. **116**, 108701 (2016).
[8] L. Gao, P. Shu, M. Tang, W. Wang, and H. Gao, Phys. Rev. E **100**, 012310 (2019).
[9] D. A. Lazar, S. Coogan, and R. Pedarsani, IEEE Trans. Autom. Contr. **66**, 2664 (2020).
[10] H. F. Po, C. H. Yeung, and D. Saad, Phys. Rev. E **103**, 022306 (2021).
[11] N. Mehr and R. Horowitz, in 2018 IEEE Conference on Decision and Control (CDC) (IEEE, Piscataway, NJ, 2018), pp. 1788–1793.
[12] D. Quercia, R. Schifanella, and L. M. Aiello, in *Proceedings of the 25th ACM Conference on Hypertext and Social Media* (ACM, 2014), pp. 116–125.
[13] Z. Zheng, Z. Huang, F. Zhang, and P. Wang, EPJ Data Sci. **7**, 23 (2018).
[14] C. H. Yeung and D. Saad, Phys. Rev. Lett. **108**, 208701 (2012).
[15] C. H. Yeung, D. Saad, and K. M. Wong, Proc. Natl. Acad. Sci. (USA) **110**, 13717 (2013).
[16] F. Altarelli, A. Braunstein, L. Dall'Asta, C. De Bacco, and S. Franz, PLoS ONE **10**, e0145222 (2015).
[17] C. De Bacco, S. Franz, D. Saad, and C. H. Yeung, J. Stat. Mech.: Theor. Expt. **2014**, P07009 (2014).
[18] J. Zhou, G. Yan, and C.-H. Lai, Europhys. Lett. **102**, 28002 (2013).
[19] V. Bonifaci, K. Mehlhorn, and G. Varma, J. Theor. Biol. **309**, 121 (2012).
[20] D. Baptista, D. Leite, E. Facca, M. Putti, and C. De Bacco, Sci. Rep. **10**, 20806 (2020).
[21] D. Baptista and C. De Bacco, R. Soc. Open Sci. **8**, 210025 (2021).
[22] A. Lonardi, E. Facca, M. Putti, and C. De Bacco, Phys. Rev. Research **3**, 043010 (2021).
[23] A. A. Ibrahim, A. Lonardi, and C. D. Bacco, Algorithms **14**, 189 (2021).
[24] A. Lonardi, M. Putti, and C. De Bacco, Sci. Rep. **12**, 7474 (2022).
[25] https://github.com/cdebacco/multiot.
[26] E. W. Dijkstra *et al.*, Numer. Math. **1**, 269 (1959).
[27] L. Guibas and J. Stolfi, ACM Trans. Graph. **4**, 74 (1985).
[28] DEFRA, Department for environment, food and rural affairs. greenhouse gas reporting: Conversion factors (2017).
[29] EEA, Energy efficiency and specific co2 emissions (2017).
[30] P. M. Dixon, J. Weiner, T. Mitchell-Olds, and R. Woodley, Ecology **68**, 1548 (1987).
[31] P. S. Chodrow, Z. Al-Awwad, S. Jiang, and M. C. González, PLoS ONE **11**, e0161738 (2016).
[32] E. Strano, S. Shai, S. Dobson, and M. Barthelemy, J. R. Soc. Interface **12**, 20150651 (2015).
[33] R. Kujala, C. Weckström, R. K. Darst, M. N. Mladenović, and J. Saramäki, Sci. Data **5**, 180089 (2018).
[34] E. Taillanter and M. Barthelemy, J. R. Soc. Interface **18**, 20210391 (2021).
[35] C. H. Yeung, Phys. Rev. E **99**, 042123 (2019).
[36] A. Lonardi, E. Facca, M. Putti, and C. De Bacco, arXiv:2112.10620.
[37] D. Hu and D. Cai, Phys. Rev. Lett. **111**, 138701 (2013).
[38] F. Corson, Phys. Rev. Lett. **104**, 048703 (2010).
[39] D. Hu, D. Cai, and A. V. Rangan, PLoS ONE **7**, 1 (2012).
[40] T. S. Tai and C. H. Yeung, Phys. Rev. E **104**, 024311 (2021).
[41] K.-K. Kleineberg, L. Buzna, F. Papadopoulos, M. Boguñá, and M. Á. Serrano, Phys. Rev. Lett. **118**, 218301 (2017).
[42] S. Wandelt, X. Sun, D. Feng, M. Zanin, and S. Havlin, Sci. Rep. **8**, 13513 (2018).
[43] Z. Wang, D. Zhou, and Y. Hu, Phys. Rev. E **97**, 032306 (2018).

Optimal Transport with Constraints: From Mirror Descent to Classical Mechanics

Abdullahi Adinoyi Ibrahim,* Michael Muehlebach,† and Caterina De Bacco‡
Max Planck Institute for Intelligent Systems, Cyber Valley, Tübingen 72076, Germany

(Received 22 August 2023; accepted 1 July 2024; published 30 July 2024)

Finding optimal trajectories for multiple traffic demands in a congested network is a challenging task. Optimal transport theory is a principled approach that has been used successfully to study various transportation problems. Its usage is limited by the lack of principled and flexible ways to incorporate realistic constraints. We propose a principled physics-based approach to impose constraints flexibly in optimal transport problems. Constraints are included in mirror descent dynamics using the D'Alembert-Lagrange principle from classical mechanics. This results in a sparse, local and linear approximation of the feasible set leading in many cases to closed-form updates.

DOI: 10.1103/PhysRevLett.133.057401

Introduction—Optimal transport in networks has important applications in different disciplines, in particular in urban transportation networks [1]. Congestion not only increases travel time for users and decreases productivity, but it also drives air pollution. Reducing congestion and making transportation more efficient are also a core objective for EU policies, as highlighted throughout the EU Transport White Paper and the Strategic Plan 2020-2024 [2,3].

The design of efficient transportation networks is a complex task that requires a multifaceted solution. One of these facets is the problem of finding optimal routes for passengers. This is a well-studied problem in operations research [4] where minimum-cost optimization is often considered to model discrete flows and can be solved using classical techniques from linear programming. In our Letter, we consider the continuous case, where flows are real-valued quantities. A variety of approaches have been suggested to model transport in networks using techniques from physics of complex systems [5,6]. Path optimality and congestion control have been studied in discrete settings [7–9] or using the cavity method [10,11]. These usually rely on *ad hoc* algorithmic updates that depend on the specific type of constraints. The computational complexity of the *ad hoc* updates is greatly influenced by the constraints. Other approaches have been proposed to investigate navigation in complex systems [12–18], where the focus lies on investigating the properties of flows, rather than their optimization, as we consider here. In addition, these models often assume that passengers follow their shortest paths, an assumption, which may not be satisfied in practice. Adaptation dynamics [19–21] have been proposed to model biological distribution networks. However, these methods fall short of describing realistic scenarios where transport flows are limited by constraints.

In the following we cast the problem of designing efficient transportation networks under the broader framework of optimal transport theory (OT) [22]. This has been used to model and optimize various aspects of transport networks such as network design [19,21,23,24] and traffic flows [25–29]. These approaches guarantee a principled and computationally efficient way of solving transportation problems on networks. In addition, they model traffic congestion with a single tuning parameter that enables a transition between opposite traffic regimes, where traffic congestion can either be consolidated or discouraged. In standard OT methods, beyond few obvious constraints (e.g., conservation of mass), the amount of flow passing through an edge of the transportation network is unconstrained. As a result, traffic tends to concentrate on path trajectories that may be structurally unfeasible, which severely limits the applicability of OT models in real-world situations, where, for example, roads have a limited capacity of vehicles traveling at the same time. This Letter proposes an approach to avoid this crucial flaw of OT models by imposing constraints. Applying this approach significantly impacts the overall network topology induced by the optimal flows, as the resulting path trajectories have different path lengths and traffic distribution than those obtained from unconstrained scenarios.

Our approach has not only a solid foundation via the principle of D'Alembert-Lagrange from classical mechanics [30], but also leads to algorithms that are computationally

*Contact author: abdullahi.ibrahim@tuebingen.mpg.de
†Contact author: michaelm@tuebingen.mpg.de
‡Contact author: 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.

efficient and have a low implementation complexity. The key idea is to consider mirror descent dynamics of an OT problem, where constraints are included on a velocity level. This leads to a sparse, local and linear approximation of the feasible set which, in many cases, allows for a closed-form update rule, even in situations where the feasible set is nonconvex.

The model—In analogy with electrical grids or hydraulic networks, we model mass flow on a transportation network using conductivities and flows on network edges. We consider a multicommodity scenario [26,31], where mass of different type $i = 1, \ldots, M$ can move along different trajectories. The flow F_e^i of mass of type i along an edge $e = (u, v)$ can be described by $F_e^i = \mu_e(p_u^i - p_v^i)/\ell_e$, where p_u^i is a pressure potential at node u for passenger of type i, ℓ_e is the length of the edge e, and μ_e its conductivity. This latter quantity can be seen as proportional to the size of an edge, and is the main variable of interest in determining optimal trajectories. Once the conductivity is known, the pressure differences can then be calculated from Kirchhoff's law, which in turn determines the flows F_e^i, see Supplemental Material (SM) [32], which includes Refs. [33–36]. In the absence of constraints, the optimal conductivities are the stationary solutions of the dynamics $\dot{\mu} = f$, where

$$f_e = \mu_e^\beta \frac{\sum_i (p_u^i - p_v^i)^2}{\ell_e^2} - \mu_e \equiv \mu_e^{\beta-2}|F_e|^2 - \mu_e, \quad (1)$$

with $F_e = (F_e^1, \ldots, F_e^M)$ and $|\cdot|$ denotes the Euclidean norm. Intuitively, this equation describes a positive feedback mechanism where conductivities increase for larger fluxes and decrease for negligible ones [19]. It can be shown that the dynamics in Eq. (1) admits a Lyapunov function $\mathfrak{L}_\beta$ which can be interpreted as a combination of the cost to operate the network and that of building the infrastructure [26], see SM [32]. Moreover, we have that $f = -S\nabla\mathfrak{L}_\beta$, where S is a diagonal matrix with diagonal entries $S_e = 2\mu_e^\beta/\ell_e$ and Eq. (1) can therefore be seen as a mirror descent for the cost function $\mathfrak{L}_\beta$ [37]. This scaling in S has the advantage of ensuring good behavior of the resulting numerical methods. One can also reinterpret Eq. (1) as a classical gradient descent by applying a suitable transformation [38], we do not explore this here.

Variants of these dynamics have been proposed to model distributions over networks [20,21,27,39,40]. The constant $\beta \in (0, 2)$ regulates the desired transportation regime. The setting $\beta < 1$ penalizes traffic congestion by distributing paths on more edges, $\beta > 1$ encourages path consolidation into fewer highways, and $\beta = 1$ is shortest pathlike.

In addition to imposing Kirchhoff's law on nodes to ensure mass conservation, solving these dynamics outputs otherwise unconstrained optimal μ_e and F_e (see SM [32]). While this may be enough in ideal cases, in more realistic scenarios it is important to further constrain the solution. For instance, structural constraints may limit the maximum amount of flow that an edge can carry, or a budget constraint may be used to limit the infrastructure cost for building the network. Hence, the dynamics $\dot{\mu} = f$ must be altered to account for these additional constraints.

There are many ways in which constraints can be added. A popular approach is to add constraints on a so-called position level, which leads to gradient inclusions in continuous time [[41], Ch 3.4], and projected gradient descent in discrete time. Unfortunately, the scope of projected gradients is limited, due to the fact that projections can only be efficiently evaluated for constraints that have a particular structure (such as a low-dimensional hyperplane, the probability simplex, or a Euclidean norm ball). When the feasible set is nonconvex and/or fails to have a simple structure, evaluating projections is a computationally daunting task. This motivates our formulation (see also Ref. [42]), which includes constraints on a velocity level and yields a sparse local and linear approximation of the feasible set. As a consequence, the updates for μ can often still be evaluated in closed form (or there is an efficient way of computing them numerically) even though the underlying feasible set is nonconvex or fails to have a simple structure. We will highlight explicit examples of such situations in the remainder of this Letter.

We define $C := \{\mu \in \mathbb{R}_{\geq 0}^E | g(\mu) \geq 0\}$ as the set of feasible conductivities $\mu = (\mu_1, \ldots, \mu_E)$, with g a constraint function that we assume continuously differentiable and E is the number of network edges. Interpreting μ as a "position" variable we can equivalently express the constraints in C in terms of a "velocity" variable by imposing $\dot{\mu}(t) \in V_\alpha(\mu(t))$, where $V_\alpha(\mu(t))$ is the set of feasible velocities and $\alpha \geq 0$ is a constant typically referred to as a "restitution" parameter or "slackness," see Appendix for details.

For $\mu(t) \notin C$ and an active constraint i, the constraint $\dot{\mu}(t) \in V_\alpha(\mu(t))$ is equivalent to $\mathrm{d}g_i(\mu(t))/\mathrm{d}t \geq -\alpha g_i(\mu(t))$, which ensures that potential constraint violations decay at the rate $\alpha > 0$. The situation is visualized in Fig. 1(a).

In order to account for the velocity constraint $\dot{\mu} \in V_\alpha(\mu)$ we augment the dynamics $\dot{\mu} = f$ with a *reaction* force R

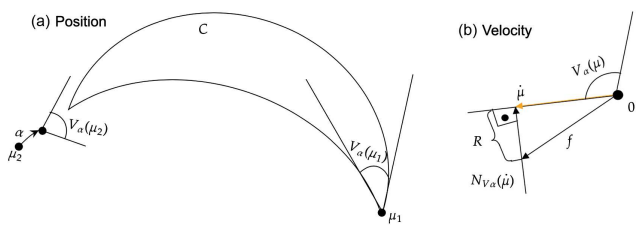


FIG. 1. (a) Visualization of the set C and the set of feasible velocities $V_\alpha(\mu_1)$ and $V_\alpha(\mu_2)$ at points μ_1 and μ_2, respectively. Point μ_1 lies on the boundary of C, while μ_2 is infeasible; α is a restitution parameter. (b) When the vector field f is pushing away from C, a force $-R \in N_{V_\alpha}(\dot{\mu})$ is added to the dynamics to ensure $\dot{\mu} \in V_\alpha(\mu)$.

that forces the solution to remain within the desired constraints:

$$\dot{\mu} = f + R, \quad \text{with} \quad -R \in N_{V_\alpha(\mu)}(\dot{\mu}), \tag{2}$$

where $N_{V_\alpha(\mu)}(\dot{\mu})$ denotes the normal cone of the set $V_\alpha(\mu)$ at $\dot{\mu}$. Because of the scaling of the gradient with S, the normal cone is defined with respect to the inner product $\langle a, b\rangle = a^T S^{-1} b$, where $a, b \in \mathbb{R}^E$ are arbitrary vectors. This has the important effect of guaranteeing that $\mathfrak{L}_\beta$ (of the unconstrained dynamics) is still a Lyapunov function also in the constrained setting and that $\mathfrak{L}_\beta(\mu(t))$ is monotonically decreasing along the trajectories of Eq. (2). A detailed derivation is included in the SM [32].

The addition of R ensures that even if f pushes μ away from C, as shown in Fig. 1(b), the force R, which is orthogonal to the set $V_\alpha(\mu)$, annihilates the component of f that would lead to a constraint violation and ensures that $\dot{\mu} \in V_\alpha(\mu)$. As discussed above, we can therefore conclude that $\mu(0) \in C \Rightarrow \mu(t) \in C$ for all $t \geq 0$ and $\mu(0) \notin C \Rightarrow \mu(t) \to C$ for $t \to \infty$.

In addition, we infer from Fig. 1 that the resulting $\dot{\mu}$ in Eq. (2) is nothing but the projection of f onto the set $V_\alpha(\mu)$ and as a result, we can rewrite $\dot{\mu}$ in the following way:

$$\dot{\mu} := \underset{v \in V_\alpha(\mu)}{\operatorname{argmin}} \frac{1}{2} \langle v - f, v - f\rangle, \tag{3}$$

which can also be equivalently reformulated as the quadratic program (QP)

$$\dot{\mu} := \underset{v \in V_\alpha(\mu)}{\operatorname{argmin}} \frac{1}{2} (v - f)^T S^{-1} (v - f). \tag{4}$$

This reformulation is not only useful for numerical computations, but also highlights that the velocity $\dot{\mu}$ is chosen, at each point in time, to match the unconstrained f. Figure 1(a) visualizes the set C and the set of feasible velocities $V_\alpha(\mu_1)$ and $V_\alpha(\mu_2)$ at points μ_1 and μ_2, respectively. Point μ_1 lies on the boundary of C, while μ_2 is infeasible. We note that the cone $V_\alpha(\mu_2)$ includes an offset, which is controlled by the restitution parameter α; this ensures that any $v \in V_\alpha(\mu_2)$ leads to a decrease in constraint violation. Figure 1(b) shows that when the vector field f is pushing away from C, a force $-R \in N_{V_\alpha}(\dot{\mu})$ is added to the dynamics. The force R annihilates the component of f that would lead to a constraint violation and ensures $\dot{\mu} \in V_\alpha(\mu)$, where $\dot{\mu}$ is chosen as close as possible to f. This can also be interpreted as Gauss's principle of least constraint. It is important to note that $V_\alpha(\mu)$ is a polyhedral set that only includes the constraints I_μ, a subset of the original constraints $g(\mu) \geq 0$. The set $V_\alpha(\mu)$ represents therefore a sparse, local, and linear approximation of the feasible set. The solution $\dot{\mu}$ of Eq. (3) can then be used to update the conductivity with a discrete-time algorithm:

$$\mu^{t+1} = \mu^t + \tau \dot{\mu}, \tag{5}$$

where $\tau > 0$ is the step size.

This general formalism can be applied to a variety of scenarios, provided one can compute ∇g, which determines the set $V_\alpha(\mu)$. We can then solve Eq. (4) by using numerical solvers tailored to the QP, which then yields the update Eq. (5). Additional details about the computational complexity for solving Eq. (5) are described in the SM [32]. However, in important special cases, the optimization Eq. (5) can be solved in closed form, as we illustrate below with three relevant examples.

Capacity constraints—In cases of structural constraints that strictly limit the amount of mass that can travel along any given edge, one can consider capacities $c_e \geq 0$ on edges and set constraints as $g_e(\mu) = c_e - \mu_e$. The velocity constraint $v \in V_\alpha(\mu)$ in Eq. (3) reads as $v_e \leq \alpha g_e(\mu_e)$, for $e \in I_\mu$, which is strictly negative, since $\alpha > 0$ (SM [32]). As previously discussed, $\alpha > 0$ is a restitution parameter that dictates the rate at which constraint violations decay. In discrete time, one should choose $\alpha > 0$ such that $\alpha\tau \leq 1$ to guarantee convergence (see Ref. [42]). We can then solve Eq. (3) in closed form for edges violating the constraint obtaining $v_e = \min\{\alpha(c_e - \mu_e), f_e\}$. In summary, for each edge e, we have

$$\dot{\mu}_e = \begin{cases} \alpha(c_e - \mu_e), & \text{if } f_e \geq \alpha(c_e - \mu_e) \text{ and } \mu_e \geq c_e, \\ f_e & \text{otherwise.} \end{cases} \tag{6}$$

Figure 2 shows the path topologies with capacity constraints on synthetic data, compared against the unconstrained case. We generate random planar networks as the Delaunay triangulation [43] of $N = 300$ points in the plane. We measure the Gini coefficient $\mathrm{Gini}(T)$ calculated on the traffic on edges, defined as the E-dimensional vector T with entries $T_e = \sum_i |F_e^i|/n$, where n is the number of passengers. The coefficient has value in [0, 1] and it determines how traffic is distributed along network edges, with $\mathrm{Gini}(T) = 0$, 1 meaning equally balanced or highly unbalanced traffic on few edges, respectively. The choice of the edge capacity c_e influences this value, with lower c_e imposing stricter constraint and thus encouraging traffic to distribute more equally along the edge, i.e., lower Gini, as shown in Fig. 2(a). Conversely, this implies longer routes for passengers, as measured by an increasing average total path length $\langle l\rangle = \sum_{e,i} \ell_e |F_e^i|/n$ compared to the unconstrained solution, as shown in Fig. 2(b).

Budget constraint—As a second example, we consider a global constraint that involves all the edges at once, a budget constraint $g_b(\mu) = b - \sum_e \mu_e$. This is relevant when a network manager has a fixed limited amount of resources $b > 0$ to invest. We note that, while the Lyapunov function $\mathfrak{L}_\beta$ contains a similar budget term—the cost to build the infrastructure—this cost is not regarded as a constraint in standard approaches [20,26] but as part of the

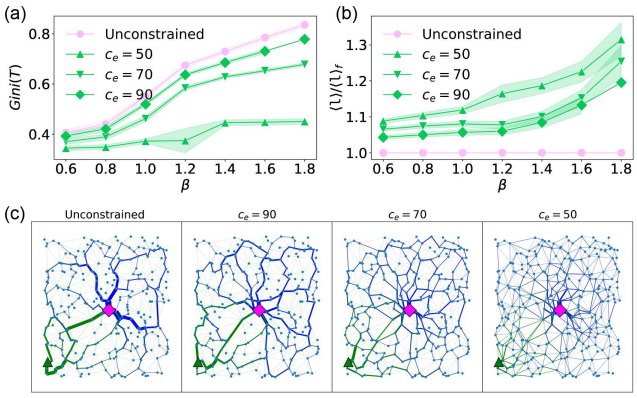


FIG. 2. Capacity constraint on synthetic networks. (a) Gini coefficient of the traffic distribution on edges. The edge capacity $c_e = c$ is selected as a percentile of the distribution of μ over edges obtained in the unconstrained case (Unconstrained). (b) Ratio of average total path length to that of Unconstrained, $\langle \mathtt{l} \rangle_f$. Markers and shadows are averages and standard deviations over 20 network realizations, with 100 randomly selected origins. All passengers have the same central destination (square magenta marker). (c) Example trajectory of one passenger type (green color), whose origin is the green triangle marker. Edge widths are proportional to the amount of passengers traveling through an edge; $\beta = 1.8$.

energy consumption, and the budget b is not a Lagrange multiplier but a measurable constant. Furthermore, unlike the previous case where including a positivity constraint $\mu_e \geq 0$ is optional (but it can in principle be imposed as well, see SM [32]), here we need to include that explicitly. In the standard OT formalism positivity is ensured, provided μ_e is initialized as a positive quantity. Adding constraint may not preserve positivity anymore during the updates, this is the case for the budget constraint, as we observed empirically. Positivity is enforced by adding $g_p(\mu) = \mu \geq 0$, i.e., $\mu_e \geq 0 \ \forall \ e$.

In this budget constraint setting, the conductivities violate the constraint whenever $\sum_e \mu_e > b$. We derive a closed-form solution as $\dot{\mu}_e = f_e - S_e \lambda_b$, if $f_e - S_e \lambda_b \geq -\alpha \mu_e$, and $\dot{\mu}_e = -\alpha \mu_e$ otherwise, where $\lambda_b \in \mathbb{R}$, a Lagrange multiplier for the budget constraint, can be numerically determined via fixed-point iteration (SM [32]).

Combining linear and nonlinear constraints—All the previous examples considered linear constraints, where it is simple to derive analytical solutions. In general, constraints can be more complicated and thus require numerical methods to solve the constrained QP in Eq. (3). In this scenario, we consider a nonlinear budget constraint of the form $g_\delta(\mu) = b - \sum_e \mu_e^\delta \geq 0$, where $\delta > 0$ is a nonlinearity parameter. Setting $\delta = 1$ gives a linear budget constraint as the one discussed earlier. A nonlinear example is a volume-preserving constraint where $\delta = 1/2$, this is relevant for biological processes such as leaf venation and vascular systems [21,44]. This nonlinear budget induces the velocity constraint $\sum_e \delta \mu_e^{\delta-1} v_e \leq \alpha g_\delta(\mu)$. In addition, we also consider a capacity constraint as in the first scenario studied above. Overall, three functions are required: (i) $g_\delta(\mu)$ to impose nonlinear budget constraint; (ii) $g_e(\mu)$ to impose edge capacity, and (iii) $g_p(\mu)$ to ensure positivity. We derive the closed-form solution as

$$\dot{\mu}_e = \begin{cases} \alpha(c_e - \mu_e) & \text{if } f_e - S_e \lambda_\delta h_e \geq \alpha(c_e - \mu_e), \ \mu_e \geq c_e \\ -\alpha \mu_e & \text{if } f_e - S_e \lambda_\delta h_e \leq -\alpha \mu_e, \qquad \mu_e \leq 0 \\ f_e - S_e \lambda_\delta h_e & \text{otherwise,} \end{cases} \tag{7}$$

where $h_e = \delta \mu_e^{\delta-1}$ and $\lambda_\delta > 0$. The value of λ_δ can be determined numerically using fixed-point iteration (SM [32]). The value $\alpha(c_e - \mu_e)$ ensures there is no violation on the edge capacity, $-\alpha \mu_e$ imposes positivity constraint, and $f_e - S_e \lambda_\delta h_e$ captures budget violation. Overall, this scenario ensures that the velocity $\dot{\mu}_e$ has an upper bound of $\alpha(c_e - \mu_e)$ and lower bound of $-\alpha \mu_e$. The choice of δ impacts the topological properties of the resulting network, e.g., the total path length. In the numerical experiments, we set the nonlinearity parameter as $\delta \in (0, 1)$.

Grenoble network—We examine the topology of various constrained solutions on the road network of the city of Grenoble [45], see Fig. 3(a). This has 640 nodes and 740 edges. As a relevant example, we set the central bus station as the destination node and select the remaining 639 nodes as origins, but our method still applies to other choices of origin-destination pairs, e.g., peripheral nodes connecting to other peripheral nodes or to various hubs. This can be specified inside Kirchhoff's law, see SM [32].

Routes generated from the nonlinear constraint scenario balance traffic more than the unconstrained case and result in longer routes, see Figs. 3(b) and 3(c). Adding a budget constraint for $\beta > 1$ results in more distributed traffic (lower Gini) without increasing much the total path length, compared to the unconstrained case. This could be used for instance to allocate to roads infrastructural works aimed at maintenance or upgrade when having a restricted budget.

Discussion—Distributing flows in a transportation network is challenging. Approaches based on optimal transport theory are promising, but they are limited by the lack of a mechanism to incorporate realistic constraints. We show how to impose arbitrary constraints on OT problems in a principled and flexible way. The constraints are lifted from a position to a velocity level and are included in the corresponding mirror descent dynamics. This results in a scalable algorithm that solves constrained OT problems in a computationally efficient manner. The algorithm relies on a sparse local approximation of the feasible set at each iteration. Thus, closed-form updates can often be derived, even if the underlying feasible set is nonconvex or nonlinear. Otherwise, one can resort to efficient numerical methods to solve at most a quadratic program. Our physics-based approach is a change of paradigm with regard to how

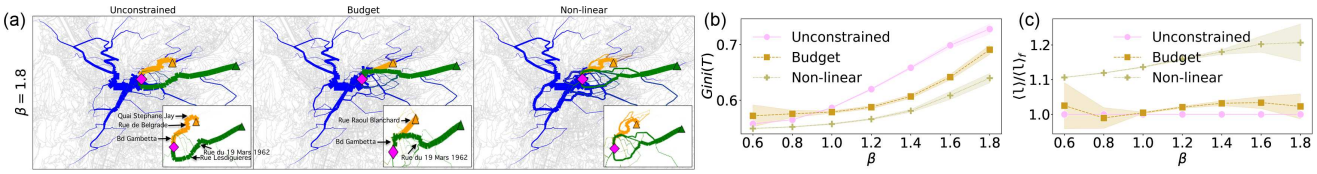


FIG. 3. Constrained OT on Grenoble road network. (a) Path trajectories for the unconstrained OT (Unconstrained), budget constraint (Budget), and a nonlinear budget plus capacity (Nonlinear). We set $b = \frac{1}{2}\sum_e \mu_e$, where μ_e is that of unconstrained, $\delta = 1/2$ and $c_e = 70$ for all edges. Example trajectories of two passenger types (green and orange), whose origin are the respective triangles. All passengers have the same central destination (magenta marker). Edge widths are proportional to the amount of passengers traveling through an edge. (b) Gini coefficient of the traffic distribution on edges. (c) Ratio of average total path length to that of Unconstrained. Markers and shadows are averages and standard deviations over 100 randomly selected destinations, respectively.

OT problems are modeled and solved numerically. This calls for a generalization of transportation problems in wider scenarios, e.g., in networks with multiple transport modes [28], with real-time traffic demands [46], or with noise-induced resonances [47].

Acknowledgments—The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting A. A. I. M. M. thanks the German Research Foundation and the Branco Weiss Fellowship, administered by ETH Zurich, for support.

Data availability—We provide an open source implementation [48].

[1] R. Arnott and K. Small, Am. Sci. **82**, 446 (1994), https://www.jstor.org/stable/29775281.
[2] E. Commission, White paper of 28 march 2011: "roadmap to a single european transport area–towards a competitive and resource efficient transport system" (2011), https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0144:FIN:EN:PDF.
[3] E. Commission, Strategic plan 2020–2024 (2020), https://commission.europa.eu/publications/strategic-plans-2020-2024_en.
[4] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, Network flows. alfred p.sloan school of management (1988), http://hdl.handle.net/1721.1/49424.
[5] R. G. Morris and M. Barthelemy, Phys. Rev. Lett. **109**, 128703 (2012).
[6] L. Gao, P. Shu, M. Tang, W. Wang, and H. Gao, Phys. Rev. E **100**, 012310 (2019).
[7] J. D. Noh and H. Rieger, Phys. Rev. E **66**, 066127 (2002).
[8] R. Dobrin and P. M. Duxbury, Phys. Rev. Lett. **86**, 5076 (2001).
[9] M. Bayati, C. Borgs, A. Braunstein, J. Chayes, A. Ramezanpour, and R. Zecchina, Phys. Rev. Lett. **101**, 037208 (2008).
[10] C. H. Yeung and D. Saad, Phys. Rev. Lett. **108**, 208701 (2012).
[11] C. H. Yeung, D. Saad, and K. M. Wong, Proc. Natl. Acad. Sci. U.S.A. **110**, 13717 (2013).
[12] A. Solé-Ribalta, S. Gómez, and A. Arenas, Phys. Rev. Lett. **116**, 108701 (2016).
[13] J. Gómez-Gardenes and V. Latora, Phys. Rev. E **78**, 065102 (R) (2008).
[14] L. Lacasa, M. Cea, and M. Zanin, Physica A (Amsterdam) **388**, 3948 (2009).
[15] K. Sneppen, A. Trusina, and M. Rosvall, Europhys. Lett. **69**, 853 (2005).
[16] M. Rosvall, A. Grönlund, P. Minnhagen, and K. Sneppen, Phys. Rev. E **72**, 046117 (2005).
[17] L. Zhao, Y.-C. Lai, K. Park, and N. Ye, Phys. Rev. E **71**, 026125 (2005).
[18] E. Estrada, J. Gómez-Gardeñes, and L. Lacasa, Proc. Natl. Acad. Sci. U.S.A. **120**, e2305001120 (2023).
[19] 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).
[20] D. Hu and D. Cai, Phys. Rev. Lett. **111**, 138701 (2013).
[21] H. Ronellenfitsch and E. Katifori, Phys. Rev. Lett. **117**, 138301 (2016).
[22] F. Santambrogio, Optimal Transport for Applied Mathematicians (Birkäuser, NY, 2015), http://ndl.ethernet.edu.et/bitstream/123456789/53568/1/Filippo%20Santambrogio.pdf.
[23] D. Baptista, D. Leite, E. Facca, M. Putti, and C. De Bacco, Sci. Rep. **10**, 20806 (2020).
[24] D. Leite and C. De Bacco, arXiv:2209.06751.
[25] V. Bonifaci, K. Mehlhorn, and G. Varma, J. Theor. Biol. **309**, 121 (2012).
[26] A. Lonardi, E. Facca, M. Putti, and C. De Bacco, Phys. Rev. Res. **3**, 043010 (2021).
[27] S. Bohn and M. O. Magnasco, Phys. Rev. Lett. **98**, 088702 (2007).
[28] A. A. Ibrahim, A. Lonardi, and C. De Bacco, Algorithms **14**, 189 (2021).
[29] A. Lonardi, D. Baptista, and C. De Bacco, Front. Phys. **11**, 65 (2023).
[30] C. Lanczos, *The Variational Principles of Mechanics* (University of Toronto Press, Toronto, 1949).
[31] 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).
[32] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.133.057401, which includes Refs. [27–31], for additional information about the derivations and a detailed discussion of the numerical simulations.

[33] A. A. Ibrahim, D. Leite, and C. De Bacco, Phys. Rev. E **105**, 064302 (2022).
[34] T. A. Davis, ACM Trans. Math. Softw. **30**, 196 (2004).
[35] W. L. Briggs, V. E. Henson, and S. F. McCormick, *A Multigrid Tutorial* (SIAM, Philadelphia, 2000).
[36] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course* (Springer Science & Business Media LLC, New York, 2004).
[37] V. Bonifaci, Comput. Optim. Applic. **79**, 441 (2021).
[38] E. Facca and M. Benzi, SIAM J. Sci. Comput. **43**, A2295 (2021).
[39] E. Katifori, G. J. Szöllősi, and M. O. Magnasco, Phys. Rev. Lett. **104**, 048704 (2010).
[40] J. R. Banavar, F. Colaiori, A. Flammini, A. Maritan, and A. Rinaldo, Phys. Rev. Lett. **84**, 4745 (2000).
[41] J. P. Aubin and A. Cellina, *Differential Inclusions* (Springer, New York, 1984).
[42] M. Muehlebach and M. I. Jordan, J. Mach. Learn. Res. **23**, 1 (2022), http://jmlr.org/papers/v23/21-0798.html.
[43] L. Guibas and J. Stolfi, ACM Trans. Graph. **4**, 74 (1985).
[44] A. Takamatsu, T. Gomi, T. Endo, T. Hirai, and T. Sasaki, J. Phys. D **50**, 154003 (2017).
[45] R. Kujala, C. Weckström, R. K. Darst, M. N. Mladenović, and J. Saramäki, Sci. Data **5**, 180089 (2018).
[46] A. Lonardi, E. Facca, M. Putti, and C. De Bacco, Phys. Rev. E **107**, 024302 (2023).
[47] F. Folz, K. Mehlhorn, and G. Morigi, Phys. Rev. Lett. **130**, 267401 (2023).
[48] Multicommodity Optimal Transport: McOpt, https://github.com/aleable/McOpt (2023).
[49] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis* (Springer Verlag, Berlin-Heidelberg, 1998).
[50] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. (Springer, New York, 2006).
[51] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. (Athena Scientific, 1999).
[52] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*, 4th ed. (Springer, New York, 2016).

End Matter

Appendix A: Details about setting the constraints—We define $C := \{\mu \in \mathbb{R}^E_{\geq 0} | g(\mu) \geq 0\}$ as the set of feasible conductivities $\mu = (\mu_1, \ldots, \mu_E)$, with g a constraint function that we assume continuously differentiable and E is the number of network edges. We focus on those edges where constraints are not satisfied, and denote the set of active constraints for a given μ as $I_\mu := \{i \in \mathbb{Z} | g_i(\mu) \leq 0\}$. Interpreting μ as a "position" variable, a constraint to ensure $\mu(t) \in C$, $\forall\ t \geq 0$, can be equivalently formulated as a constraint on its velocity $\dot{\mu}(t) \in T_C(\mu(t))$, $\forall\ t \geq 0$, with $\mu(0) \in C$, where $T_C(\mu)$ denotes the tangent cone of the feasible set at μ, see Ref. [49]. However, it will be convenient to slightly extend the notion of tangent cone to also account for infeasible initial conditions (this is particularly important for the discretization), which is achieved by imposing $\dot{\mu}(t) \in V_\alpha(\mu(t))$, where $V_\alpha(\mu) := \{v \in \mathbb{R}^E | \nabla g_i(\mu)^T v \geq -\alpha g_i(\mu),\ i \in I_\mu\}$, and $\alpha \geq 0$ is a constant typically referred to as a "restitution" parameter or "slackness." We note that $V_\alpha(\mu)$ generalizes the notion of the tangent cone, since for $\mu \in C$, $V_\alpha(\mu) = T_C(\mu)$. We assume mild regularity conditions (constraint qualification). A sufficient condition is, for example, the existence of $v \in \mathbb{R}^E$ such that $\nabla g_i(\mu)^T v > 0$ for all $i \in I_\mu$.

For $\mu(t) \notin C$, the constraint $\dot{\mu}(t) \in V_\alpha(\mu(t))$ is equivalent to $\mathrm{d}g_i(\mu(t))/\mathrm{d}t \geq -\alpha g_i(\mu(t))$, $i \in I_{\mu(t)}$, which ensures that potential constraint violations decay at the rate $\alpha > 0$.

Appendix B: Details about our method—From a variational optimization perspective, our approach is related to successive linear and sequential quadratic programming [50–52]. The underlying idea of these methods is to linearize the objective function and the constraints about the current iterate and to solve a local linear and/or quadratic program. Our Letter improves upon these ideas and tailors these to optimal transport problems in the following way: (i) we linearize a subset of constraints at every iteration, which means that the subproblem Equation (3) typically includes very few constraints and can be solved efficiently; (ii) we introduce a non-Euclidean inner product that is adapted to optimal transport problems and is used to show that $\mathfrak{L}_\beta$ is a Lyapunov function; (iii) we provide closed-form updates in various problem instances that are practically relevant.

Optimal transport with constraints: from mirror descent to classical mechanics
Supporting Material (SM)

Abdullahi Adinoyi Ibrahim,* Michael Muehlebach,† and Caterina De Bacco‡
Max Planck Institute for Intelligent Systems, Cyber Valley, Tübingen 72076, Germany
(Dated: June 11, 2024)

DETAILED MATHEMATICAL DERIVATIONS OF OPTIMAL TRANSPORT WITH CONSTRAINTS

Here we present in more details the mathematical derivations of the results presented in the main text. Specifically, we consider the three examples of constraints described in the main manuscript: capacity on edges, budget, and a third constraint that combines a linear capacity constraint and a non-linear budget constraint.

In the following we denote as E the set of network edges and $E = |E|$ is the number of edges.

Capacity constraint

The first case considered is that of a local and linear constraint where we assign a capacity on individual edges such that conductivities cannot be larger than the prescribed capacity. This is relevant in situations where structural constraints prevent a large amount of mass to travel on individual network edges without compromising the infrastructure. Mathematically, we define for each $e \in E$ the constraint as:

$$g_{c_e}(\mu) = c_e - \mu_e \quad , \tag{S1}$$

where c_e is the capacity imposed on edge e. This is a parameter that a user can enter as input and can be different for each edge. In the numerical experiments in the main manuscript we assume c_e to be equal for all edges for simplicity, but the theory here is not impacted by this choice.
By considering a vectorial representation of the constraint where $g_c(\mu) \in \mathbb{R}^E$ is the vector with entries $g_{c_e}(\mu)$, this definition also implies that we have a constant derivative $\nabla_{\mu_e} g_{c_e}(\mu) = -1 < 0$. The constraint $v \in V_\alpha(\mu)$ required to solve the minimization in Eq. (3) implies:

$$\nabla g_c(\mu)^T v \geq -\alpha g_c(\mu) \implies -v \geq -\alpha(c-\mu) \quad \implies \quad v \leq \alpha(c-\mu) \quad . \tag{S2}$$

Solving the quadratic program minimization is simple in this case. For an edge that violates the constraint there are two possibilities: either i) $\mu_e^{\beta-2}|F_e|^2 - \mu_e \geq \alpha(c_e - \mu_e)$ or ii) $\mu_e^{\beta-2}|F_e|^2 - \mu_e \leq \alpha(c_e - \mu_e)$. In case i) we obtain that $v_e = \alpha\,(c_e - \mu_e)$; while in ii) we have $v_e = (\mu_e^{\beta-2}|F_e|^2 - \mu_e) = f_e$. However, case i) results in a reduction of the constraint violation, as we have $\mu_e^{(t+1)} = \mu_e^{(t)} + \tau\, v_e = \mu_e^{(t)} + \tau\,\alpha\,(c_e - \mu_e)$, where $\tau > 0$ is the algorithmic step size. Hence, $\mu_e^{(t+1)} - c_e \leq (1-\alpha\tau)\,(\mu_e^{(t)} - c_e)$, which means that the constraint violation $\mu_e^{(t)} - c_e$ decreases at the exponential rate $\alpha > 0$. Thus, α controls how quickly the constraint violations decay. It controls the trade-off between reducing the objective function (encouraged by small α) and converging to the feasible set (encouraged by larger α) [1]. In discrete time, $\alpha\tau$ should be chosen so that $0 < \alpha\tau < 1$ to guarantee convergence. Hence, solving the quadratic program for the setting of capacity constraints gives $v = \min\{\alpha\, g_c(\mu), f\}$. In summary, for e such that $\mu_e \geq c_e$ (constraint violated), we have:

$$\dot{\mu}_e = \begin{cases} \alpha\,(c_e - \mu_e) & \text{if} \quad f_e \geq \alpha\,(c_e - \mu_e) \\ f_e & \text{if} \quad f_e < \alpha\,(c_e - \mu_e) \end{cases} \quad . \tag{S3}$$

The algorithmic update is then $\mu_e^{(t+1)} = \mu_e^{(t)} + \alpha\,\tau\,\dot{\mu}_e$, with $0 \leq \alpha\,\tau \leq 1$ and $\dot{\mu}_e$ as in Eq. (S3).

Note that in the analytical result of Eq. (S3) we did not impose any additional positivity constraint $\mu_e \geq 0$. This was not necessary in our empirical results, as we never found it violated, provided one initializes $\mu_e \geq 0$ at the

* AAI : abdullahi.ibrahim@tuebingen.mpg.de
† MM : michaelm@tuebingen.mpg.de
‡ CDB : caterina.debacco@tuebingen.mpg.de

first iteration. We will show the importance of this additional constraint in subsequent sections when considering constraints other than the capacity. To impose a positivity constraint, we need to enforce an additional constraint of the form $g_p(\mu) = \mu \geq 0$. In the velocity space, this translates to $v \geq -\alpha\,\mu$. Element-wise, the solution will be of the form $v_e = \max\{f_e, -\alpha\,\mu_e\}$, $\forall e \in E$ such that $\mu_e \leq 0$. The analytical solution in addition to positivity constraint is summarized as:

$$\dot{\mu}_e = \begin{cases} \alpha\,(c_e - \mu_e) & \text{if} \quad f_e \geq \alpha\,(c_e - \mu_e) \\ \\ -\alpha\,\mu_e & \text{otherwise} \end{cases} \tag{S4}$$

for all $e \in E$ such that $\mu_e \leq 0$.

Budget constraint

Here we illustrate our formalism to fix the global network budget b. Formally, we have:

$$g_b(\mu) = b - \sum_{e \in E} \mu_e \quad . \tag{S5}$$

In words, the conductivities $\mu = (\mu_1, \ldots, \mu_E)$ violate the constraint whenever their sum is greater than the input budget $b > 0$. As this involves all the conductivities at once, we need to solve Eq. (3) in vectorial form, i.e., for an input array $v = (v_1, \ldots, v_E)$ of dimension E. We also have $\nabla g(\mu) = (\partial g/\partial \mu_1, \ldots, \partial g/\partial \mu_E)$, hence

$$\nabla g_b(\mu)^T v = \sum_{e \in E} \frac{\partial g_b(\mu)}{\partial \mu_e} v_e = \sum_{e \in E} (-1 \cdot v_e) = -\sum_{e \in E} v_e \geq -\alpha\, g_b(\mu) \implies \sum_{e \in E} v_e \leq \alpha\, g_b(\mu) \quad . \tag{S6}$$

This means that some v_e are allowed to be positive, as long as their contribution is compensated by other negative ones, such that their overall sum is lower than $\alpha\, g_b(\mu)$. Notice that beyond this budget constraint we need to guarantee the fundamental constraint that conductivities have to be positive quantities. Formally, this can be enforced by adding the following additional constraint:

$$g_p(\mu) = \mu \geq 0 \quad . \tag{S7}$$

In the velocity domain this translates into $\nabla g_p(\mu)^T v = v \geq -\alpha\,\mu$; element-wise, this means $v_e \geq -\alpha\,\mu_e$, $\forall e \in E$ such that $\mu_e \leq 0$.

To derive the closed-form solution in this budget constraint case, we thus minimize

$$\arg\min_{v_e} \left\{ \frac{1}{2} \sum_{e \in E} S_e^{-1} (v_e - f_e)^2 \right\} \quad , \tag{S8}$$

subject to the following two constraints:

$$\sum_{e \in E} v_e \leq \alpha \left(b - \sum_{e \in E} \mu_e \right) \quad , \quad \text{if} \quad b \leq \sum_{e \in E} \mu_e \tag{S9}$$

$$v_e \geq -\alpha\,\mu_e, \quad \forall e \in E \text{ such that } \mu_e \leq 0 \quad . \tag{S10}$$

To derive the closed-form solution in this case, we can add a Lagrange multiplier for the budget constraint and solve an auxiliary constraint minimization problem with a modified cost function defined as:

$$L(v, \lambda_b) = \frac{1}{2} \sum_{e \in E} S_e^{-1} (v_e - f_e)^2 + \lambda_b \left(\sum_{e \in E} v_e - \alpha \left(b - \sum_{e \in E} \mu_e \right) \right) \quad . \tag{S11}$$

We then want to solve:

$$\min_{v:\, v_e \geq \alpha\,\mu_e,\, \forall e \in E: \mu_e \leq 0} \; \max_{\lambda_b \geq 0} L(v, \lambda_b) \quad . \tag{S12}$$

Defining $\vec{1}_E = (1, \ldots, 1)$ the E-dimensional vector with entries all equal to 1 and using a vectorial representation (where $\vec{1}_E^T v = \sum_{e \in E} v_e$), this is equivalent to solve:

$$\arg\min_{v:\, v_e \geq \alpha\,\mu_e,\, \forall e \in E: \mu_e \leq 0} \left\{ \frac{1}{2} |S^{-1/2}(v - f)|^2 + \lambda_b \vec{1}_E^T v - \alpha\,\lambda_b\, g_b(\mu) \right\} \quad , \tag{S13}$$

where λ_b denotes the optimal multiplier. Equivalently, the above problem can be reformulated as

$$\underset{v:\, v_e \geq \alpha\, \mu_e,\, \forall e \in E: \mu_e \leq 0}{\arg\min} \left\{ \frac{1}{2} |S^{-1/2}(v - f + S\lambda_b \vec{1}_E)|^2 \right\} = \underset{v:\, v_e \geq \alpha\, \mu_e,\, \forall e \in E: \mu_e \leq 0}{\arg\min} \left\{ \frac{1}{2} |v - f + S\lambda_b \vec{1}_E|^2 \right\} \quad . \tag{S14}$$

For an edge such that $\mu_e \leq 0$, this has the following closed-form solution:

$$\dot{\mu}_e = \begin{cases} f_e - S_e \lambda_b & \text{if} \quad f_e - S_e \lambda_b \geq -\alpha\, \mu_e \\ \\ -\alpha\, \mu_e & \text{if} \quad f_e - S_e \lambda_b < -\alpha\, \mu_e \end{cases} \quad . \tag{S15}$$

Estimation of λ_b

Imposing the budget constraint, and defining $I_p(\lambda_b) = \{e \in E \,|\, f_e - S_e \lambda_b < -\alpha\, \mu_e \text{ and } \mu_e \leq 0\}$ as the set of edges that violate the positivity constraint (both in position and velocity), we obtain that the condition $\sum_{e \in E \setminus I_p(\lambda_b)} (f_e - S_e \lambda_b) - \alpha \sum_{e \in I_p(\lambda_b)} \mu_e \leq \alpha\, g_b(\mu)$ must be satisfied to ensure Eq. (S9). This inequality determines the value λ_b. In addition we can make the following case distinction (complementary slackness) $\lambda_b = 0 \iff \sum_{e \in E \setminus I_p(0)} f_e - \alpha \sum_{e \in I_p(0)} \mu_e \leq \alpha g_b(\mu)$ and $\lambda_b > 0 \iff \sum_{e \in E \setminus I_p(\lambda_b)} (f_e - S_e \lambda_b) - \alpha \sum_{e \in I_p(\lambda_b)} \mu_e = \alpha g_b(\mu)$. In the former case the solution v to Eq. (S8) is given by Eq. (S15) with $\lambda_b = 0$. In the latter case we compute λ_b with a fixed-point method and define:

$$k_b(\lambda_b) = \frac{\sum_{e \in E \setminus I_p(\lambda_b)} f_e - \alpha \sum_{e \in I_p(\lambda_b)} \mu_e - \alpha\, (b - \sum_{e \in E} \mu_e)}{\sum_{e \in E \setminus I_p(\lambda_b)} S_e} \quad . \tag{S16}$$

The multiplier λ_b is then computed as $\lambda_b^{(a+1)} = k_b(\lambda_b^{(a)})$, where initial value of $\lambda_b^{(0)}$ can be chosen for instance as $\lambda_b^{(0)} = \min_{e: \mu_e \leq 0 \text{ and } f_e + \alpha\, \mu_e \geq 0} \{f_e + \alpha\, \mu_e\}$.

Combination of linear and non-linear constraints

We now consider a more complex scenario where we combine the capacity constraint with a non-linear generalization of the budget constraint. Specifically, we consider three functions for the constraints, a local capacity constraint $g_c(\mu) : \mathbb{R}^E \to \mathbb{R}^E$, a local positivity constraint $g_p(\mu) : \mathbb{R}^E \to \mathbb{R}^E$ and a global budget constraint $g_\delta(\mu) : \mathbb{R}^E \to \mathbb{R}^1$. These functions are defined as:

$$g_e(\mu) = c - \mu \tag{S17}$$

$$g_p(\mu) = \mu \tag{S18}$$

$$g_\delta(\mu) = b - \sum_{e \in E} \mu_e^\delta \quad , \tag{S19}$$

where $b > 0$ and $\delta > 0$ are a budget and a scaling parameter, respectively. We recover the linear budget constraint for $\delta = 1$.

The constraint on v that result from the capacity constraint in Eq. (S17) required to solve Eq. (3) have been derived in Section "Capacity constraint". The function $g_p(\mu)$ imposes the positivity constraint, which means that each individual edge has to have $\mu_e \geq 0$. The constraint $g_p(\mu)$ induces the following velocity constraint

$$\nabla g_{p_e}(\mu)^T\, v_e \geq -\alpha\, \mu_e \implies v_e \geq -\alpha \mu_e \quad , \tag{S20}$$

for all $e \in E$ such that $\mu_e \leq 0$.

Similarly, we solve the non-linear budget constraint as follows

$$\nabla g_\delta(\mu)^T\, v \geq -\alpha\, (b - \sum_{e \in E} \mu_e^\delta) \implies \sum_{e \in E} \delta \mu_e^{\delta - 1} v_e \leq \alpha\, (b - \sum_{e \in E} \mu_e^\delta) \quad , \tag{S21}$$

as long as $b \leq \sum_{e \in E} \mu_e^\delta$. To derive the closed-form solution in this case, we minimize

$$\min_{v_e} \left\{ \frac{1}{2} \sum_{e \in E} S_e^{-1} (v_e - f_e)^2 \right\} \quad , \tag{S22}$$

subject to the following three constraints:

$$\sum_{e\in E} \delta \mu_e^{\delta-1} v_e \leq \alpha \left(b - \sum_{e\in E} \mu_e^{\delta} \right) \quad , \quad \text{if} \quad b \leq \sum_{e\in E} \mu_e^{\delta} \tag{S23}$$

$$v_e \geq -\alpha\, \mu_e, \quad \forall e \in E \text{ such that } \mu_e \leq 0 \quad , \tag{S24}$$

$$v_e \leq \alpha\, (c_e - \mu_e), \; \forall e \in E \text{ such that } \mu_e \geq c_e \quad . \tag{S25}$$

To derive the closed-form solution in this case, we can add a Lagrange multiplier for the non-linear constraint and solve an auxiliary constraint minimization problem with a modified cost function defined as:

$$L_n(v, \lambda_\delta) = \frac{1}{2} \sum_{e\in E} S_e^{-1} (v_e - f_e)^2 + \lambda_\delta \left(\sum_{e\in E} \delta \mu_e^{\delta-1} v_e - \alpha \left(b - \sum_{e\in E} \mu_e^{\delta} \right) \right) \quad , \tag{S26}$$

where $\lambda_\delta \geq 0$. We then want to solve:

$$\min_{\substack{v: \\ v_e \leq \alpha\,(c_e - \mu_e),\, \forall e\in E:\, \mu_e \geq c_e \\ v_e \geq -\alpha \mu_e,\, \forall e \in E: \mu_e \leq 0}} \; \max_{\lambda_\delta \geq 0} L_n(v, \lambda_\delta) \quad . \tag{S27}$$

Defining $h = \delta\, (\mu_1^{\delta-1}, \ldots, \mu_E^{\delta-1})$ and using a vectorial representation (where $h^T v = \sum_{e\in E} \delta \mu_e^{\delta-1} v_e$), this is equivalent to solving

$$\arg\min_{\substack{v: \\ v_e \leq \alpha\,(c_e - \mu_e),\, \forall e\in E:\, \mu_e \geq c_e \\ v_e \geq -\alpha \mu_e,\, \forall e \in E: \mu_e \leq 0}} \left\{ \frac{1}{2} |S^{-1/2}(v - f)|^2 + \lambda_\delta h^T v - \alpha\, \lambda_\delta\, g_\delta(\mu) \right\} \quad , \tag{S28}$$

where $\lambda_\delta \geq 0$ denotes the optimal Lagrange multiplier. Equivalently, by completing the square and ignoring terms that do not depend on v, the above problem can be re-written as

$$\arg\min_{\substack{v: \\ v_e \leq \alpha\,(c_e - \mu_e),\, \forall e\in E:\, \mu_e \geq c_e \\ v_e \geq -\alpha \mu_e,\, \forall e \in E: \mu_e \leq 0}} \left\{ \frac{1}{2} |S^{-1/2}(v - f + S\lambda_\delta h)|^2 \right\} = \arg\min_{\substack{v: \\ v_e \leq \alpha\,(c_e - \mu_e),\, \forall e\in E:\, \mu_e \geq c_e \\ v_e \geq -\alpha \mu_e,\, \forall e \in E: \mu_e \leq 0}} \left\{ \frac{1}{2} |v - f + S\lambda_\delta h|^2 \right\} \quad . \tag{S29}$$

The analytical solution to Eq. (S29) is given by

$$\dot{\mu}_e = \begin{cases} \alpha\,(c_e - \mu_e) & \text{if } \; f_e - S_e \lambda_\delta\, h_e \geq \alpha\,(c_e - \mu_e)\;, \quad c_e \leq \mu_e \quad , \\ -\alpha\, \mu_e & \text{if } \; f_e - S_e \lambda_\delta\, h_e < -\alpha\, \mu_e, \quad \mu_e \leq 0 \quad , \\ f_e - S_e \lambda_\delta\, h_e & \text{otherwise} \quad . \end{cases} \tag{S30}$$

The analytical solution $\dot{\mu}_e$ is bounded, and a typical plot of μ_e^t with respect to f_e^t is shown in Fig. S1. We note that the value of μ_e^t is also dependent on δ and b, which determine λ_δ as discussed in the next section.

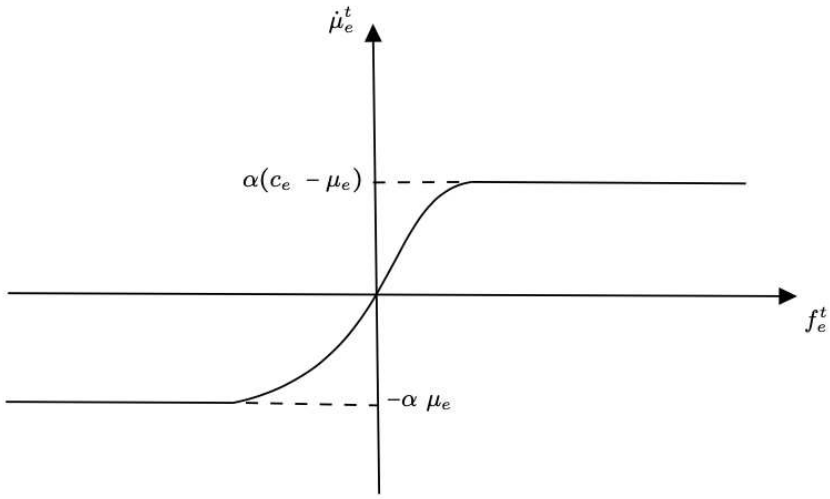


FIG. S1. This plot shows $\dot{\mu}_e^t$ as a function of f_e^t (typical situation, the function also depends on δ and b). The solution is expected to move at most by $\alpha\,(c_e - \mu_e)$ and at least by $\alpha\,\mu_e$.

Computation of λ_δ

By imposing the non-linear budget constraint, and defining $I_p(\lambda_\delta) = \{e \in E \mid f_e - S_e\lambda_\delta\, h_e < -\alpha\,\mu_e, \text{ and } \mu_e \le 0\}$ the set of edges that violate the positivity constraint, and $I_c(\lambda_\delta) = \{e \in E \mid f_e - S_e\,\lambda_\delta\, h_e \ge \alpha\,(c_e - \mu_e),\, c_e \le \mu_e\}$ the set of edges that violate the capacity constraint. We obtain that $\sum_{e\in E\setminus I_{pc}(\lambda_\delta)}(f_e - S_e\lambda_\delta) - \alpha\sum_{e\in I_c(\lambda_\delta)}(c_e - \mu_e) - \alpha\sum_{e\in I_p(\lambda_\delta)}\mu_e \le \alpha\, g_\delta(\mu)$, must be satisfied for the minimizer in Eq. (S22), where $I_{pc}(\lambda_\delta) = I_p(\lambda_\delta) \cap I_c(\lambda_\delta)$. We again make a case distinction on $\lambda_\delta \ge 0$. If $\sum_{e\in E\setminus I_{pc}(0)} f_e - \alpha\sum_{e\in I_c(0)}(c_e - \mu_e) - \alpha\sum_{e\in I_p(0)}\mu_e \le \alpha g_\delta(\mu)$ holds, $\lambda_\delta = 0$. Otherwise $\lambda_\delta > 0$ and $\sum_{e\in E\setminus I_{pc}(\lambda_\delta)}(f_e - S_e\lambda_\delta) - \alpha\sum_{e\in I_c(\lambda_\delta)}(c_e - \mu_e) - \alpha\sum_{e\in I_p(\lambda_\delta)}\mu_e = \alpha\, g_\delta(\mu)$, which is solved by fixed-point iteration. To that extent we introduce

$$k_\delta(\lambda_\delta) = \frac{\sum_{e\in E\setminus I_{pc}(\lambda_\delta)} f_e - \alpha\sum_{e\in I_c(\lambda_\delta)}(c_e - \mu_e) - \alpha\sum_{e\in I_p(\lambda_\delta)}\mu_e - \alpha\left(b - \sum_{e\in E}\mu_e^\delta\right)}{\sum_{e\in E\setminus I_{pc}(\lambda_\delta)} S_e} \quad , \tag{S31}$$

and iterate λ_δ as follows $\lambda_\delta^{(a+1)} = k_\delta(\lambda_\delta^{(a)})$ until convergence.
The guess for an initial value is $\lambda^{(0)} - \delta = \min_{e:\mu_e\le 0 \text{ and } f_e + \alpha\,\mu_e \ge 0}\{f_e + \alpha\,\mu_e\}$.

CONSTRAINED OT ADMITS LYAPUNOV FUNCTION

This section shows that the Lyapunov function of the unconstrained case is still valid when adding the auxiliary force R that imposes the constraints.
The Lyapunov function for the dynamics $\dot{\mu}$ in the unconstrained case is the one given in [2]:

$$\mathfrak{L}_\beta = \frac{1}{2}\sum_{j\in M}\sum_{v\in N} p_v^j(\mu)\, q_v^j + \frac{1}{2(2-\beta)}\sum_{e\in E}\ell_e\mu_e^{2-\beta} \quad , \tag{S32}$$

where N denote the set of nodes and $q^j \in N$ denote the inflow-outflow rate of each passenger type j such that $\sum_v q_v^j = 0$.
To prove that the Lyapunov function is well-defined, we show the following expressions i) $\mathfrak{L}_\beta \ge 0$, ii) $\dot{\mathfrak{L}}_\beta \le 0$ and iii) $\dot{\mathfrak{L}}_\beta = 0$ if and only if μ is a stationary point for the dynamics.
The first (energy dissipation) and second (transport cost) terms of Eq. (S32) are non-negative, hence $\mathfrak{L}_\beta$ satisfies the inequality $\mathfrak{L}_\beta \ge 0$.

Now we prove claim ii), i.e. $\dot{\mathfrak{L}}_\beta \leq 0$. First, notice that

$$\dot{\mathfrak{L}}_\beta = \nabla \mathfrak{L}_\beta^T \, \dot{\mu} \tag{S33}$$
$$= -\langle f, \dot{\mu} \rangle \tag{S34}$$
$$= -\langle \dot{\mu}, \dot{\mu} \rangle + \langle R, \dot{\mu} \rangle \quad , \tag{S35}$$

where in Eq. (S34) we used $\partial \mathfrak{L}_\beta / \partial \mu_e = -\frac{\ell_e}{2\,\mu_e^\beta} f_e$ [2] and in Eq. (S35) we used $\dot{\mu} = f + R$.
The inequality $-\langle \dot{\mu}, \dot{\mu} \rangle = -\dot{\mu}^T S^{-1} \dot{\mu} \leq 0$ is valid because it results in a non-positive sum of squares. Thus, the remaining task is to prove that $\langle R, \dot{\mu} \rangle \leq 0$. The stationarity condition of Eq. (3) can be expressed as:

$$\left(S^{-1}(\dot{\mu} - f)\right)^T (v - \dot{\mu}) \geq 0 \,, \quad \forall v \in V_\alpha(\mu) \quad , \tag{S36}$$

where the first factor is the gradient of the cost in Eq. (3) with respect to v, the second factor is the variation of v and the positivity is due to $\dot{\mu}$ being the minimizer. Using $\dot{\mu} = f + R$, we get $\langle R, v - \dot{\mu} \rangle \geq 0$ for all $v \in V_\alpha(\mu)$.
Now, if $\mu \in C$ (μ is feasible), then $V_\alpha(\mu)$ is a (convex) cone and therefore $0 \in V_\alpha(\mu)$. Hence, we can choose $v = 0$ in the previous expression, yielding $\langle R, -\dot{\mu} \rangle \geq 0 \implies \langle R, \dot{\mu} \rangle \leq 0$. Hence $\dot{\mathfrak{L}}_\beta \leq 0$.

To prove claim iii), we have established $\langle R, \dot{\mu} \rangle \leq 0$ and assuming that $\mu(t) > 0$, we deduce the following:

$$\dot{\mathfrak{L}}_\beta = 0 \iff \langle \dot{\mu}, \dot{\mu} \rangle = 0, \; \langle R, \dot{\mu} \rangle = 0 \quad , \tag{S37}$$
$$\iff \dot{\mu} = 0 \quad , \tag{S38}$$

where $\dot{\mu} = 0$ means $0 = -S \nabla \mathfrak{L}_\beta + R$, with $-R \in N_{V_\alpha(\mu)}(0)$. Additionally, we have established $\mu(t) \in C$, $\forall t$ and therefore $V_\alpha(\mu) = T_C(\mu)$. As a result:

$$\dot{\mu} = 0 \iff -S \nabla \mathfrak{L}_\beta \in N_{T_C(\mu)}(0) \quad , \tag{S39}$$
$$\iff \langle -S \nabla \mathfrak{L}_\beta, v \rangle \leq 0, \quad \forall v \in T_C(\mu) \quad , \tag{S40}$$
$$\iff \nabla \mathfrak{L}_\beta(\mu)^T v \geq 0, \quad \forall v \in T_C(\mu) \quad . \tag{S41}$$

This means that μ corresponds to a stationary point, specifically a local minimum. (Note that we have used the simplifying assumption that $\mu > 0$ in the above argument.)

ALGORITHMIC IMPLEMENTATION

This section presents the algorithmic implementation of the constrained OT method in Eq. (3).
We denote as I the set of indices denoting the constraints, so that each constraint function is written as $g_i(\mu)$, with $i \in I$. Furthermore, Kirchhoff's law is defined as:

$$\sum_{e \in E} B_{ve} F_e^j = \sum_{e \in E} B_{ve} \frac{\mu_e \left(p_u^j - p_v^j\right)}{\ell_e} = q_v^j \,, \quad j = 1, \ldots, M \,; e = (u, v) \quad , \tag{S42}$$

where p_u^j is the pressure potential of type j on node u; B_{ve} is the network incidence matrix with entries $B_{ve} = 1, -1, 0$ if node v is the start, end of edge e, or neither of these, respectively; and q_v^j is the amount of mass of type j entering or exiting at node v. The quantities q_v^j are fixed and given by the problem formulation. Similarly, both the lengths $\{\ell_e\}_{e \in E}$ and the matrix B are fixed as determined by the input graph $G(V, E)$. Hence, the quantities that need to be evaluated are the conductivities $\{\mu_e\}_{e \in E}$ and the pressure potentials $\left\{p_u^j\right\}_{u \in V}$, $\forall j = 1, \ldots, M$, which in turn determine the fluxes vectors $\{F_e\}_{e \in E}$. Given the μ, we can obtain the pressure potentials as $p_v^j = \sum_u (L^j)^\dagger_{vu} q_u^j$, where L^j are elements of the μ^j/ℓ-weighted graph Laplacian, and $\dagger$ denotes the Moore-Penrose inverse.

The algorithmic implementation of the constrained OT method is described in Algorithm 1.

Algorithm 1 Constrained OT Method

Input: Graph $G(V,E)$, M, β, $\alpha\tau \in (0,1]$, g, $\{q_v^j\}_{v\in V, j=1,\dots,M}$
Initialize: $t \leftarrow 0$, μ^t (e.g. sampling as i.i.d. $\mu_e^t \sim \mathrm{Unif}(0,1)$)
while convergence not achieved **do**
  $\{p_u^j\}_{u\in V, j=1,\dots,M} \leftarrow$ solve Kirchhoff's law (Eq. (S42)) for $\{p_u^j\}$ given μ^t and $\{q_v^j\}_{v\in V, j=1,\dots,M}$
  $f^t \leftarrow$ compute the gradients
  $I_{\mu^t} \leftarrow \{\}, w^t \leftarrow \{\}$
  for i in I **do**
    if $g_i(\mu) \le 0$ **then**
      $I_{\mu^t} \leftarrow I_{\mu^t} \cup \{i\}$
      $w^t \leftarrow w^t \cup \{\nabla g_i(\mu^t)\}$
    end if
  end for
  $\dot{\mu}^t \leftarrow$ solve the dynamics in Eq. (4) as follows ▷ (Note: The objective and constraints are of (max) size $|E|$).

$$\dot{\mu}^t \leftarrow \text{SOLVE}\left\{\begin{array}{c} \arg\min_{v\in V_\alpha(\mu^t)} \left\{\frac{1}{2}\left|S^{-1/2}(v-f)\right|^2\right\} \\ \text{subject to}: \\ w_t^T v \ge -\alpha\{g_i(\mu^t)\}_{i\in I_{\mu^t}} \end{array}\right\}.$$

  Update the dynamics: $\mu^{t+1} \leftarrow \mu^t + \tau\dot{\mu}^t, \quad t \leftarrow t+1$
end while
Output: Fluxes $\{F_e\}_{e\in E}$ and conductivities $\{\mu_e\}_{e\in E}$ at convergence

To determine convergence we use the result in [2] that the stationary solution of the dynamics minimizes the transport cost:

$$J_\beta = \sum_{e\in E} \ell_e |F_e|^\Gamma \quad , \tag{S43}$$

where $\Gamma = 2(2-\beta)/(3-\beta)$.

COMPUTATIONAL COMPLEXITY

The computational complexity of solving the OT-constrained problems described in the main manuscript depends on a variety of factors, in particular on the choices of the transportation regime β and on the specific constraints selected. Both of these factors can make it harder to search the solution space. Giving a precise theoretical characterization of the overall complexity is challenging, but we can discuss the complexity of individuals steps.

The complexity of the unconstrained problem is dominated by the cost of solving Kirchhoff's law. In our implementation, we carry out this operation by means of a sparse direct solver (UMFPACK, Unsymmetric MultiFrontal sparse LU Factorization [3]) that performs a LU decomposition for each column of the matrix version of the right-hand side of Eq. (S42). The total computational complexity of this process scales as $O(Mn_{nnz})$, where n_{nnz} is the number of non-zero entries of the Laplacian L^j. This computational complexity could in principle be further reduced using multigrid methods [4]; we do not explore this here.

The constrained case requires further solving the quadratic program, which, if solved via an interior point method, roughly takes $O(E^{3.5})$ elementary operations (there are E decision variables and at most E constraints, where E is the number of edges) [5]. We note however, that this is a conservative estimate and that interior point methods are in practice often much faster than their worst-case complexity estimate. In case of capacity constraints, the solution of the quadratic program can be carried out in $O(E)$ steps in the worst-case, when all constraints are active. In addition, in case of the (nonlinear) budget and capacity constraints each fixed-point iteration has complexity $O(E)$, and in our experiments we required less than ten fixed-point iterations for convergence.

In addition to these theoretical considerations, we show results from numerical experiments to assess how various configurations scale with the network sizes N and E. Specifically, we investigate the running (in seconds) of four methods, namely the Unconstrained, Capacity contraint, Budget constraint, and Non-linear constraint, on network of varying sizes (generated synthetically as in the other experiments presented in this paper) and for two regimes of β, as shown in Fig. S2. The methods were implemented with Algorithm 1.

We observe that Non-linear constraint is the type of constraint that is the slowest to run, possibly because of solving numerically the auxiliary optimization problem, which can be done analytically in the other cases.

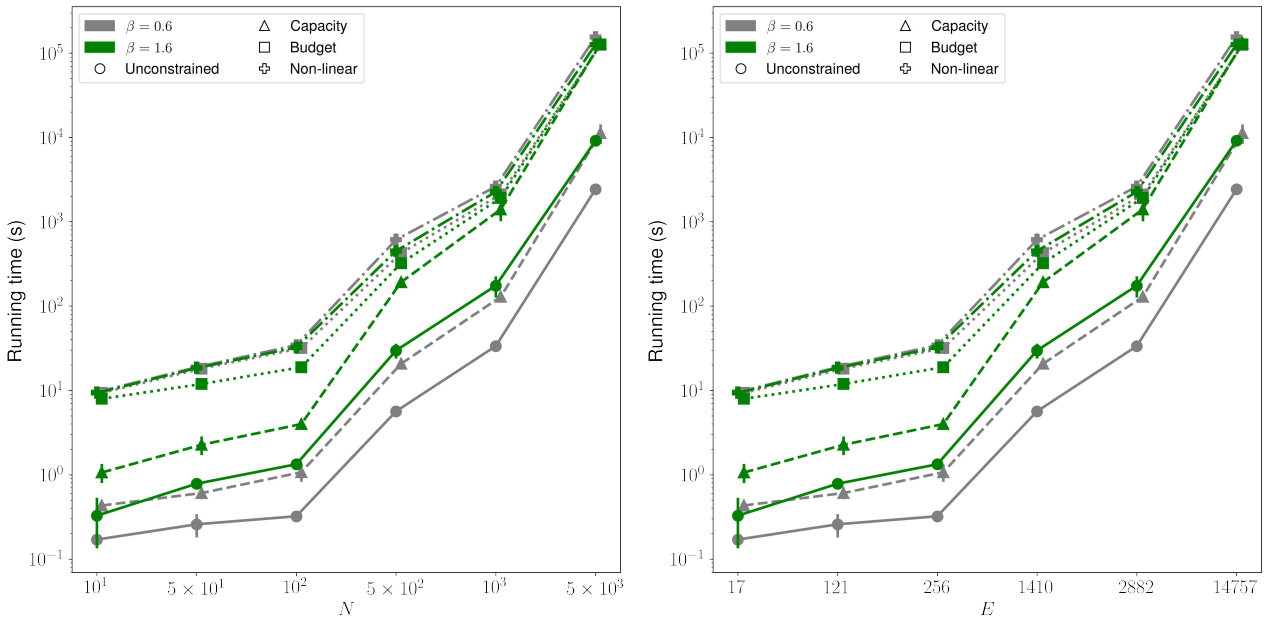


FIG. S2. Running time as a function of N and E. We show two scenarios of β, i.e., penalizes traffic congestion ($\beta < 1$) and encourages path consolidation ($\beta > 1$). Markers and error bar are averages and standard deviations over 10 network realizations. Other settings: $\delta = 1/2$, $c_e = 75$, and $b = \frac{1}{2}\sum_e \mu_e$, where μ_e is that of Unconstrained.

ADDITIONAL EXPERIMENTAL RESULTS

Synthetic network generation and OD pairs

We generated a planar network to mimic a real transport network. To construct this network, we positioned $N = 300$ nodes randomly within a unit square domain (i.e., $[0,1] \times [0,1]$) area, and its Delaunay triangulation is extracted. As a last step, we ensure only edges within a specified distances are retained. This layout forms our initial network structure, resembling a planar graph.
To set the traffic demands (i.e., origin-destination pair) we rewire all origin-destination pairs to a central destination. However, our method applies to any other choice of traffic demands as origin-destination pairs. Specifically, we can:

- rewire each origin to a random and unique destination (i.e., heterogeneous destinations),
- also route a ratio of these origins to a common central destination (i.e., say 45% has a monocentric destination), whereas the remaining ratio of origins (55%) have heterogeneous and random destinations.

These scenarios do not require any modification to the structure of the proposed algorithm, and can be accounted for when computing the pressure potentials with Kirchhoff's law (see [6] for examples).

Additional results on real and synthetic networks

This section provides more results to support the ones presented in the main paper.
We measure the Gini coefficient and average path length on the synthetic network, shown in Fig. S3, and that of Grenoble network in Fig. S4. Fig. S5 shows the toplogies for $\beta = 0.6$ and $\beta = 1.0$, in addition to Fig. 3 of main paper. We illustrate (as addition to Fig. 3 of main paper) the edge-wise differences between the algorithms and the unconstrained case in Fig. S6.

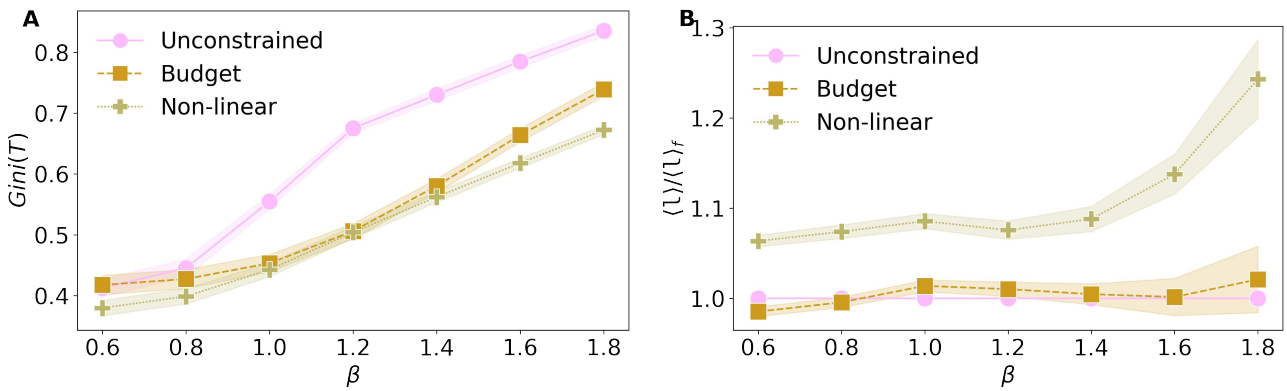


FIG. S3. Results over varying β on synthetic networks. (A) Gini coefficient of the traffic distribution on edges. (B) Ratio of average total path length to that of the unconstrained OT method, denoted as $\langle \mathtt{l} \rangle_f$. Markers and shadows are averages and standard deviations over 20 network realizations, with 100 randomly selected origins for each network realization. All passengers have the same central destination. Settings: $N = 300$, $\delta = 1/2$, $c_e = 70$, $b = \frac{1}{2}\sum_e \mu_e$, where μ_e is that of unconstrained.

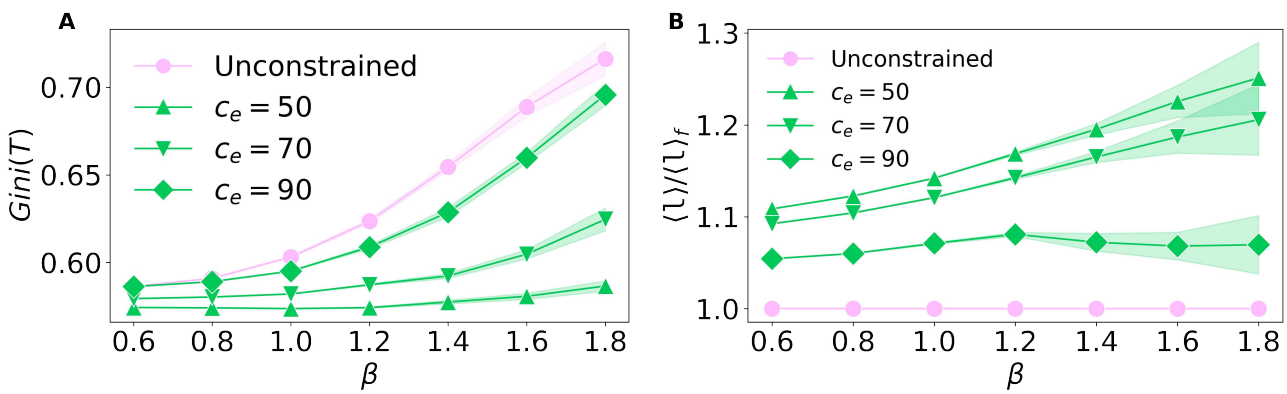


FIG. S4. Results on Grenoble bus network. (A) Gini coefficient of the traffic distribution on the network edges. The edge capacity c_e is the percentile of μ from f, and varied between low, medium and high capacities. Varying c_e helps to understand how the size of highway impacts the traffic. Setting a low capacity optimizes traffic better than high values. (B) The ratio of average total path length to that of the unconstrained OT method. These results are averaged over 100 randomly selected origin-destination pairs. The origin-destination pairs have been selected so that all the passenger types have a central destination. Markers and shadows indicates average and standard deviation, respectively.

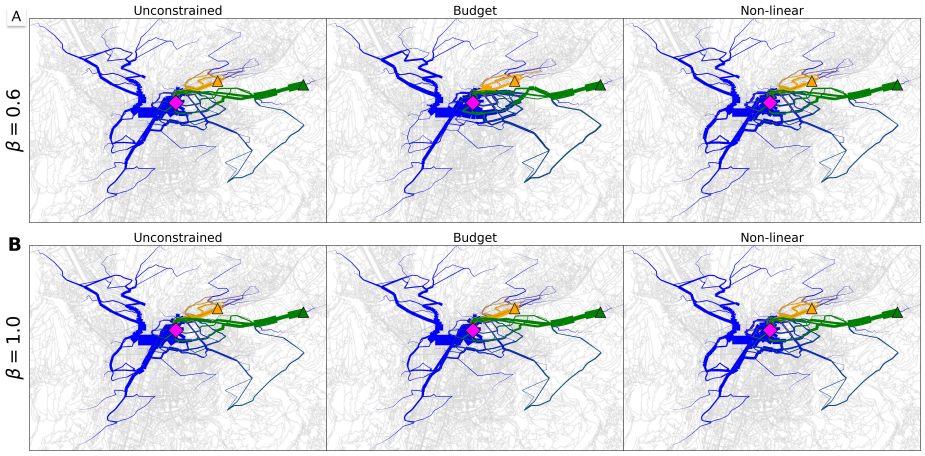


FIG. S5. Constrained OT on Grenoble road network. (A-B) Path trajectories of the Grenoble road network for the unconstrained OT (Unconstrained), using a budget constraint (Budget); a capacity constraint and a non-linear budget (Non-linear). Other settings are the same as in Fig. 3.

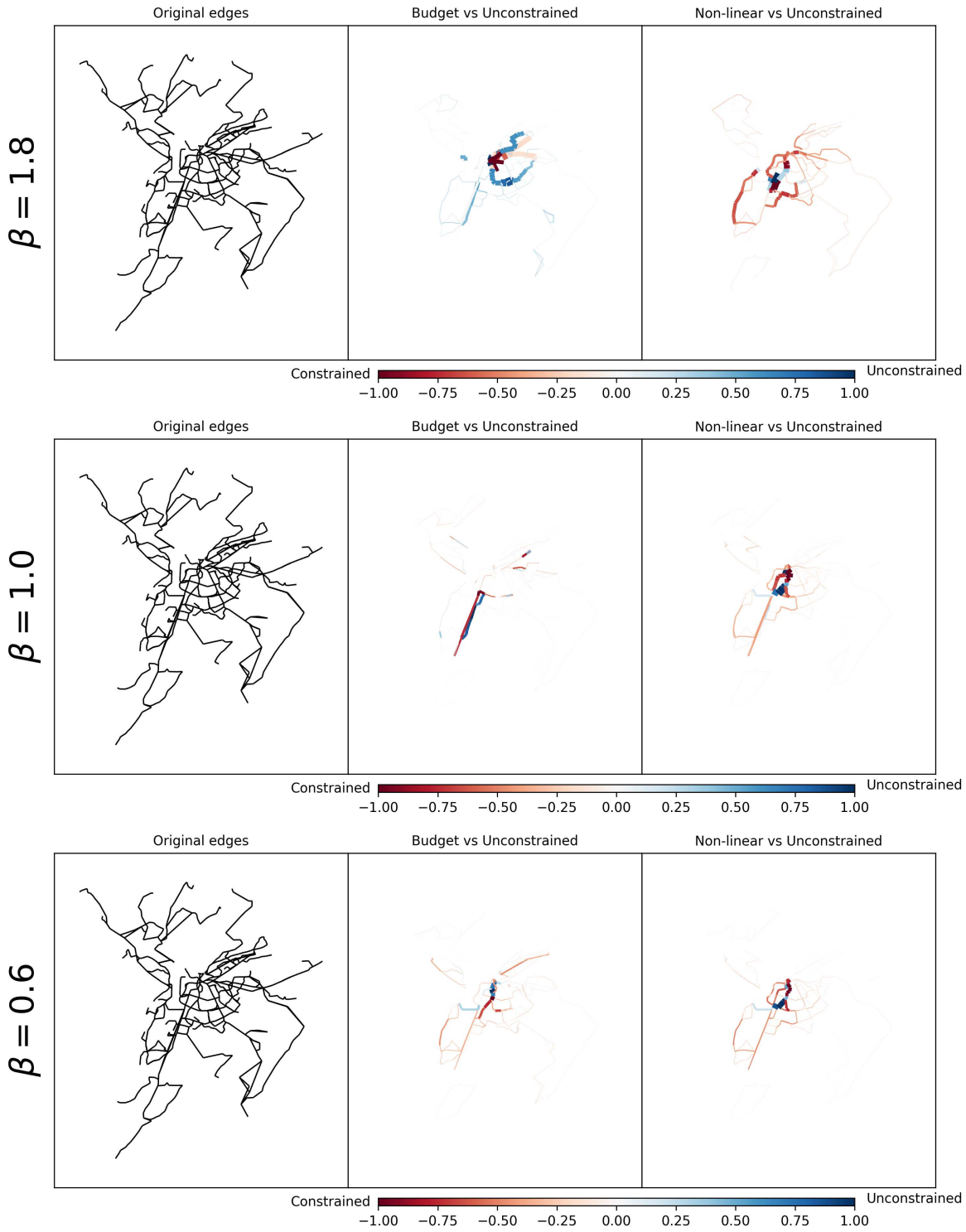


FIG. S6. Edge-wise difference only plots on Grenoble bus network. We compute the difference of traffic on edges as $T_e^c - T_e^u$ where T_e^c and T_e^u denote the traffic for constrained and unconstrained methods, respectively. Other settings are the same as in Fig. 3.

[1] M. Muehlebach and M. I. Jordan, Journal of Machine Learning Research **23**, 1 (2022).
[2] A. Lonardi, E. Facca, M. Putti, and C. De Bacco, Physical Review Research **3**, 043010 (2021).
[3] T. A. Davis, ACM Trans. Math. Softw. **30**, 196–199 (2004).
[4] W. L. Briggs, V. E. Henson, and S. F. McCormick, *A multigrid tutorial* (SIAM, 2000).
[5] Y. Nesterov, *Introductory lectures on convex optimization: A basic course* (Springer Science & Business Media LLC, 2004).
[6] A. A. Ibrahim, D. Leite, and C. De Bacco, Physical Review E **105**, 064302 (2022).

scientific reports

OPEN

Community detection in networks by dynamical optimal transport formulation

Daniela Leite[1,3]✉, Diego Baptista[1,3], Abdullahi A. Ibrahim[1], Enrico Facca[2] & Caterina De Bacco[1]

Detecting communities in networks is important in various domains of applications. While a variety of methods exist to perform this task, recent efforts propose Optimal Transport (OT) principles combined with the geometric notion of Ollivier–Ricci curvature to classify nodes into groups by rigorously comparing the information encoded into nodes' neighborhoods. We present an OT-based approach that exploits recent advances in OT theory to allow tuning between different transportation regimes. This allows for better control of the information shared between nodes' neighborhoods. As a result, our model can flexibly capture different types of network structures and thus increase performance accuracy in recovering communities, compared to standard OT-based formulations. We test the performance of our algorithm on both synthetic and real networks, achieving a comparable or better performance than other OT-based methods in the former case, while finding communities that better represent node metadata in real data. This pushes further our understanding of geometric approaches in their ability to capture patterns in complex networks.

Complex networks are ubiquitous, hence modeling interactions between pairs of individuals is a relevant problem in many disciplines[1,2]. Among the variety of analyses that can be performed on them, community detection[3–6] is a popular application that involves finding groups (or communities) of nodes that share similar properties. The detected communities may reveal important structural properties of the underlying system. Community detection has been used in diverse areas including, discovering potential friends on social networks[7], evaluating social networks[8], personalized recommendation of item to user[9], detecting potential terrorist activities on social platforms[10], fraud detection in finance[11], study epidemic spreading process[12] and so on.

Several algorithms have been proposed to tackle this problem which utilize different approaches, such as statistical inference[13,14], graph modularity[15], statistical physics[16], information theory[17] and multifractal topological analysis[18]. Here, instead, we adopt a recent approach connecting community detection with geometry, where communities are detected using geometric methods like the Ollivier–Ricci curvature (ORC) and we exploit a dynamical approach of optimal transport theory to calculate this efficiently and flexibly across various transportation regimes.

In Riemannian geometry, the sign of the curvature quantifies how geodesic paths converge or diverge. In networks, the ORC plays a similar role: edges with negative curvature are traffic bottlenecks, whereas positively curved ones allow mass to flow more easily along the network. Defining communities as structures that allow robust transport of information, we could cluster edges based on their curvature: those with positive curvature can be clustered together, while those with negative curvature may be seen as "bridges" connecting different communities. The idea of using Ricci curvature to find communities on networks was first proposed by Jost and Liu [19] and then further explored in subsequent works[20–23]. Our work follows a similar approach as in[22,23] to calculate the OR curvature, but generalizes it for the cases of branched[24,25] and congested[26] optimal transport problems, building from recent results[27,28]. Specifically, our algorithm allows to efficiently tune the sensitivity to detecting communities in a network, through a parameter that controls the flow of information shared between nodes. We perform a comprehensive comparison between the proposed algorithm and existing ones on synthetic and real data. Our algorithm, named ORC-Nextrout, detects communities in synthetic networks with similar or higher accuracy compared to other OT-based methods in the regime where inference is not trivial, i.e. the inference problem is neither too easy nor too difficult to solve, and thus communities are only partially retrieved. This is also observed in a variety of real networks, where the ability to tune between different transportation regimes

[1]Max Planck Institute for Intelligent Systems, Cyber Valley, 72076 Tübingen, Germany. [2]Univ. Lille, Inria, CNRS, UMR 8524 - Laboratoire Paul Painlevé, 59000 Lille, France. [3]These authors contributed equally: Daniela Leite and Diego Baptista. ✉email: daniela.leite@tuebingen.mpg.de

allows finding at least one result that outperforms other methods, including approaches based on statistical inference and modularity.

Related work. The idea of exploring the geometrical properties of a graph, and in particular curvature, has been explored in different branches of network science, ranging from biological[29] to communication[30] networks. Intuitively, the Ricci curvature can be seen as the amount of volume through which a geodesic ball in a curved Riemannian manifold deviates to the standard ball in Euclidean spaces[31]. When defined in graphs, it indicates whether edges (those with positive values for the curvature) connect nodes inside a cluster, or if they rather bond different clusters together (those with negative values for the curvature).

Previous works[32–35] extended the idea of the OR curvature. In[32], the authors introduced the concept of "resistance curvature" for both nodes and edges. Taking inspiration from electrical circuits, this approach assigns a resistance being applied by the whole network from a current that flows between any two edges and correlates this to known concepts of OR discrete curvature. The resistance curvature provides a natural way to define the Ricci flow. In[33] the authors proposed a *dynamical* version of the OR curvature, where a continuous-time diffusion process is defined for every node, at different time scales. In this context, the dynamical perspective is used to frame probability masses at nodes in terms of diffusion processes, e.g. those deployed in random walks. In our work instead, the dynamics enters to solve efficiently the underlying optimization problem required to compute the OR curvature. Regardless of the choice of the distribution that characterizes mass on nodes, this quantity is then used to define the curvature of the edges of the graph. Previous works have typically defined the OR curvature in terms of the 1-Wasserstein distance. In contrast, we take a more general approach and explore the usage of the β-Wasserstein, where $\beta \in (0, 2]$, to account for a variety of OT problems, ranging from branched to congested transportation.

Other discrete graph curvature approaches include the Ollivier–Ricci (OR) curvature based on the Optimal Transport theory introduced by Ollivier,[36,37] and Forman–Ricci curvature introduced by Forman[38]. While the graph Laplacian-based Forman curvature is computationally fast and less geometrical, we focus on the OT-based approach due to its more geometric nature. Some applications of the Ollivier–Ricci curvature include network alignment[39] and community detection[22,23,40].

On the other hand, community detection in networks is a fundamental area of network science, with a wide range of approaches proposed for this task[3,4,41]. These include methods based on statistical mechanical models[16,17,42], probabilistic generative models[13,43–45], nonnegative matrix factorization[46], spectral methods[47,48], multifractal topological analysis[18] and modularity optimization[15,49,50]. In contrast, our work is inspired by recent OT-based methods[22,23] for community detection. These methods consider the OR curvature to sequentially identify and prune negatively curved edges from a network to identify communities. While our approach also considers OR curvature to prune edges, it controls the flow of information exchanged between nodes by means of a traffic-penalization parameter, making the edge pruning completely dynamic. This is detailed in "β-Wasserstein community detection algorithm" section.

β-Wasserstein community detection algorithm

In this section, we describe how our approach solves the community detection problem. As previously stated, we rely on optimal transport principles to find the communities. To solve the optimal transport problem applied in our analysis we use the discrete *Dynamic Monge-Kantorovich* model (*DMK*), as proposed by Facca et al.[51,52] to solve transportation problems on networks.

We denote a weighted undirected graph as $G = (V, E, W)$, where V, E, W are the set of nodes, edges, and weights, respectively. We use the information of the neighborhood of a node i, $\mathcal{N}(i) = \{j \in V | (i, j) \in E\}$, to decide whether a node belongs to a given community. We do this by comparing a distribution defined on $\mathcal{N}(i)$ with those defined on other nodes close to i. There are several choices that can be made for this. For instance, one could frame this in the context of diffusion processes on networks and relate the distribution to random walkers traveling along the network with a certain jump probability[33]. Here we follow previous work[39] and assign it as m_i^α, where $m_i^\alpha(k) := \alpha$ if $k = i$ and $m_i^\alpha(k) := (1-\alpha)/|\mathcal{N}(i)|$ if $k \in \mathcal{N}(i)$. Intuitively, the distribution m assigns a unit of mass to i and its connections: α controls how much weight node i should have, and once this is assigned, its neighbors receive the remaining mass in an even way. We use $\alpha = 0$ in all the experiments reported in this manuscript, i.e. the mass is equally distributed on the neighbors. This corresponds to a one-step transition probability for a random walker in the context of diffusion processes.

The next step is to compare the distribution m_i^α of node i to that of its neighbors. Consider an edge $(i, j) \in E$ and m_j, the distribution defined on node j, neighbor of i. We assume that if i and j belong to the same community, then both nodes may have several neighbors in common and, therefore, m_i and m_j should be similar. Note that this is valid for both assortative and disassortative community structures. In the former case, nodes are more likely to interact within the same community, while in the latter case we have the opposite, nodes are more likely to interact across communities[2,4]. When there is a consistent community pattern for all groups (e.g. all communities are assortative), this idea of comparing the distributions m_i^α may be appropriate to detect communities. On the contrary, it may be difficult to perform this task in networks with mixed connectivity patterns, where some communities are assortative and others are disassortative. This makes it difficult to detect communities as edges within an assortative community are shortened, likewise edges between a node in a disassortative and a node in an assortative one. This may confuse the algorithm, as both types of edges are shortened. A careful treatment of these cases is an interesting direction for future work.

To estimate the similarity between m_i and m_j we use OT principles. Specifically, we compute the cost of transforming one distribution into the other. This is related to the cost of moving the mass from one neighborhood to the other, and it is assumed to be the weighted shortest-path distance between nodes belonging to $\mathcal{N}(i)$ and

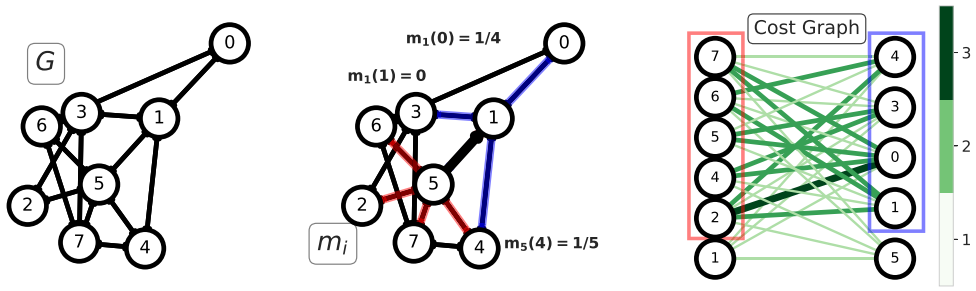


Figure 1. Left: an example graph G where edges have unitary weights. Center: the edge (1, 5) (bold black line) is selected to define the OT problem between m_1, m_5; neighborhoods of nodes 1 and 5 are highlighted with blue and red edges and are used to build the corresponding distributions m_1, m_5. Right: The complete bipartite graph B_{15} where the OT problem is defined. The color intensity of the edges represents the distance between the associated nodes on the graph G, as shown by the color bar. m_1 and m_5 are both defined for $\alpha = 0$, i.e. no mass is left in 1 and 5.

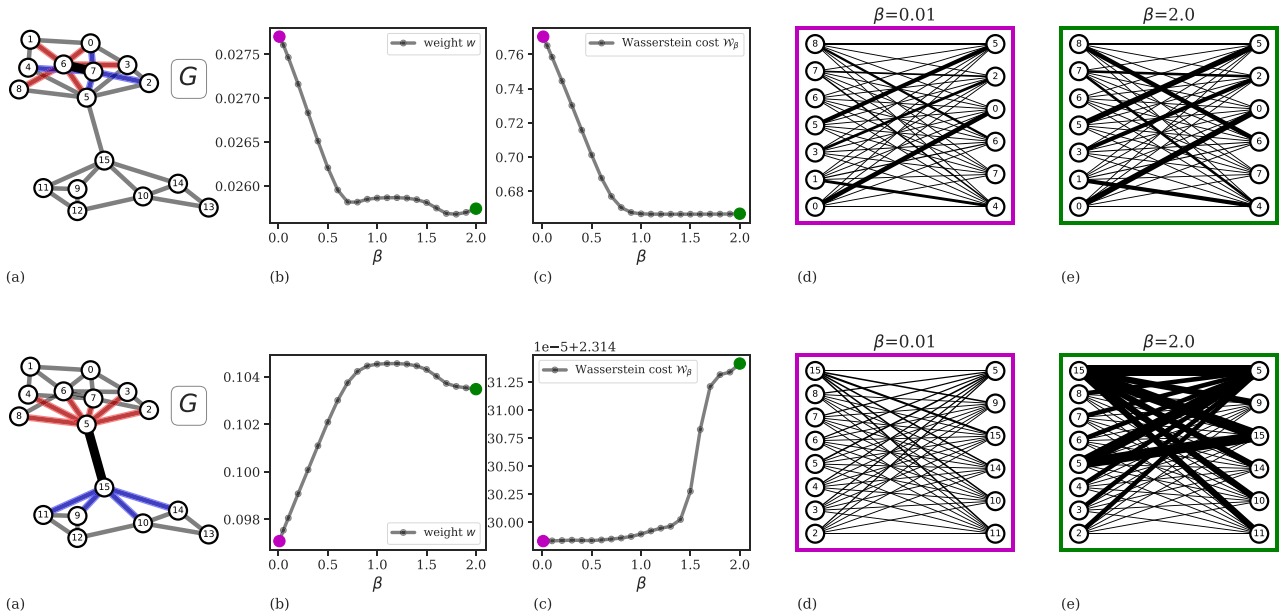


Figure 2. Visualization of how β impacts intra-community and inter-community edge weights. (**a**) Examples of intra-community (top panel) and inter-community (bottom panel) structures between nodes 6 and 7, and nodes 5 and 15, respectively. (**b**) The weight of edge (6, 7) decreases when $0 < \beta < 0.6$, while for $0.5 < \beta < 2.0$ it reaches a minimum, and then slightly increases again. Similar but opposite pattern is observed for the edge (5, 15). (**c**) The β-Wasserstein cost: for intra-community edges, $\beta > 1$ consolidates traffic in the network as the Wasserstein cost stabilizes, making it minimum for the extreme value $\beta = 2$, whereas it is maximized in the case of the inter-community edge. (**d**, **e**) Example cost graphs B_{67} (top) and B_{515} (bottom) with fluxes solution of the OT problem (edge thickness is proportional to the amount of flux) in the regimes of small (**d**) and high (**e**) values of β.

$\mathcal{N}(j)$. A schematic representation of the algorithm can be seen in Fig. 1. The OT problem is solved in an auxiliary graph, the complete bipartite network $B_{ij} = (V_{ij}, E_{ij}, \omega_{ij})$ where $V_{ij} := (V_i, V_j) := (\mathcal{N}(i) \cup \{i\}, \mathcal{N}(j) \cup \{j\})$, E_{ij} is made of all the possible edges between V_i and V_j. The weights of the edges are given by the weighted shortest path distance d between two nodes measured on the input network G.

The similarity between m_i and m_j is the Wasserstein cost $\mathscr{W}(m_i, m_j, \omega_{ij})$ of the solution of the transportation problem. In its standard version, this number is the inner product between the solution Q, a vector of flows defined on edges, and the cost ω_{ij}. In our case, since the DMK model allows to control the flow of information through a hyperparameter $\beta \in (0, 2]$, we define the *β-Wasserstein cost*, $\mathscr{W}_\beta(m_i, m_j, \omega_{ij})$, as the inner product of the solution $Q = Q(\beta)$ of the DMK model and the cost ω_{ij}. For $\beta = 1$ we compute the 1−Wasserstein distance between m_i and m_j, while for $\beta \neq 1$ the influence of β in the solution of the transportation problem can be seen in Fig. 2. When $\beta < 1$, more edges of B tend to be used to transport the mass, thus we observe congested transportation[26]. When $\beta > 1$ fewer edges are used, hence we observe branched transportation, and the

β-Wasserstein cost coincides with a branched transport distance[25,53]. The idea of tuning β to interpolate between various transportation regimes has been used in several works and engineering applications[27,54–59].

Calculating the Wasserstein cost is necessary to determine our main quantity of interest, the discrete Ollivier–Ricci curvature, defined as

$$\kappa_\beta(i,j) := 1 - \frac{\mathscr{W}_\beta(m_i, m_j, \omega_{ij})}{d_{ij}} \quad , \tag{1}$$

where d_{ij} is the weighted shortest path distance between i and j as measured in G. Intuitively, if i and j are in the same communities, several $k \in V_i$ and $\ell \in V_j$ will be also directly connected. Thus, the Wasserstein distance between m_i and m_j will be shorter than d_{ij}, yielding a positive $\kappa_\beta(i,j)$. Instead, when i and j are in different communities, their respective neighbors will be unlikely connected, hence $d_{ij} < \mathscr{W}_\beta(m_i, m_j, \omega_{ij})$, yielding a negative $\kappa_\beta(i,j)$.

The Ricci flow algorithm on a network is defined by iteratively updating the weights of the graph G[22,23]. These are updated by combining the curvature and shortest path distance information[36]. We redefine these updates using our proposal for the Ollivier–Ricci curvature:

$$w_{ij}^{(t+1)} := d_{ij}^{(t)} - \kappa_\beta^{(t)}(i,j) \cdot d_{ij}^{(t)}, \tag{2}$$

where $w_{ij}^{(t+1)}$ is the weight of edge (i,j) at time t, $w_{ij}^{(0)} = d_{ij}^{(0)}$, and $d_{ij}^{(t)}$ is the shortest path distance between nodes i and j at iteration t. At every time step t, the weights are normalized by their total sum.

The algorithm ORC-Nextrout dynamically changes the weights of the graph G to isolate communities: intra-community edges will be shortened, while inter-community ones will be enlarged. These changes are reached after a different number of iterations of the whole routine depending on the input data. To find the communities, we apply a *network surgery* criterion on the edges based on the stabilization of the modularity of the network, as proposed by Ni et al.[22]. Notice that our algorithm does not need prior information about the number of communities: edges will be enlarged or shortened depending on the optimal transport principles, agnostic to community labeling. The computational complexity of the algorithm is dominated by that of solving the DMK model, which takes $O(|E|^{2.36})$ (estimated numerically) and by computing weighted shortest path distances d_{ij}, which costs $O(|V|^2 \log|V| + |V||E|)$[60]. A pseudo-code of the implementation is shown in Algorithm 1.

Algorithm 1 ORC-Nextrout

Input: $G = (V, E, W)$, traffic rate β, $MaxIterNum \in \mathbb{N}$
Output: updated W
Initialize: edge weights $\mathbf{w}^0 = W$; neighborhood distributions $\mathbf{m}$
for $t \in range(MaxIterNum)$ **do**
  Compute all-pair-shortest-path matrix d^t
  for $e = (i,j) \in E$ **do**
    Build B_{ij}
    Get $Q(\beta) \in \mathbb{R}^{|E_{ij}|}$, $Q(\beta) = DMK(B_{ij}, m_i, m_j, \beta, d^t)$
    Compute $\kappa_\beta(e)$ using $d^t(e)$ and $Q(\beta)$
    Compute $\mathbf{w}_\beta(e)$ using $\kappa_\beta(e)$ and $d^t(e)$
  end for
  Update $\mathbf{w}^t = \mathbf{w}_\beta$
end for

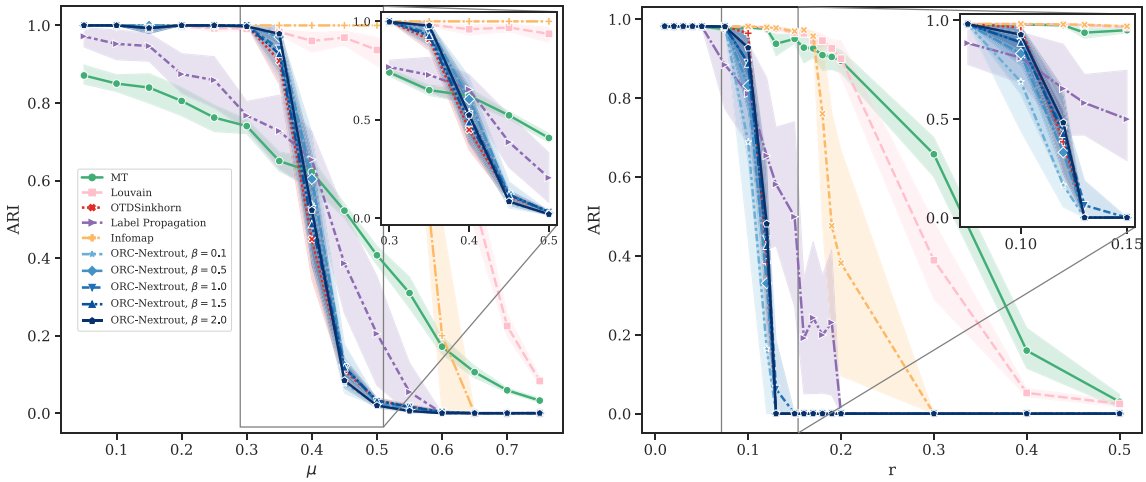
Results on community detection problems

Synthetic networks. To investigate the accuracy of our model in detecting communities, we consider synthetic networks generated using the *Lancichinetti–Fortunato–Radicchi* (LFR) benchmark[61] and the *Stochastic Block Model* (SBM)[62]. Both models provide community labels used as *ground-truth* information during the classification tasks.

Lancichinetti–Fortunato–Radicchi benchmark: this benchmark generates undirected unweighted networks G with disjoint communities. It samples node degrees and community sizes from power law distributions, see Fig. 4 for an example. One of its advantages is that it generates networks with heterogeneous distributions of degrees and community sizes. The main parameters in input are the number of nodes N, two exponents τ_1 and τ_2 for the power law distributions of the node degree and community size respectively, the expected degree d of the nodes, the maximum number of communities on the network K_{max} and a fraction μ of inter-community edges incident to each node. To test the performance of our algorithm, we use the set of LFR networks used and provided by the authors of[22]. We set $\tau_1 = 2, \tau_2 = 1, d = 20, K_{max} = 50$ and $\mu \in [0.05, 0.75]$.

Stochastic Block Model: this model probabilistically generates networks with non-overlapping communities. One specifies the number of nodes N and the number of communities K, together with the expected degree d of a node and a ratio $r \in [0,1]$. Networks are generated by connecting nodes with a probability $r * p_{intra}$ if they belong to different communities; p_{intra} if they are part of the same community, where $p_{intra} = d \times K/N$. Notice that the smaller the ratio r is, the fewer inter-community connections would exist, which leads to networks with a more distinct community structure.

We set $N = 500, K = 3, d = 15$ and $r \in [0.01, 0.5]$ and generate 10 random networks per value of r.



(a) LFR **(b)** SBM

Figure 3. Results on LFR and SBM synthetic data. Performance in detecting ground-truth communities is measured by the ARI score. Markers and shadows are the averages and standard deviations over 10 network realisations with the same value of the parameter used in generation. Markers' shape denote different algorithms. (**a**) LFR graph with $N = 500$ nodes and different values of K ranging from (17, 22). (**b**) SBM with $N = 500$ nodes, $K = 3$ communities and average degree $d = 15$. The parameter r is the ratio of inter-community with intra-community edges. The inset on each plot zooms in the central parts of the plots.

Results. To evaluate the performance of our method in recovering the communities, we use the *Adjusted Rand Index* (ARI)[63]. ARI compares the community partition obtained with the *ground truth* clustering. It takes values ranging from 0 to 1, where ARI = 0 is equivalent to random community assignment, and ARI = 1 denotes perfect matching with the ground truth communities, hence the higher this value, the better the recovery of communities. A more detailed presentation of this metric is given in the Supplementary Information.

We test our algorithm for different types of information spreading in our OT-based model, as controlled by the parameter β, using the software developed in[64] (available at https://gitlab.com/enrico_facca/dmk_solver). We used $\beta = 1$, i.e., standard Wasserstein distance; $\beta \in \{0.1, 0.5\}$ for congested transportation, enforcing broad spreading across neighbors; and $\beta \in \{1.5, 2\}$ to favor branching schemes, where fewer edges are used to decide which community a node should belong to. For OT-based algorithms where we update the weights in Eq. (2) for 15 times ($MaxIterNum = 15$ in Algorithm 1). Since in some cases the ARI score does not consistently increase with the number of iterations, we show results only for the iteration that maximizes the score.

The results in Fig. 3 show the performance on both LFR and SBM benchmarks with OT-based methods, our method for various β and one based on the Sinkhorn algorithms (OTDSinkhorn)[65,66]. Our main goal is to assess the impact of tuning between different transportation regimes (as done by β) in terms of community detection via OT principles. Nevertheless, to better contextualize the performance of OT-based algorithms in the wide spectrum of community detection methods, we also include comparisons with algorithms that are not OT-based. Namely, we consider a probabilistic model with latent variables (MT)[13], two modularity-based algorithms, Label Propagation[14] and Louvain[50], and with a flow-based algorithm, Infomap[17]. Our algorithm outperforms OTD-Sinkhorn for various values of β in an intermediate regime where OT-based inference is not trivial, i.e. detecting communities is neither too easy nor too difficult. This occurs in both the LFR and SBM benchmark, as shown in Fig. 3. For lower and higher values of the parameters, performance is similar and close to the two extremes of ARI = 0 and 1. OT-based methods have a similar sharp decay in performance from the regime where inference is easy to the more difficult one, as also observed in[22]. The other community detection methods have smoother decay, but with lower performance in the regime where OT-based approaches strive, except for Label Propagation and MT, which are more robust in this sense. In the intermediate regime where inference is not trivial (i.e. along the sharp decay of OT-based methods), we observe that different values of β give higher performance than OTDSinkhorn in most cases. For SBM the highest performance is consistently achieved for high $\beta = 2$, while for LFR the best β varies with μ. A qualitative example where ORC-Nextrout is performing better than OTDSinkhorn, in an instance of LFR of this intermediate regime, is shown in Fig. 4. Note that in this case, ORC-Nextrout perfectly recovers the 21 communities described by the ground-truth network, whereas OTDSinkhorn merges three of the central communities into one group, therefore recovering only 19 groups.

These results suggest that practitioners may choose the β that gives the best performance in detecting communities, e.g. the one that maximizes ARI or other relevant metrics depending on the application at hand. We show examples of this on real data below.

Analysis of real networks. Next, we evaluate our model on various real datasets[67] containing node metadata that can be used to assess community recovery. While failing to recover communities that align well with node metadata should not be automatically interpreted as a model's failure[68] (e.g. the inferred communities and the chosen node metadata may capture different aspect of the data), having a reference community structure to

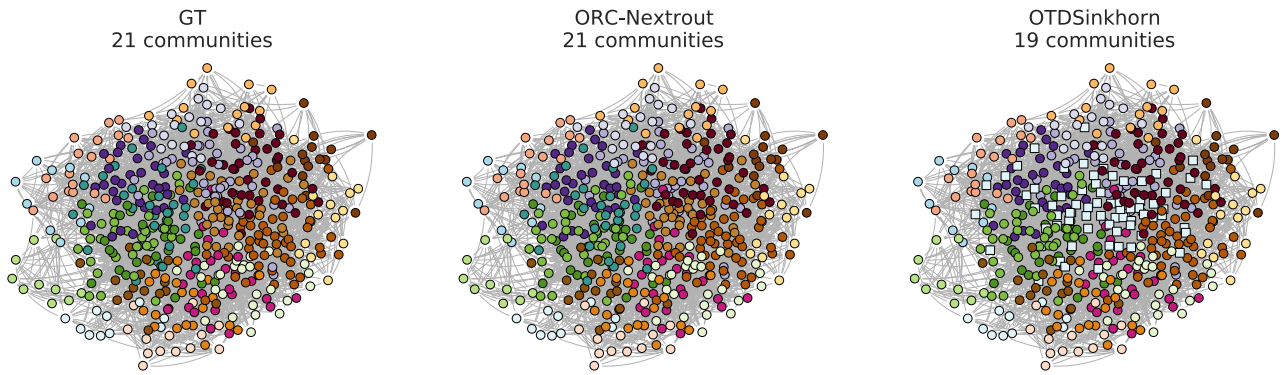


Figure 4. Example of community structure on a synthetic LFR network. The rightmost panel shows the ground-truth community structures to be predicted in an LFR network generated using $\mu = 0.35$. This network is one sample of the synthetic data used in Fig. 3. Square-shaped markers denote nodes that are assigned to communities different from those in ground-truth. In the middle and last panels, ORC-Nextrout with $\beta = 2$ perfectly retrieves the 21 communities, while OTDSinkhorn predicts only 19 communities with an ARI score of 0.73, wrongly assigning ground-truth dark green and light brown (square-shaped) nodes to the light blue community.

Dataset	N	E	K	AvgDeg	AvgBtw	AvgClust
Les Miserables	77	254	11	6.6	0.0219	0.5731
Dolphins	62	159	4	5.1	0.0393	0.2590
American football	115	613	12	10.7	0.0133	0.4032
Political books	105	441	3	8.4	0.0202	0.4875

Table 1. Real networks description. We report statistics for the real networks used in our experiments. N and E denote the number of nodes and edges, respectively. K is the number of communities in the ground truth data. AvgDeg, AvgBtw and AvgClust are the average degree, betweenness centrality and average clustering coefficient, respectively.

compare against allows one to inspect quantitatively difference between models. These real networks differ on structural features like number of nodes, average degree, number of communities, and other standard network properties as detailed in Table 1. Specifically, we consider (i) a network of co-appearances of characters in the novel *Les Misérables*[69] (Les Miserables). Edges are built between characters that encounter each other. (ii) A network of 62 bottlenose dolphins in a community living off Doubtful Sound, in New Zealand[70] (Dolphins). The nodes represent dolphins, and the edges indicate frequent associations between them. This network is clustered into four groups, conjectured as clustered from one population and three sub-populations based on the interactions between dolphins of different sex and ages[71]. The dolphins were observed between 1994 and 2001. (iii) A network of Division I matches of American Football during a regular season in the fall of 2000[49] (American football). Nodes represent teams, and edges are games between teams. Teams can be clustered according to their football college conference memberships. (iv) A network of books on US politics published around the 2004 presidential election and sold by an online bookseller[72] (Political books). Nodes represent the books, and the edges between books are frequent co-purchasing of books by the same buyers. Books are clustered based on their political spectrum as neural, liberal, or conservative.

OT-based algorithms outperform other community detection algorithms in detecting communities aligned with node metadata for two of the four studied datasets, as shown in Fig. 5. In particular, ORC-Nextrout has the highest accuracy performance considering the best performing β in these cases. The impact of tuning this parameter is noticeable from these plots, as the best-performing value varies across datasets. In Les Miserables and Dolphins networks, $\beta < 1$ has better performance, while in American Football the best performing value is for $\beta > 1$. Performance is similar across OT-based methods in the Political books network. In Fig. 6 we show the communities detected by the best-performing ORC-Nextrout version together with OTDSinkhorn and Infomap in Les Miserables and Political books (see Supplementary Material for the remaining datasets). Focusing on Les Miserables, we see how ORC-Nextrout successfully detects three characters in the green communities, in particular a highly connected node in the center of the figure (in dark green). Notice that these are placed in the same community (pink or black) by OTDSinkhorn. Thus ORC-Nextrout achieves a higher ARI than OTDSinkhorn. Both OT-based approaches retrieve well communities exhibiting clustering patterns with many connections within the community. Instead, they both divide the communities with a hub and spokes structure due to the lack of common connections within the group.

The communities detected in both datasets highlight the tendency of OT-based methods to extract a larger number of communities than those observed from node metadata. Among these extra communities, some are made of a few nodes (e.g. the light-blue and violet), while others are made of one isolated node each (highlighted

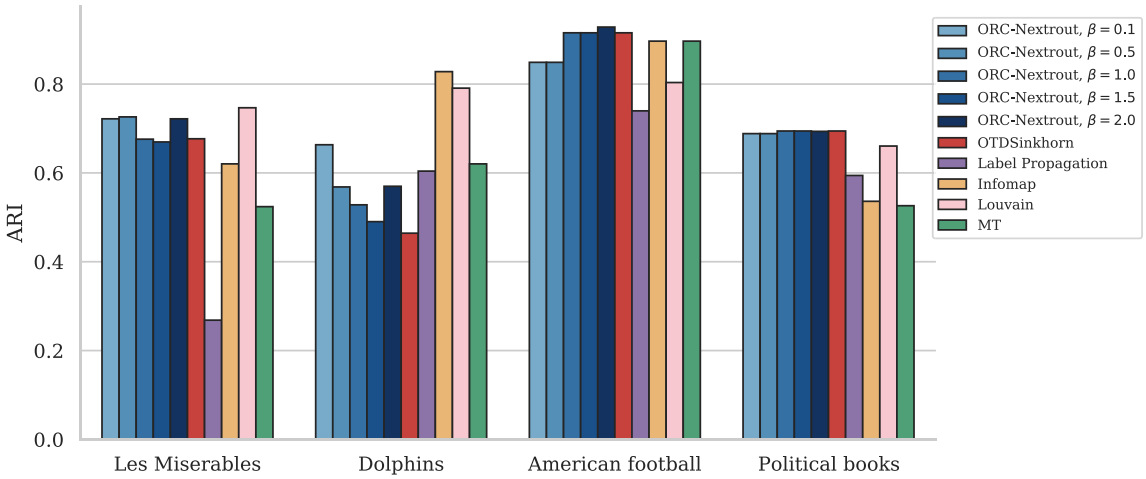


Figure 5. Results on real data. Performance in terms of recovering communities using metadata information is calculated in terms of the ARI score. ORC-Nextrout shows competing results against all methods with different optimal β across datasets.

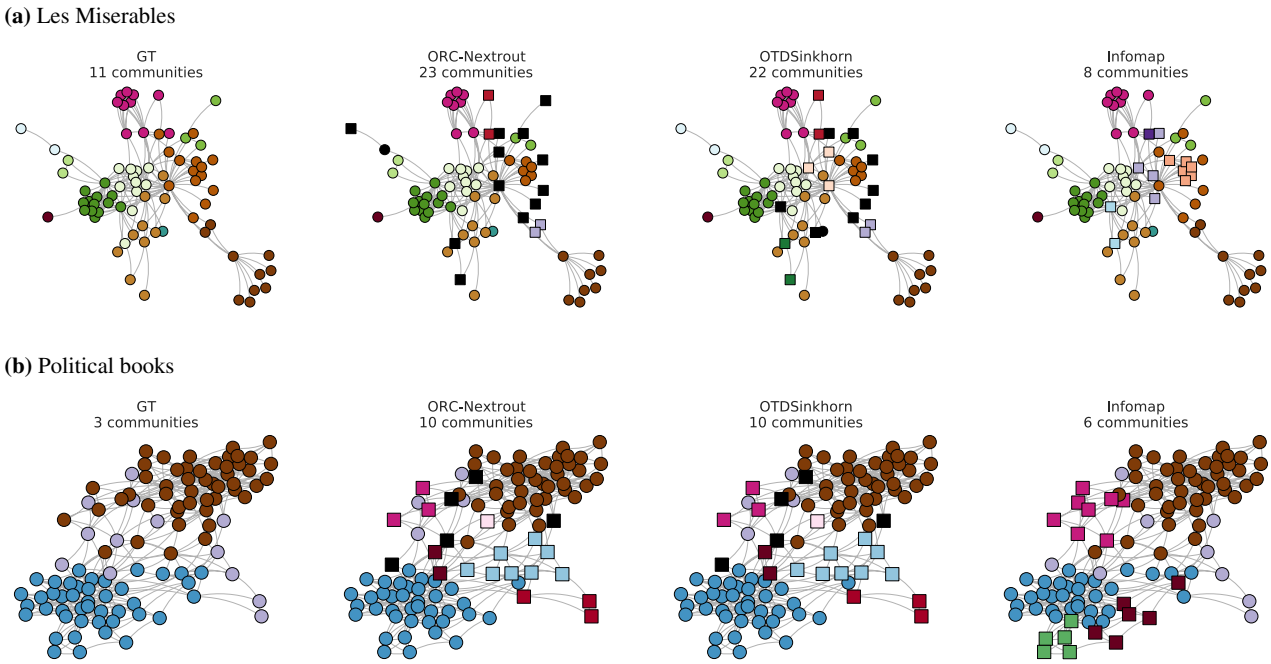


Figure 6. Communities in real networks. We show the communities inferred for Les Miserables (**a**) and Political books (**b**) by ORC-Nextrout ($\beta = 0.5, 0.1$ for top and bottom rows respectively), OTDSinkhorn and Infomap and compare against those extracted using node attributes (GT). The visualization layout is given by the *Fruchterman-Reingold force-directed* algorithm[73], therefore, groups of well-connected nodes are located close to each other. Dark nodes represent individual nodes who are assigned to isolated communities by OT-based methods. Square-shaped markers denote nodes assigned to communities different from those obtained from node metadata.

in black). This is related to the fact that OT-based methods perform particularly well for networks with internally densely connected community structures, but may be weaker for community structures that are sparsely connected[23]. One could potentially assign these nodes to larger communities, for instance, by preferential attachment as done in[23], thus in practice reducing the number of communities. Devising a principled method or criterion to do this automatically is an interesting topic for future work. This tendency is further corroborated by the fact that OT-based algorithms recover robustly the two communities that are mostly assortative (blue and brown in the figure) in the Political books network, while they struggle to recover the disassortative community depicted in the center (violet). This community has several connections with nodes in the other two communities and has been separated into smaller groups by OT-based approaches, as described above. This also highlights the

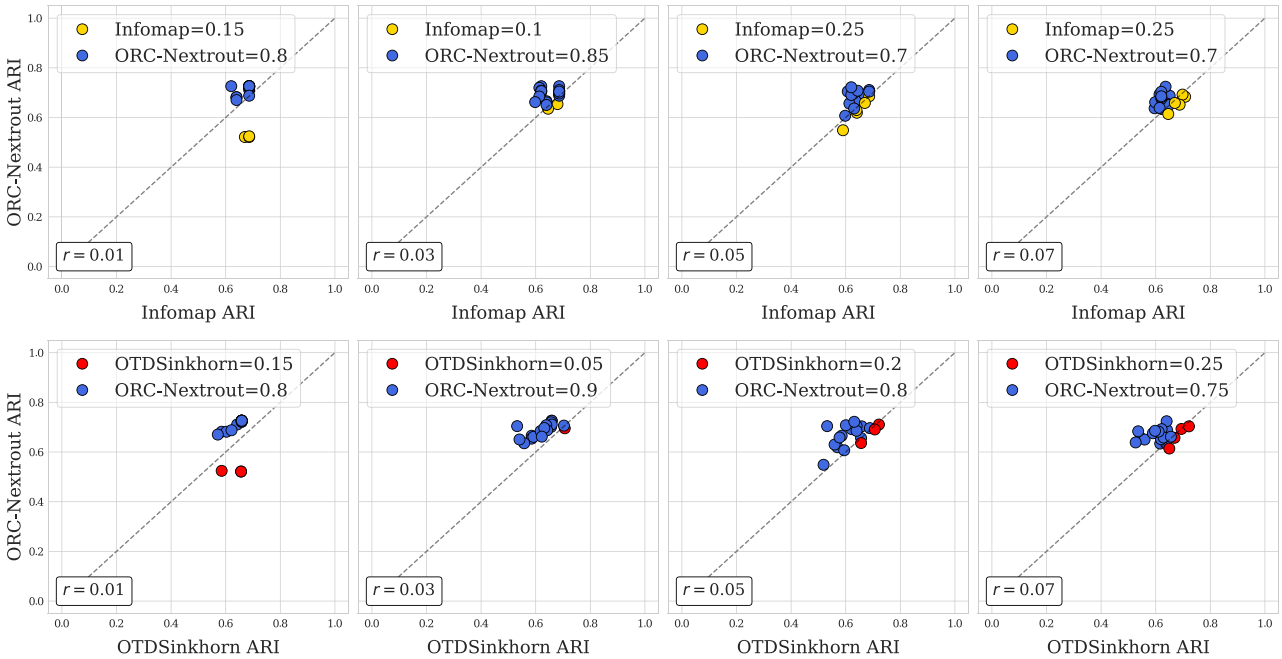


Figure 7. Removing intra-community edges test on Les Miserables data. Markers correspond to 20 instances of semi-synthetic networks generated from real data. Their (x, y) coordinates are ARI scores of the indicated method on the axes. Colors are given by the best performing algorithm, e.g. if $x > y$, the color of the associated method to x is chosen. The legend shows the percentage of times that the corresponding method outperforms the other. The parameter r describes the proportion of entries for the adjacency matrix A that have been changed. This increases from left to right.

need for methods that are robust against situations where mixed connectivity patterns arise, i.e. a combination of assortative and disassortative communities coexisting in a network.

Two tests on semi-synthetic networks. To further investigate the different performance gaps between the various approaches, we expand the comparison between the OT-based methods and Infomap on two semi-synthetic scenarios generated from Les-Miserables (Fig. 6a) and Dolphins datasets, where the largest ARI differences were observed. Specifically, we add random noise to the existing set of connections to understand if the performance gap can also be observed in more challenging scenarios. We add noise to the real data in two different ways (more details can be found in the Supplementary Material):

1. *Flipping entries:* from a given network, we generate a new one by flipping R entries of its adjacency matrix A uniformly at random. This means that if $A_{ij} = 1$, this is changed to $A_{ij} = 0$, and vice versa. The flipping of an entry A_{ij} occurs with probability $p = 0.1$.
2. *Removing intra-community edges:* from a given network, we build a new one with the same inter-community structure but modified intra-community one by removing R within-community edges uniformly at random, based on the ground truth communities. To avoid generating disconnected networks, we only sample edges that are not connected to any leaves.

Both of these procedures make inference harder, but they act differently. The first process is meant to add random noise independently of the community structure (flips are made uniformly at random), while the second aims at targeting the community structure by weakening the assortative structure. We choose R to be $r \times |V|^2$ for the first test and $r \times |E|$ for the second, where we vary $r \in [0, 1]$ to study the impact of these parameters on inferring the communities as measured by the ARI score. We generate 20 samples for each of these two mechanisms built using the Les-Miserables dataset ($|V| = 77, |E| = 254$) shown in Fig. 6a.

We show the results obtained in the test of removing intra-community edges as scatter plots in Fig. 7. We use these plots to compare the algorithms on trial-by-trial community detection tasks: a point on each plot is an instance of a semi-synthetic network with (x, y) coordinates being the ARI scores of ORC-Nextrout (y) and either OTDSinkhorn or Infomap (x). If $y > x$, then ORC-Nextrout outperforms the other method in this particular dataset (blue markers), and vice versa if $x < y$. We then compute the percentage of times that ORC-Nextrout outperforms the other (as indicated in the legend). We find that ORC-Nextrout outperforms both OTDSinkhorn and Infomap clearly and consistently across different values of r ranging from $r = 0.01$ ($R \approx 3$) to $r = 0.07$ ($R \approx 20$). In at least 70% of the cases, ORC-Nextrout gives more accurate results than the other two algorithms. This suggests that ORC-Nextrout is more robust against perturbations of the community structure.

Similar patterns are seen in the case where edges are removed at random and in semi-synthetic networks generated from the Dolphins dataset, see Supplementary Material.

Conclusion

Community detection on networks is a relevant and challenging open area of research. Several methods have been proposed to tackle this issue, with no "best algorithm" that fits well for every type of data. We focused here on a recent line of work that exploits principles from Optimal Transport theory combined with the geometric concept of Ollivier-Ricci curvature applied to discrete graphs. Our method is flexible in that it tunes between different transportation regimes to extract the information necessary to compute the OR curvature on edges. On synthetic data, our model is able to identify communities more robustly than other OT-based methods based on the standard Wasserstein distance in the regime where inference is not trivial. On real data, our model shows better or comparable performance in recovering community structure aligned with node metadata compared to other approaches, thanks to the ability to tune the parameter β.

A relevant advantage of OT-based methods is that the number of communities is automatically learned from data, contrarily to other approaches that need this as an input parameter. In this respect, our model has the tendency to overestimate this number, similarly to other OT-based methods. Understanding how to properly incorporate small-size communities into larger ones in a principled and automatic way is an interesting topic for future work. Similarly, it would be interesting to quantify the extent to which various β capture different network topologies. To address this, one could, for instance, use methods to calculate the structural distance between networks[74] and correlate this against the values of the best performing β. Similarly, as our approach allows obtaining different sets of weights on edges, depending on β, it would be interesting to investigate how different values of this parameter impact network properties that are governed by the weight distribution, such as multi-fractality[75].

There are a number of directions in which this model could be extended. Nodes can be connected in more than one way, as in multilayer networks. Our model could be extended by considering a different β for each edge type, as done in[59,76]. Similarly, real networks are often rich in additional information, e.g. attributes on nodes. It would be interesting to incorporate a priori additional information to inform community detection[43,77]. This information can potentially be used to mitigate the problem of overestimation of the number of communities, as explained above. Finally, we have focused here on the flexibility of solving various transportation problems to provide different computations of the OR curvature. Different results could also be obtained by choosing different input mass distributions on nodes' neighborhoods, as done in[33]. It would be interesting to combine these two approaches to reveal further insights of the role that the OR curvature plays in detecting communities in networks.

Methods

Optimal transport formulation. Consider the probability distributions q that take pairs of vertices and also satisfy the constraints $\sum_i q_{ij} = m_j, \sum_j q_{ij} = m_i$. In other words, these are the joint distributions whose marginals are m_i and m_j. We call these distributions *transport plans* between m_i and m_j. The Optimal Transport problem we are interested in is that of finding a transport plan q^* that minimizes the quantity $\sum_{i\sim j} q_{ij} d_{ij}$, where $i \sim j$ means that the nodes i and j are neighbors and d_{ij} is the cost of transporting mass from i to j, e.g. the distance between these two nodes. The quantity $\mathscr{W}_\beta(m_i, m_j, d) := \sum_{i,j} q^*_{ij} d_{ij}$, defined for this optimal q^*, is the *Wasserstein distance* between m_i and m_j.

The dynamical Monge–Kantorovich model. It was recently proved[51,52] that solutions of the optimal transport problem previously stated can be found by turning that problem into a system of differential equations. This section is dedicated to describe this dynamical formulation.

Let $G = (V, E, W)$ be a weighted graph, with N the number of nodes and E the number of edges in G. Let $\mathbf{B}$ be the *signed incidence* matrix of G. Let f^+ and f^- be two N-dimensional discrete distributions such that $\sum_{i\in V} f_i = 0$ for $f = f^+ - f^-$; let $\mu(t) \in \mathbb{R}^E$ and $u(t) \in \mathbb{R}^N$ be two time-dependent functions defined on edges and nodes, respectively. The discrete *Dynamical Monge-Kantorovich model* can be written as:

$$f_i = \sum_e B_{ie} \frac{\mu_e(t)}{w_e} \sum_j B_{ej}\, u_j(t), \tag{3}$$

$$\mu'_e(t) = \left[\frac{\mu_e(t)}{w_e} \left| \sum_j B_{ej}\, u_j(t) \right| \right]^\beta - \mu_e(t), \tag{4}$$

$$\mu_e(0) > 0, \tag{5}$$

where $|\cdot|$ is the absolute value element-wise. Equation (3) corresponds to Kirchhoff's law, Eq. (4) is the discrete dynamics with β a traffic rate controlling the different routing optimization mechanisms; Eq. (5) is the initial distribution for the edge conductivities.

For $\beta = 1$ the dynamical system described by Eqs. (3)–(5) is known to reach a steady state, i.e., the updates of μ_e and u_e converge to stationary functions μ^* and u^* as t increases. The flux function q defined as $q^*_e := \mu^*_e |u^*_i - u^*_j| / w_e$ is the solution of the optimal transport problem presented in the previous section. Notice

that μ and u depend on the chosen traffic rate β, and thus, so does $q = q(\beta)$. Therefore, we can introduce a generalized version of the distance $\mathscr{W}$:

$$\mathscr{W}_\beta(m_i, m_j, w) := \sum_{i,j} q^*_{ij}(\beta)\ w_{ij}.$$

We then redefine the proposed Ollivier-Ricci curvature as:

$$\kappa_\beta(i,j) := 1 - \frac{\mathscr{W}_\beta(m_i, m_j, w)}{d_{ij}}.$$

Probability distributions on neighborhoods. ORC-Nextrout takes in input a graph and a forcing term. While the graph encapsulates the neighborhood information provided by the nodes i and j, the forcing function is related to the distributions that one needs to transport. Analogously to what was proposed by[22], we define this graph to be the weighted *complete bipartite* $B_{ij} = (V_{ij}, E_{ij}, \omega_{ij})$. The weights in ω_{ij} change iteratively based on the curvature. Notice that a bipartite graph must satisfy $\mathscr{N}(i) \cap \mathscr{N}(j) = \varnothing$, which does not hold true if i and j have common neighbors (this is always the case since $i \in \mathscr{N}(j)$). Nonetheless, this condition does not have great repercussions in the solution of the optimal transport problem since the weights corresponding to these edges (of the form (i, i)) are equal to 0. As for the forcing function, we define it to be $f := f^+ - f^- = m_i - m_j$.

Other methods. To evaluate the performance of ORC-Nextrout, we compare with some of the well-established community detection algorithms including: Infomap[17], MULTITENSOR[13] (MT), discrete Ricci flow[22] (OTDSinkhorn), Label propagation[14] and Louvain[50]. We briefly describe each of these algorithms as follows;

- The *Discrete Ricci flow* (here addressed as OTDSinkhorn)[22] is an iterative node clustering algorithm that deforms edge weights as time progresses, by shrinking sparsely traveled links and stretching heavily traveled edges. These edge weights are iteratively updated based on neighborhood transportation Wasserstein costs, similarly to what is proposed in this manuscript. After a predefined number of iterations, heavily traveled links are removed from the graph. Communities are then obtained as the connected components of this modified network.
- MULTITENSOR (MT)[13] is an algorithm to find communities in multilayer networks. It is a probabilistic model with latent variables regulating community structure and runs with a complexity of $O(EK)$ with assortative structure (as we consider here), where K is the number of communities. This model assumes that the nodes inside the communities can belong to multiple groups (mixed-membership). In this implementation we use their validity for single layer networks (a particular case of a multilayer network).
- *Infomap*[17] employs information theoretic approach for community detection. This method uses the map equation to attend patterns of flow on a network. This flow is simulated using random walkers' traversed paths. Based on the theoretic description of these paths, nodes with quick information flow are then clustered into the same groups. The algorithm runs in $O(E)$. In the experiments, we fix the number of random initialization of the random walkers to be equal to 10. The inferred partition is then the one minimizing the entropy.
- *Label propagation*[14] assigns each node to the same community as the majority of its neighbors. Its working principle start by initializing each node with a distinct label and converges when every node has same label as the majority of its neighboring node. The algorithm has a complexity scaling as $O(E)$.
- *Louvain*[50] is a fast algorithm used to find communities on networks by maximizing the modularity of the associated partitions. It consists of two phases. First, it assigns every node on the network into a different community. Then, it aggregates nodes and neighbors based on gains of modularity. This last step is repeated until no further improvement can be achieved.

Data availability

The real datasets analyzed during the current study are available at http://www-personal.umich.edu/~mejn/netdata/. The synthetic data generated are available from the corresponding author upon request.

Code availability

Open source codes and executables are available at https://github.com/Danielaleite/ORC-Nextrout. https://gitlab.com/enrico_facca/dmk_solver.

Received: 17 May 2022; Accepted: 21 September 2022
Published online: 07 October 2022

References

1. Huang, X., Chen, D., Ren, T. & Wang, D. A survey of community detection methods in multilayer networks. *Data Min. Knowl. Disc.* **35**, 1–45 (2021).
2. Newman, M. *Networks* (Oxford University Press, Oxford, 2018).
3. Fortunato, S. Community detection in graphs. *Phys. Rep.* **486**, 75–174 (2010).
4. Fortunato, S. & Hric, D. Community detection in networks: A user guide. *Phys. Rep.* **659**, 1–44 (2016).
5. Weber, M., Jost, J. & Saucan, E. Detecting the coarse geometry of networks. in *NeurIPS 2018 workshop* (2018).
6. Samal, A. *et al.* Comparative analysis of two discretizations of Ricci curvature for complex networks. *Sci. Rep.* **8**, 1–16 (2018).

7. Zhu, J., Wang, B., Wu, B. & Zhang, W. Emotional community detection in social network. *IEICE Trans. Inf. Syst.* **100**, 2515–2525 (2017).
8. Wang, D., Li, J., Xu, K. & Wu, Y. Sentiment community detection: exploring sentiments and relationships in social networks. *Electron. Commer. Res.* **17**, 103–132 (2017).
9. Li, C. & Zhang, Y. A personalized recommendation algorithm based on large-scale real micro-blog data. *Neural Comput. Appl.* **32**, 11245–11252 (2020).
10. Waskiewicz, T. Friend of a friend influence in terrorist social networks. In *Proceedings on the international conference on artificial intelligence (ICAI)*, 1 (The Steering Committee of The World Congress in Computer Science, Computer..., 2012).
11. Pinheiro, C.A.R. Community detection to identify fraud events in telecommunications networks. in *SAS SUGI proceedings: Customer intelligence* (2012).
12. Chen, J., Zhang, H., Guan, Z.-H. & Li, T. Epidemic spreading on networks with overlapping community structure. *Physica A* **391**, 1848–1854 (2012).
13. De Bacco, C., Power, E. A., Larremore, D. B. & Moore, C. Community detection, link prediction, and layer interdependence in multilayer networks. *Phys. Rev. E* **95**, 042317 (2017).
14. Raghavan, U. N., Albert, R. & Kumara, S. Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E* **76**, 036106 (2007).
15. Clauset, A., Newman, M. E. & Moore, C. Finding community structure in very large networks. *Phys. Rev. E* **70**, 066111 (2004).
16. Reichardt, J. & Bornholdt, S. Statistical mechanics of community detection. *Phys. Rev. E* **74**, 016110 (2006).
17. Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci.* **105**, 1118–1123 (2008).
18. Yang, R., Sala, F. & Bogdan, P. Hidden network generating rules from partially observed complex networks. *Commun. Phys.* **4**, 1–12 (2021).
19. Jost, J. & Liu, S. Ollivier's Ricci curvature, local clustering and curvature-dimension inequalities on graphs. *Discrete Comput. Geom.* **51**, 300–322 (2014).
20. Weber, M., Jost, J. & Saucan, E. Forman–Ricci flow for change detection in large dynamic data sets. *Axioms* **5**, 26 (2016).
21. Weber, M., Saucan, E. & Jost, J. Characterizing complex networks with Forman–Ricci curvature and associated geometric flows. *J. Complex Netw.* **5**, 527–550 (2017).
22. Ni, C.-C., Lin, Y.-Y., Luo, F. & Gao, J. Community detection on networks with Ricci flow. *Sci. Rep.* **9**, 1–12 (2019).
23. Sia, J., Jonckheere, E. & Bogdan, P. Ollivier–Ricci curvature-based method to community detection in complex networks. *Sci. Rep.* **9**, 1–12 (2019).
24. Santambrogio, F. Optimal channel networks, landscape function and branched transport. *Interfaces Free Bound.* **9**, 149–169 (2007).
25. Facca, E., Cardin, F. & Putti, M. Branching structures emerging from a continuous optimal transport model. *J. Comput. Phys.* **447**, 110700 (2021).
26. Brasco, L., Carlier, G. & Santambrogio, F. Congested traffic dynamics, weak flows and very degenerate elliptic equations. *J. Math. Pures Appl.* **93**, 652–671 (2010).
27. Baptista, D., Leite, D., Facca, E., Putti, M. & De Bacco, C. Network extraction by routing optimization. *Sci. Rep.* **10**, 1–13 (2020).
28. Facca, E., Karrenbauer, A., Kolev, P. & Mehlhorn, K. Convergence of the non-uniform directed Physarum model. *Theoret. Comput. Sci.* **816**, 184–194 (2020).
29. Sandhu, R. *et al.* Graph curvature for differentiating cancer networks. *Sci. Rep.* **5**, 1–13 (2015).
30. Wang, C., Jonckheere, E. & Banirazi, R. Interference constrained network control based on curvature. In *2016 American control conference (ACC)* 6036–6041 (IEEE, 2016).
31. Ni, C.-C., Lin, Y.-Y., Gao, J., Gu, X. D. & Saucan, E. Ricci curvature of the internet topology. In *2015 IEEE conference on computer communications (INFOCOM)* 2758–2766 (IEEE, 2015).
32. Devriendt, K. & Lambiotte, R. Discrete curvature on graphs from the effective resistance. *J. Phys. Complex.* (2022).
33. Gosztolai, A. & Arnaudon, A. Unfolding the multiscale structure of networks with dynamical Ollivier–Ricci curvature. *Nat. Commun.* **12**, 1–11 (2021).
34. Paulin, D. Mixing and concentration by Ricci curvature. *J. Funct. Anal.* **270**, 1623–1662 (2016).
35. Veysseire, L. Coarse Ricci curvature for continuous-time Markov processes. arXiv:1202.0420 (2012)
36. Ollivier, Y. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.* **256**, 810–864 (2009).
37. Ollivier, Y. A survey of Ricci curvature for metric spaces and Markov chains. In *Probabilistic approach to geometry* 343–381 (Mathematical Society of Japan, 2010)
38. Forman, R. Bochner's method for cell complexes and combinatorial Ricci curvature. *Discret. Comput. Geom.* **29**, 323–374 (2003).
39. Ni, C.-C., Lin, Y.-Y., Gao, J. & Gu, X. Network alignment by discrete Ollivier–Ricci flow. In *International symposium on graph drawing and network visualization*,447–462 (Springer, 2018).
40. Ye, Z., Liu, K. S., Ma, T., Gao, J. & Chen, C. Curvature graph network. In *International conference on learning representations* (2019)
41. Lancichinetti, A. & Fortunato, S. Community detection algorithms: A comparative analysis. *Phys. Rev. E* **80**, 056117 (2009).
42. Decelle, A., Krzakala, F., Moore, C. & Zdeborová, L. Inference and phase transitions in the detection of modules in sparse networks. *Phys. Rev. Lett.* **107**, 065701 (2011).
43. Contisciani, M., Power, E. A. & De Bacco, C. Community detection with node attributes in multilayer networks. *Sci. Rep.* **10**, 1–16 (2020).
44. Ball, B., Karrer, B. & Newman, M. E. Efficient and principled method for detecting communities in networks. *Phys. Rev. E* **84**, 036103 (2011).
45. Peixoto, T.P. Bayesian stochastic blockmodeling. *Adv. Netw. Clust. Blockmodeling* 289–332 (2019)
46. Yang, J. & Leskovec, J. Overlapping community detection at scale: A nonnegative matrix factorization approach. In *Proceedings of the sixth ACM international conference on Web search and data mining* 587–596 (2013).
47. Newman, M. E. Spectral methods for community detection and graph partitioning. *Phys. Rev. E* **88**, 042822 (2013).
48. Sussman, D. L., Tang, M., Fishkind, D. E. & Priebe, C. E. A consistent adjacency spectral embedding for stochastic blockmodel graphs. *J. Am. Stat. Assoc.* **107**, 1119–1128 (2012).
49. Girvan, M. & Newman, M. E. Community structure in social and biological networks. *Proc. Natl. Acad. Sci.* **99**, 7821–7826 (2002).
50. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).
51. Facca, E., Cardin, F. & Putti, M. Towards a stationary Monge–Kantorovich dynamics: The physarum polycephalum experience. *SIAM J. Appl. Math.* **78**, 651–676 (2018).
52. Facca, E., Daneri, S., Cardin, F. & Putti, M. Numerical solution of Monge–Kantorovich equations via a dynamic formulation. *J. Sci. Comput.* **82**, 1–26 (2020).
53. Xia, Q. Optimal paths related to transport problems. *Commun. Contemp. Math.* **5**, 251–279 (2003).
54. Baptista, D. & De Bacco, C. Principled network extraction from images. *R. Soc. Open Sci.* **8**, 210025 (2021).
55. Baptista, D. & De Bacco, C. Convergence properties of optimal transport-based temporal networks. In *International conference on complex networks and their applications* 578–592 (Springer, 2021)
56. Lonardi, A., Facca, E., Putti, M. & De Bacco, C. Designing optimal networks for multicommodity transport problem. *Phys. Rev. Res.* **3**, 043010 (2021).

57. Lonardi, A., Putti, M. & De Bacco, C. Multicommodity routing optimization for engineering networks. *Sci. Rep.* **12**, 1–11 (2022).
58. Lonardi, A., Facca, E., Putti, M. & De Bacco, C. Infrastructure adaptation and emergence of loops in network routing with time-dependent loads. arXiv:2112.10620 (2021)
59. Ibrahim, A. A., Lonardi, A. & Bacco, C. D. Optimal transport in multilayer networks for traffic flow optimization. *Algorithms* **14**, 189 (2021).
60. Fredman, M. L. & Tarjan, R. E. Fibonacci heaps and their uses in improved network optimization algorithms. *J. ACM (JACM)* **34**, 596–615 (1987).
61. Lancichinetti, A., Fortunato, S. & Radicchi, F. Benchmark graphs for testing community detection algorithms. *Phys. Rev. E* **78**, 046110 (2008).
62. Holland, P. W., Laskey, K. B. & Leinhardt, S. Stochastic blockmodels: First steps. *Soc. Netw.* **5**, 109–137 (1983).
63. Hubert, L. & Arabie, P. Comparing partitions. *J. Classif.* **2**, 193–218 (1985).
64. Facca, E. & Benzi, M. Fast iterative solution of the optimal transport problem on graphs. *SIAM J. Sci. Comput.* **43**, A2295–A2319 (2021).
65. Cuturi, M. Sinkhorn distances: Lightspeed computation of optimal transport. *Adv. Neural Inf. Process. Syst.* **26** (2013)
66. Flamary, R. *et al.* Pot: Python optimal transport. *J. Mach. Learn. Res.* **22**, 1–8 (2021).
67. Network data. http://www-personal.umich.edu/~mejn/netdata/
68. Peel, L., Larremore, D. B. & Clauset, A. The ground truth about metadata and community detection in networks. *Sci. Adv.* **3**, e1602548 (2017).
69. Knuth, D. E. *The Stanford GraphBase: A platform for combinatorial computing* Vol. 1 (AcM Press, New York, 1993).
70. Lusseau, D. *et al.* The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behav. Ecol. Sociobiol.* **54**, 396–405 (2003).
71. Lusseau, D. & Newman, M. E. Identifying the role that animals play in their social networks. *Proc R Soc Lond Ser B Biol Sci* **271**, S477–S481 (2004).
72. Books about us politics dataset. http://www.orgnet.com/.
73. Fruchterman, T. M. & Reingold, E. M. Graph drawing by force-directed placement. *Softw. Pract. Exp.* **21**, 1129–1164 (1991).
74. Xiao, X., Chen, H. & Bogdan, P. Deciphering the generating rules and functionalities of complex networks. *Sci. Rep.* **11**, 1–15 (2021).
75. Xue, Y. & Bogdan, P. Reliable multi-fractal characterization of weighted complex networks: Algorithms and implications. *Sci. Rep.* **7**, 1–22 (2017).
76. Ibrahim, A. A., Leite, D. & De Bacco, C. Sustainable optimal transport in multilayer networks. *Phys. Rev. E* **105**, 064302 (2022).
77. Newman, M. E. & Clauset, A. Structure and inference in annotated networks. *Nat. Commun.* **7**, 11863 (2016).

Acknowledgements
The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting Daniela Leite, Diego Baptista and Abdullahi A. Ibrahim.

Author contributions
All authors contributed to developing the models, conceived the experiments, analyzing the results and reviewing the manuscript. D.L., D.B. and A.A.I. conducted the experiments.

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/s41598-022-20986-y.

Correspondence and requests for materials should be addressed to D.L.

Reprints and permissions information is available at 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

Supplementary information for
Community Detection in networks by Dynamical Optimal Transport Formulation

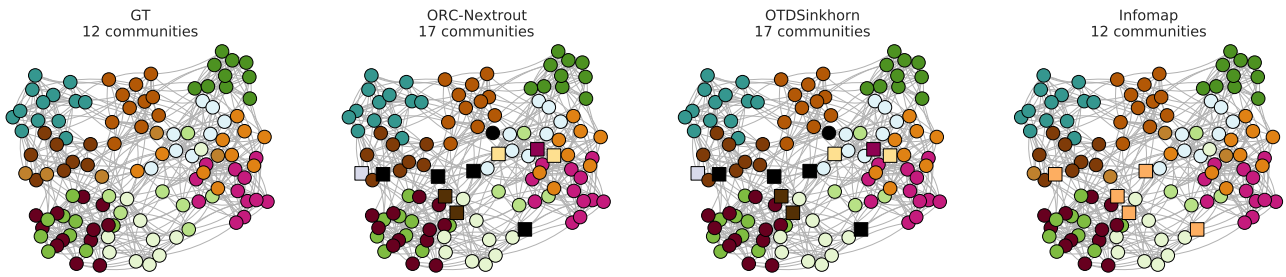
Daniela Leite[1,*,+], Diego Baptista[1,*], Abdullahi A. Ibrahim[1], Enrico Facca[2], and Caterina De Bacco[1]

[1]Max Planck Institute for Intelligent Systems, Cyber Valley, 72076 Tübingen, Germany
[2]Univ. Lille, Inria, CNRS, UMR 8524 - Laboratoire Paul Painlevé, F-59000 Lille, France
[+]daniela.leite@tuebingen.mpg.de
[*]These authors contributed equally to this work

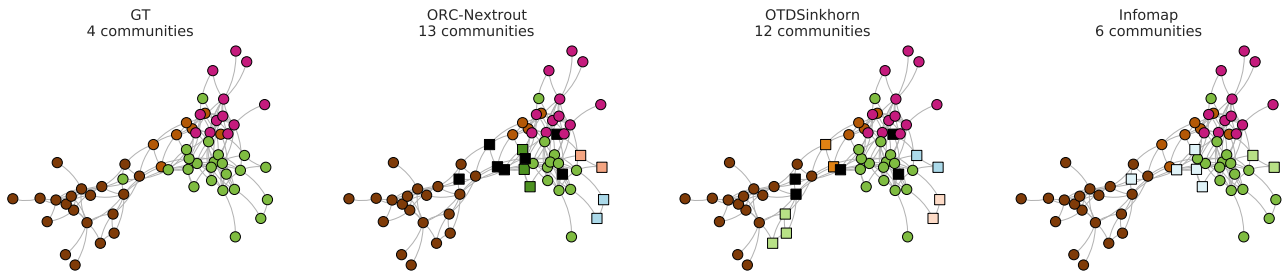
Supplementary Note 1: Real networks

This section presents additional plots for real networks on communities detected by some algorithms studied in this paper. Supplementary Fig. 1a and Supplementary Fig. 1b exhibit the detected communities for American football and Political books, respectively. For each dataset we compare three algorithms with the ground truth, and with ORC-Nextrout having a single setting of β.

(a) American Football



(b) Dolphins



Supplementary Figure 1. Communities in real networks. We show the communities for American football (a) and Dolphins (b) inferred by ORC-Nextrout ($\beta = 2.0, 1.5$ for top and bottom rows respectively), OTDSinkhorn and Infomap and compare against those extracted using node attributes (GT). Dark nodes represent individual nodes who are assigned to isolated communities by OT-based methods. Square-shaped markers denote nodes assigned to communities different than those obtained from node metadata.

Supplementary Note 2: ARI score

We dedicate this section to present the main score used in the experiments described in this manuscript.
The *Rand Index*[1] (RI) is a measure of the similarity between two sets of clusters. Let $S = \{x_1, x_2, ..., x_n\}$ be a set of n different elements, and let $X = \{X_1, X_2, .., X_m\}$, $\hat{X} = \{\hat{X}_1, \hat{X}_2, .., \hat{X}_p\}$, be two partitions of S into m and p subsets, respectively. Intuitively,

X can thought of as the "correct" separation of S into classes, while $\hat{X}$ would be a "prediction" of it. We would like to understand the quality of the prediction $\hat{X}$ in terms of the ground truth community information X. Consider

- TP (*true positive*) is the number of times that a pair of elements (x_i, x_j) belonging to the **same** class X_k gets assigned to the **same** class $\hat{X}_l$.
- TN (*true negative*) is the number of times that a pair of elements (x_i, x_j) belonging to **different** classses in X gets assigned to **different** classes in $\hat{X}$.
- FP (*false positive*) is the number of times that a pair of elements (x_i, x_j) belonging to **same** class X_k is (*falsely*) assigned to **different** classes $\hat{X}_l$ and $\hat{X}_m$.
- FN (*false negative*) is the number of times that a pair of elements (x_i, x_j) belonging to **different** classes X_k and X_l is (*falsely*) assigned to the **same** class $\hat{X}_m$.

One can think of the the words *positive* and *negative* referring to whether two elements in S belong to the same or to different classes, respectively. The words *true* and *false* would then judge the performance of the prediction: if it maches the nature of the elements under inspection, then the word *true* is associated to it; *negative*, otherwise.

The Rand Index is then computed using the formula:

$$RI(X, \hat{X}) = \frac{TP + TN}{TP + FP + FN + TN}.$$

The *Adjusted Rand Index* (ARI) is the *corrected for chance* version of the RI:

$$ARI(X, \hat{X}) = \frac{RI - \mathbb{E}[RI]}{1 - \mathbb{E}[RI]},$$

where $\mathbb{E}[RI]$ is the expected value of the RI under the assumptions that the partitions X and $\hat{X}$ are sampled from the generalized hypergeometric distribution. Closed forms for the terms shown in the ARI formula can be computed. See[1] for a more detailed presentation of the RI and ARI scores.

Supplementary Note 3: Two tests on random networks generated from real structures

Algorithm and adjacency matrix

The pseudo-code of the random processes defined on Section ***Two tests on semi-synthetic networks*** of the manuscript is shown in the Algorithms 1 and 2.

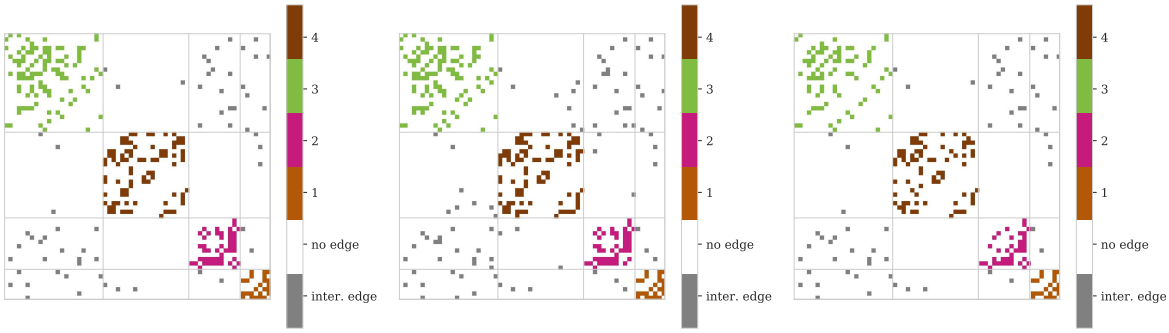
Algorithm 1 Flipping entries of the adjacency matrix

```
Input: G = (V, E, W), flipping proportion r ∈ [0, 1], flipping probability p ∈ [0, 1]
Output: G' = (V, E')
Build adjacency matrix of G: A
Make a copy of A: A'
Compute number of nodes in G: N
for t ∈ range(r * N^2) do
  for i ∈ range(N) do
    for j ∈ range(i + 1, N) do
      Assume P_ij(A[i][j]) = p
      Sample A'[i][j] from P
    end for
  end for
end for
Symmetrize A'
Build G' using the connections between nodes described by A'.
```

We show two examples of the outputs of these algorithms on Supplementary Figure 2 together with the adjacency matrix that was used to originally build them. Notice that the matrix in panel (b) is different from that in panel (a) both in terms of intra and inter-community blocks, whereas that in (c) only differs along the within-community entries. This indicates that the first method alters the overall configuration of the edges in the network by adding random noise, whereas the second changes the original network only by reducing the intra-community relationships.

Algorithm 2 Removing intra-community edges

Input: $G = (V, E, W)$, removal proportion $r \in [0, 1]$,
Output: $G' = (V, E')$
Build list of edges of G: $E(G)$
Make a copy of E: $E'(G)$
Compute number of edges in G: M
Remove from $E'(G)$ all the edges $e = (i, j)$ such that either i or j is a leaf in G
Remove $r * M$ elements from $E'(G)$ uniformly at random
Define G' using V and $E'(G)$



(a) Original adjacency matrix **(b)** Flipping entries **(c)** Removing intra-community edges

Supplementary Figure 2. Adjacency matrices for original and perturbed Dolphins network. We show the adjacency matrix of (a) the original dataset, (ii) a perturbed network built from the previous one by flipping entries at random and (c) a perturbed graph obtained by removing intra-community edges. We used $r = 0.05$ and $r = 0.15$ to generate the matrices shown in (b) and (c), respectively. Nodes are grouped by communities to highlight the block structure. Gray lines denote the boundaries of these blocks. Diagonal blocks represent the communities. Off-diagonal blocks show connections between communities. Colored entries are in agreement with those of community layout shown in Supplementary Fig. 1b. Inter-community connections are highlighted in gray.

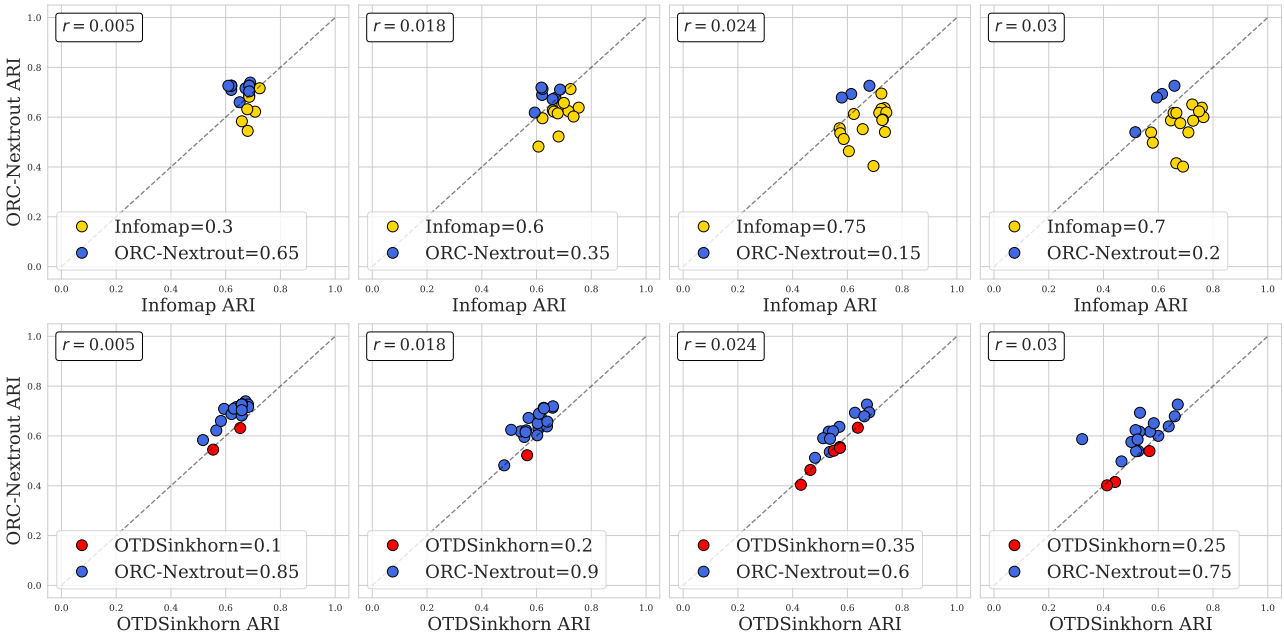
Results

We show results obtained on the Les-Miserables dataset for the test where we flip at random the entries of the adjacency matrix A (see Supplementary Figure 4). In this case, ORC-Nextrout outperforms Infomap only in the case where $r = 0.005$, i.e. when 5% of elements of A are changed. As r increases, Infomap increases its accuracy. On the other hand, ORC-Nextrout shows a better performance than OTDSinkhorn consistently across values of r.

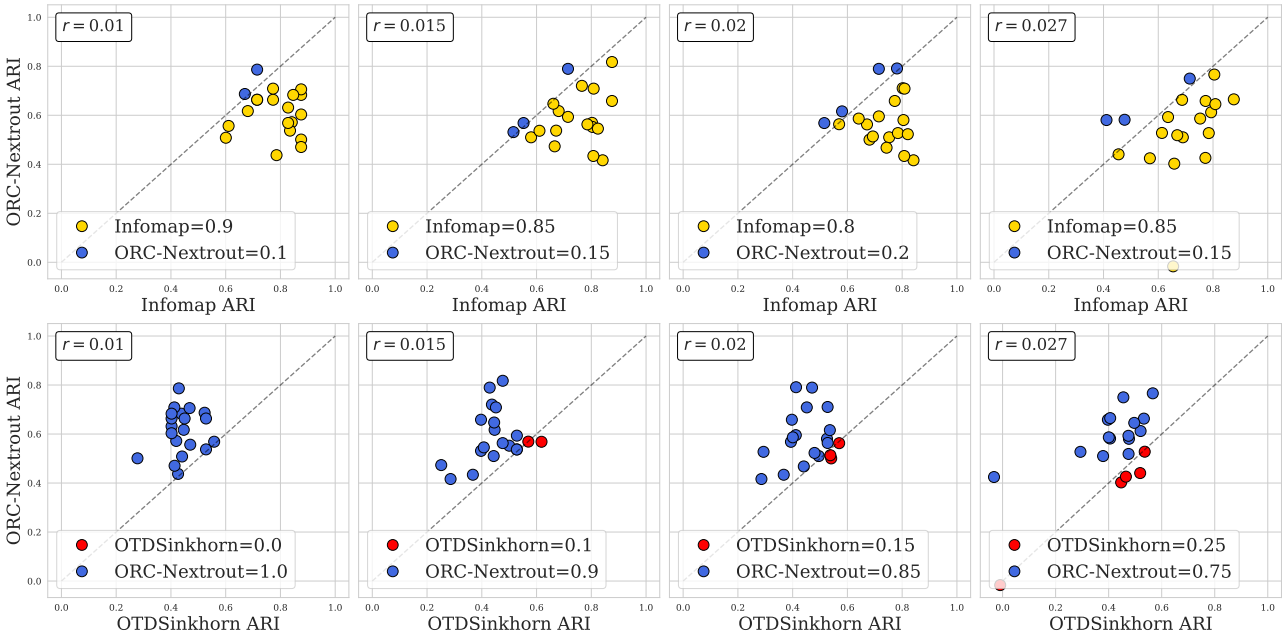
Lastly, in Supplementary Figure 4. we show the results obtained for both tests on the Dolphins dataset. It can be seen that ORC-Nextrout outperforms OTDSinkhorn consistently across values of r in both cases. On the other hand, Infomap has a higher accuracy in both tests, as expected given the results shown in Figure 5 of the main manuscript.

References

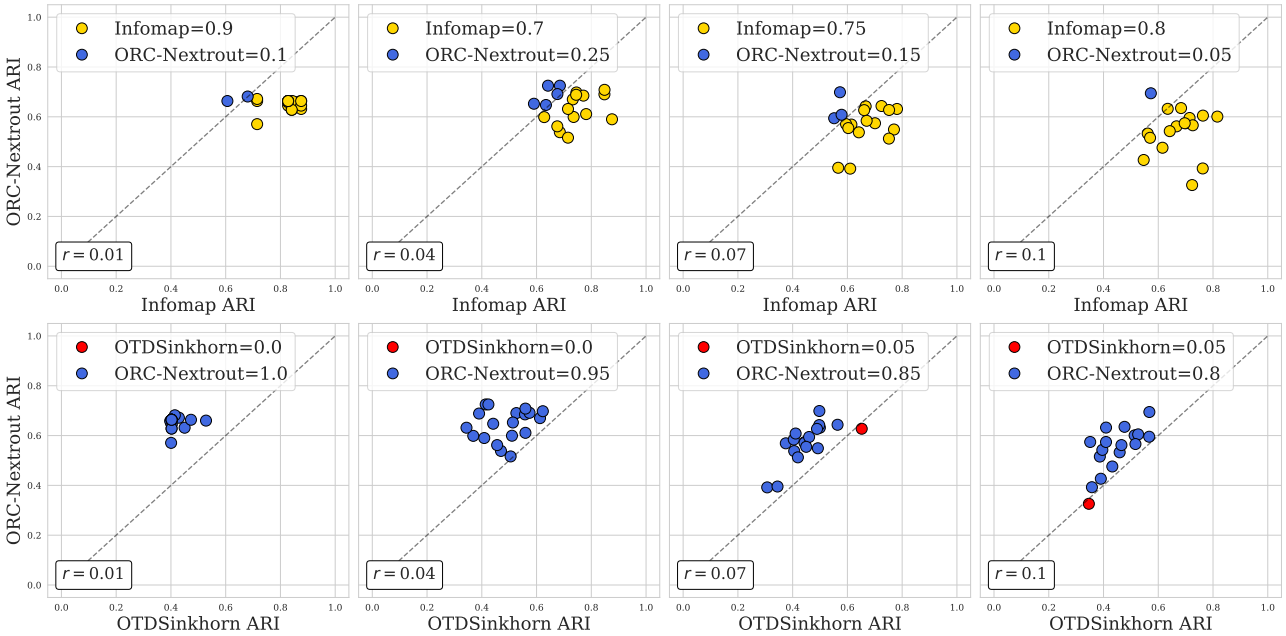
1. Hubert, L. & Arabie, P. Comparing partitions. *J. classification* **2**, 193–218 (1985).



Supplementary Figure 3. Flipping-entries test on Les Miserables data. Markers correspond to 20 instances of semi-synthetic networks generated from real data. Their (x, y) coordinates are the ARI scores of the method indicated on axes. Colors are given by the best performing algorithm, e.g. if $x > y$, the color of the method associated to x is chosen. The legend shows the percentage of times that the corresponding method outperforms the other. The parameter r describes the proportion of entries of the adjacency matrix A that have been changed. This increases from left to right.



Supplementary Figure 4. Flipping-entries test on Dolphins data. Markers' description is similar as in Supplementary Figure 3.



Supplementary Figure 5. Removing intra-community edges test on Dolphins data. Markers' description is similar as in Supplementary Figure 3.