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

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.

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

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 $dg_i(\mu(t))/dt \geq -\alpha g_i(\mu(t))$, 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
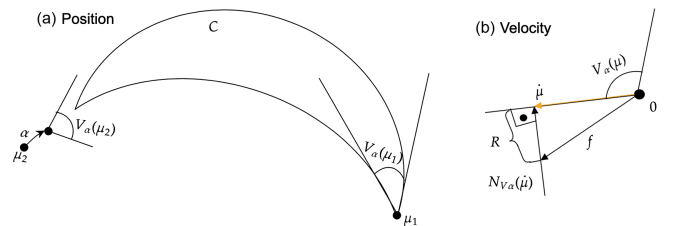

FIG. 1. (a) Visualization of the set C and the set of feasible velocities $V_\alpha(\mu_1)$ and $V_\alpha(\mu_2)$ at points μ_1 and μ_2, respectively. Point μ_1 lies on the boundary of C, while μ_2 is infeasible; α is a restitution parameter. (b) When the vector field f is pushing away from C, a force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

that forces the solution to remain within the desired constraints:

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

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:

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

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

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

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:

$$\mu^{t+1} = \mu^t + \tau \dot{\mu}, \qquad (5)$$

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

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

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
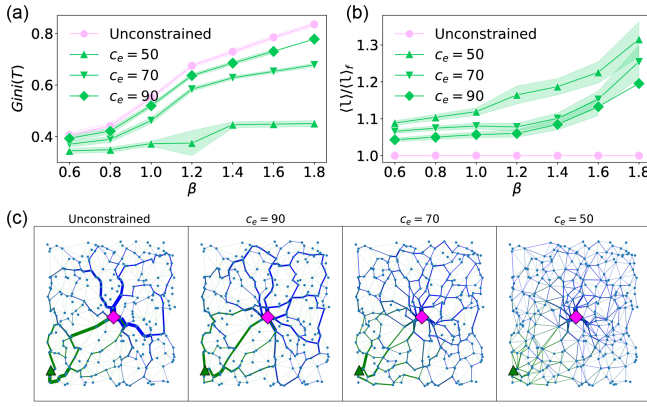
FIG. 2. Capacity constraint on synthetic networks. (a) Gini coefficient of the traffic distribution on edges. The edge capacity $c_e = c$ is selected as a percentile of the distribution of μ over edges obtained in the unconstrained case (Unconstrained). (b) Ratio of average total path length to that of Unconstrained, $\langle 1 \rangle_f$. 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

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

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
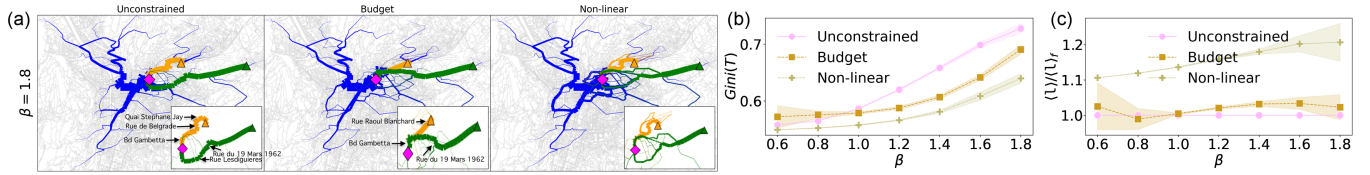
FIG. 3.   Constrained OT on Grenoble road network. (a) Path trajectories for the unconstrained OT (Unconstrained), budget constraint (Budget), and a nonlinear budget plus capacity (Nonlinear). We set $b = \frac{1}{2} \sum_e \mu_e$, where μ_e is that of unconstrained, $\delta = 1/2$ and $c_e = 70$ for 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].

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 $C \coloneqq \{\mu \in \mathbb{R}^E_{\geq 0} | g(\mu) \geq 0\}$ 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 $I_\mu \coloneqq \{i \in \mathbb{Z} | g_i(\mu) \leq 0\}$. 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 $V_\alpha(\mu) \coloneqq \{v \in \mathbb{R}^E | \nabla g_i(\mu)^T v \geq -\alpha g_i(\mu), i \in I_\mu\}$, 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:

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

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:

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

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:

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

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:

$$\dot{\mu}_e = \begin{cases} f_e - S_e\lambda_b & \text{if } f_e - S_e\lambda_b \geq -\alpha\,\mu_e \\[2mm] -\alpha\,\mu_e & \text{if } f_e - S_e\lambda_b < -\alpha\,\mu_e \quad . \end{cases} \tag{S15}$$

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:

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

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

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

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

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

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

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

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

$$\dot{\mu}_e = \begin{cases} \alpha (c_e - \mu_e) & \text{if } f_e - S_e \lambda_\delta h_e \geq \alpha (c_e - \mu_e) \, , \quad c_e \leq \mu_e \quad , \\[2mm] -\alpha \mu_e & \text{if } f_e - S_e \lambda_\delta h_e < -\alpha \mu_e, \quad \mu_e \leq 0 \quad , \\[2mm] f_e - S_e \lambda_\delta h_e & \text{otherwise} \quad . \end{cases} \tag{S30}$$

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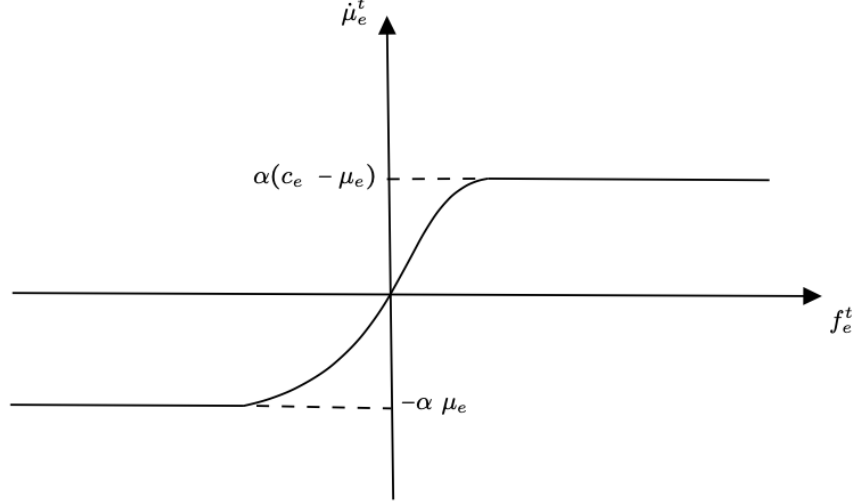
FIG. S1. This plot shows $\dot{\mu}_e^t$ as a function of f_e^t (typical situation, the function also depends on δ and b). The solution is expected to move at most by $\alpha\,(c_e - \mu_e)$ and at least by $\alpha\,\mu_e$.

<center>Computation of λ_δ</center>

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. $\dot{\mathfrak{L}}_\beta \le 0$. 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 $-\langle \dot{\mu}, \dot{\mu} \rangle = -\dot{\mu}^T S^{-1} \dot{\mu} \le 0$ is valid because it results in a non-positive sum of squares. Thus, the remaining task is to prove that $\langle R, \dot{\mu} \rangle \le 0$. The stationarity condition of Eq. (3) can be expressed as:

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

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 $\langle R, v - \dot{\mu} \rangle \ge 0$ 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 $\langle R, -\dot{\mu} \rangle \ge 0 \implies \langle R, \dot{\mu} \rangle \le 0$. Hence $\dot{\mathfrak{L}}_\beta \le 0$.

To prove claim iii), we have established $\langle R, \dot{\mu} \rangle \le 0$ 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:

$$\dot{\mu} = 0 \iff -S\nabla \mathfrak{L}_\beta \in N_{T_C(\mu)}(0) \quad , \tag{S39}$$

$$\iff \langle -S\nabla \mathfrak{L}_\beta, v \rangle \le 0, \quad \forall v \in T_C(\mu) \quad , \tag{S40}$$

$$\iff \nabla \mathfrak{L}_\beta(\mu)^T v \ge 0, \quad \forall v \in T_C(\mu) \quad . \tag{S41}$$

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:

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

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 $\{p_u^j\}_{u \in V}$, $\forall j = 1, \dots, M$, which in turn determine the fluxes vectors $\{F_e\}_{e \in E}$. Given the μ, we can obtain the pressure potentials as $p_v^j = \sum_u (L^j)_{vu}^\dagger q_u^j$, 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

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

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

14:     Update the dynamics: $\mu^{t+1} \leftarrow \mu^t + \tau\dot{\mu}^t, \quad t \leftarrow t + 1$
15: **end while**
16: **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.
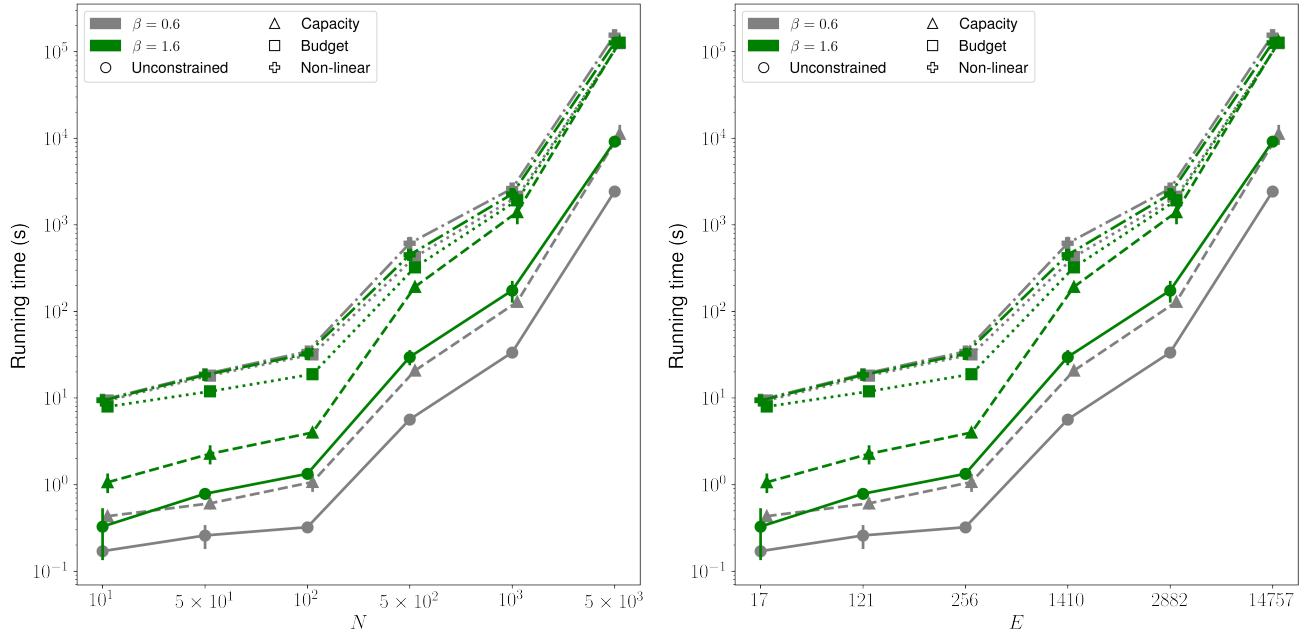

FIG. S2. Running time as a function of N and E. We show two scenarios of β, i.e., penalizes traffic congestion ($\beta < 1$) and encourages path consolidation ($\beta > 1$). Markers and error bar are averages and standard deviations over 10 network realizations. Other settings: $\delta = 1/2$, $c_e = 75$, 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.
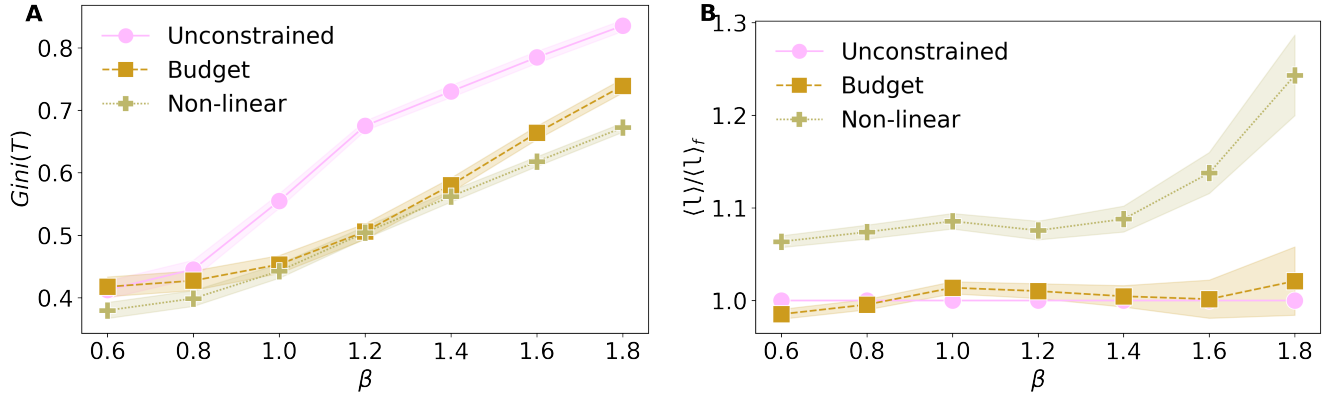
FIG. S3. Results over varying β on synthetic networks. (A) Gini coefficient of the traffic distribution on edges. (B) Ratio of average total path length to that of the unconstrained OT method, denoted as $\langle \mathtt{l} \rangle_f$. Markers and shadows are averages and standard deviations over 20 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.
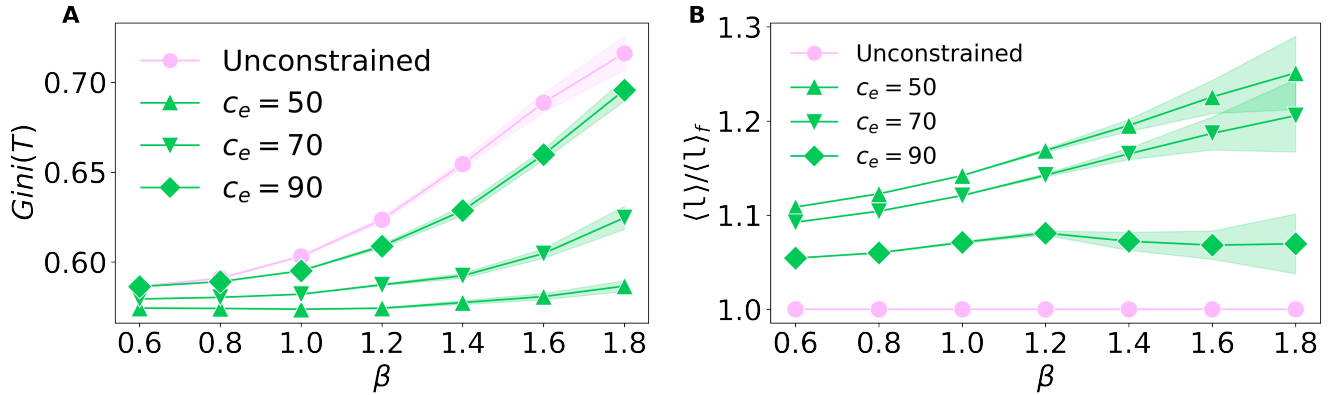

FIG. S4. Results on Grenoble bus network. (A) Gini coefficient of the traffic distribution on the network edges. The edge capacity c_e is the percentile of μ from f, and varied between low, medium and high capacities. Varying c_e helps to understand how the size of highway impacts the traffic. Setting 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.
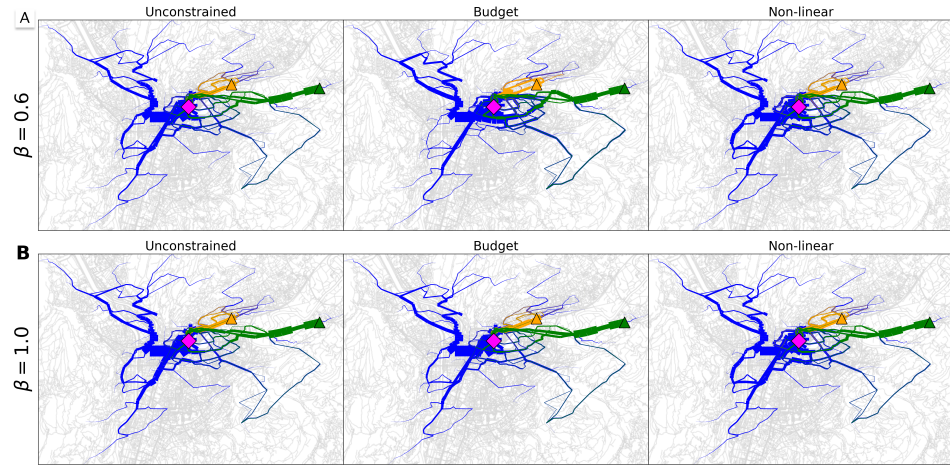
FIG. S5. Constrained OT on Grenoble road network. (A-B) Path trajectories of the Grenoble road network for the unconstrained OT (Unconstrained), using a budget constraint (Budget); a capacity constraint and a non-linear budget (Non-linear). Other settings are the same as in Fig. 3.
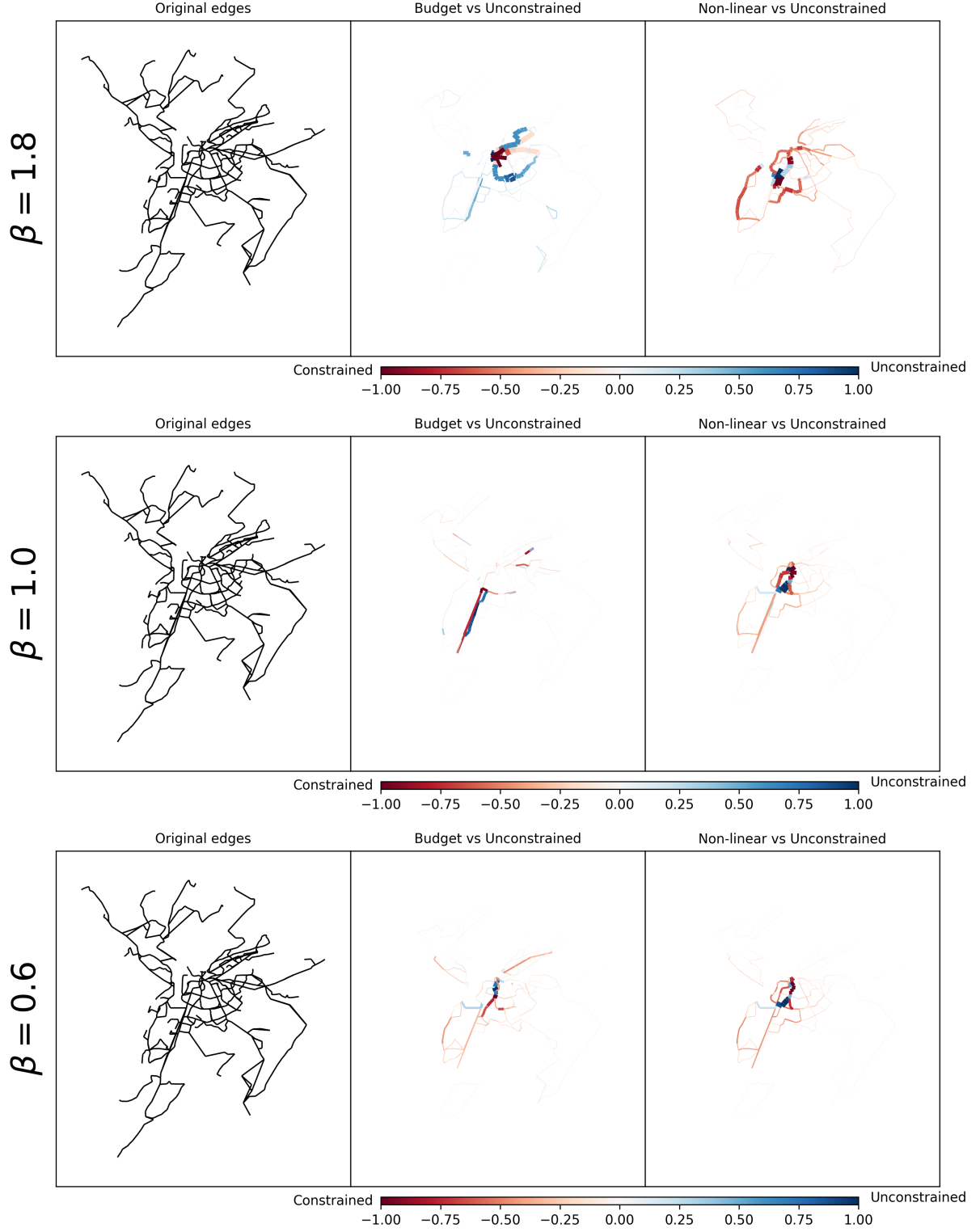
FIG. S6. Edge-wise difference only plots on Grenoble bus network. We compute the difference of traffic on edges as $T_e^c - T_e^u$ where T_e^c and T_e^u denote the traffic for constrained and unconstrained methods, respectively. Other settings are the same as in Fig. 3.

[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).