Bilevel Optimization for Traffic Mitigation in Optimal Transport Networks

Alessandro Lonardi [*] and Caterina De Bacco [†]
Max Planck Institute for Intelligent Systems, Cyber Valley, Tübingen 72076, Germany

(Received 28 June 2023; revised 21 September 2023; accepted 22 November 2023; published 26 December 2023)

Global infrastructure robustness and local transport efficiency are critical requirements for transportation networks. However, since passengers often travel greedily to maximize their own benefit and trigger traffic jams, overall transportation performance can be heavily disrupted. We develop adaptation rules that leverage optimal transport theory to effectively route passengers along their shortest paths while also strategically tuning edge weights to optimize traffic. As a result, we enforce both global and local optimality of transport. We prove the efficacy of our approach on synthetic networks and on real data. Our findings on the international European highways suggest that thoughtfully devised routing schemes might help to lower car-produced carbon emissions.

DOI: 10.1103/PhysRevLett.131.267401

Introduction.—Transport networks are ubiquitous in nature and engineering, spanning from living organisms to cities and telecommunications. Many of these systems can be modeled by adaptation rules that follow the principle of minimum energy, regulating edge flows to optimize transportation costs. Examples in biology are plants, whose profiles emerge from a trade-off between minimization of hydraulic resistance and carbon cost [1], and leaves, shaped by the interplay of nutrients' transport efficiency and robustness to damage [2–4].

Similarly, adaptation rules have been employed to model traffic flows in urban transportation by jointly minimizing the energy dissipated by the passengers and the construction cost of the infrastructure [5–11]. While these models set forth a first approach to simulate traffic flows using adaptation, they crucially neglect that passengers in a transportation network do not move cohesively to minimize a unique cost. Instead, they choose their routes greedily to maximize their benefit (Wardrop's first principle) [12–14]. As a consequence, transport networks may be globally inefficient.

In this Letter, we propose a set of adaptation equations to find traffic flows that mitigate congestion, considered as a proxy for global efficiency, while trading off against the shortest routes.

We frame the problem in a bilevel optimization setup, which poses a competition between greedy passengers and a network manager. The passengers minimize their origin-destination path cost seeking for the user equilibrium [15] (lower-level problem), whereas the network manager guarantees global efficiency by mitigating traffic bottlenecks on edges to achieve the system optimum (upper-level problem), while implicitly accounting for passengers' shortest path. We tackle the optimization problem by alternating optimal transport- (OT) inspired adaptation rules for the lower-level optimization and a projected stochastic gradient descent (PSGD) scheme for the upper-level optimization.

In detail, greedy passenger flows are found by solving a dynamical system that governs the evolution of edge capacities, variables that control passenger allocation, so that these travel on their shortest paths. Adaptation rules are a well-established mechanism for route assignment on networks [3,5–11,16–21] and in continuous domains [22–26]. Classically, user equilibrium greedy flows can be found with the Frank-Wolfe algorithm [27] or, alternatively, with recent methods accounting for passengers' travel budgeting [28]. Here, we propose a model that exploits OT theory to prove that, at convergence, passengers move along the shortest path. Particularly, our dynamical system admits a Lyapunov functional [24] that asymptotically converges to the shortest path (Wasserstein) distance between entry and exit distributions of passengers [16,17,29].

Traffic mitigation is performed by minimizing a quadratic loss function that penalizes edges whose traffic exceeds a prefixed threshold. The minimization problem can be treated analytically by assuming that the network edges are endowed with capacities and weights (resistances) and their flows are the gradient of a scalar potential, as for electrical networks. We derive closed-form gradients for the weights, which can be interpreted as the cost that passengers pay for traveling. In practice, network managers would implement these weights by strategically designing incentives or disincentives, e.g., assigning road tolls, to encourage passengers to relocate from jammed edges. The task of traffic mitigation has been addressed using several methods. These include belief propagation [30–32], adaptive dynamical networks [33], Markov chain Monte Carlo schemes [34], cellar automata [35,36], and heuristic routing models [37].

A bilevel optimization problem similar to the one studied here was solved using message passing [38]. While the problem's setting is similar to ours, the methodologies differ since we alternate adaptation rules for the capacities with global descent for the weights, whereas message passing uses local updates for flows. Our approach outputs individual passengers' optimal paths, whereas the formulation in Li *et al.* [38] can only extract aggregate routes.

We find that our method effectively trades off traffic mitigation against the shortest passenger routes. Namely, both on synthetic topologies and real roads, it returns optimal transport networks where congestion is heavily reduced. We argue that this result is beneficial for reducing the carbon footprint of roads. We also show that the uncoordinated actions of the network manager and passengers can be counterproductive, i.e., they may increase traffic, with an outcome opposite to that intended.

Problem.—We take a network $G(V,E)$ where $M \geq 1$ groups of greedy passengers i can travel from origin nodes O^i (one node per group) to possibly multiple destination nodes D^i. Stationary numbers of entry and exit passengers are stored in a mass matrix with entries $\tilde{S}^i_v > 0$ for each $v = O^i$, $\tilde{S}^i_v < 0$ for $v \in D^i$, and $\tilde{S}^i_v = 0$ otherwise. We assume that the system is isolated, i.e., that passengers entering the network must also exit. This condition is $\sum_v \tilde{S}^i_v = 0$ for all i. When traveling along an edge, passengers pay a cost $\tilde{w}_e > 0$, and finally, each edge is equipped with a capacity that controls the rate at which passengers i are allocated along each edge e; $\tilde{c}^i_e \geq 0$. Intuitively, one could think of capacities as the space occupied by passengers of type i, i.e., larger space accommodates more passengers. All problem variables have been introduced with units, however, these can be nondimensionalized to derive scale-independent adaptation rules (see Supplemental Material [39]). We denote dimensional quantities with a tilde and dimensionless ones without.

Lower-level optimization.—The lower-level problem allows us to find the cheapest routes from O^i to D^i. In order to model traffic flows, we introduce the fluxes F^i_e, specifying the displacement of S^i along an edge e. In analogy with electrical networks, we assume that there exists an auxiliary pressure potential p^i_v on each node v due to index i. We interpret them as the travel demand from passengers traveling from v. With this, we define the potential-based fluxes for all $e = (u,v)$ and i, i.e., Poiseuille's law, as

$$F^i_e = \frac{c^i_e}{w_e}(p^i_u - p^i_v). \tag{1}$$

Fluxes must obey Kirchhoff's law. We can write it as $\sum_e B_{ve} F^i_e = S^i_v$, where B is a conventionally oriented incidence matrix of the network. Substituting Eq. (1) in Kirchhoff's law, the potential becomes a function of c and w, namely, $p^i_v = \sum_u (L^{i^\dagger})_{vu} S^i_u$, where $\dagger$ denotes the Moore-Penrose inverse and $L^i_{uv} = \sum_e (c^i_e/w_e) B_{ue} B_{ve}$ are entries of the network weighted Laplacian. With this substitution, $F \equiv F(c,w)$ is also a function of only c and w, the only independent problem's variables.

For any fixed set of weights, we write the lower-level problem as

$$J(c,w) = \sum_{ei} w_e |F^i_e|, \tag{2}$$

$$\min_{c\geq 0} J(c,w). \tag{3}$$

The convex OT cost J in Eq. (2) is the sum over M indexes of the w shortest path costs $J^i = \sum_e w_e |F^i_e|$ [16,17]. Its only minimizer is the overlap of M shortest paths from all O^i to D^i, which are found with c using Eq. (1) and Kirchhoff's law.

Upper-level optimization.—The upper-level problem formalizes the task of the network manager of tuning w to mitigate traffic jams triggered by the passengers. We measure traffic by penalizing congested links where $\sum_i |F^i_e|$ exceeds a threshold $\theta \geq 0$, above which infrastructural failures may occur. Conveniently, we introduce $\Delta_e = \sum_i |F^i_e| - \theta$.

Analogously to Eqs. (2) and (3), for any set of capacities, the upper-level optimization is

$$\Omega(c,w) = \frac{1}{2}\sum_e \Delta_e^2 H(\Delta_e), \tag{4}$$

$$\min_{w\geq\epsilon} \Omega(c,w), \tag{5}$$

where H is the Heaviside step function. In Eq. (4), other objective functions, e.g., the hinge loss, can be utilized [38,46]; we do not explore this here. Furthermore, the weights are constrained to be larger than a small $\epsilon > 0$. This means that passengers cannot profit ($w < 0$) or travel for free ($w = 0$). Practically, this ensures that the Laplacian L is well defined.

Bilevel optimization.—We combine the two optimization problems into one. Suppose that the network manager is regularly informed of the passengers' routes and, using such information, the weights are tuned to mitigate traffic. After each update, passengers reroute according to the updated weights.

Formally, this translates into the problem

$$\min_{w\geq\epsilon} \Omega(w;\hat{c}), \tag{6}$$

$$\text{such that } \hat{c} = \operatorname*{argmin}_{c\geq 0} J(c;w), \tag{7}$$

where the equality in Eq. (7) comes from the convexity of J [16,17]. In Eq. (6) we explicitly state the dependence on w as a variable and on c as a parameter [conversely for Eq. (7)].

Optimal transport dynamics.—To find the shortest paths required for the lower-level problem, we couple fluxes and capacities with the ordinary differential equations

$$\frac{dc_e^i}{dt} = \frac{F_e^{i2}}{c_e^i} - c_e^i, \tag{8}$$

where fluxes obey Kirchhoff's law. In Eq. (8), edges with high flux enlarge, whereas those where the negative decaying term prevails shrink. Crucially, asymptotic solutions converge to the minimum OT cost J in Eq. (2), being the Wasserstein distance between passengers' entry and exit distributions [39] and whose minimizers are origin-destination shortest paths. Beside Kirchhoff's law and positivity, capacities in Eqs. (2) and (3) are otherwise unconstrained. One can potentially add additional constraints, e.g., a limited budget, by employing recent ideas in the context of adaptation equations [40]. We do not explore this here.

Projected stochastic gradient descent.—Minimization of Eq. (6) is performed using stochastic gradient descent with a projection step to enforce $w \geq \epsilon$. Importantly, we can derive a closed-form expression for the gradients $\Psi_e = \partial\Omega/\partial w_e$ [39]. To explore the nonconvex landscape of the minimization in Eqs. (6) and (7), we update the weights with dropout at each step, i.e., setting to zero $|E|(1-q)$ random gradients, where $0 \leq q \leq 1$. For $q = 1$ we get vanilla gradient descent.

Bilevel optimization scheme.—In order to find the optimal c and w, and hence F, we iterate between Eq. (8) and PSGD recursively. The scheme is repeated until J and Ω converge. A diagram outlining the optimization method is in Fig. 1; we also provide an open-source code (Bilevel routing on networks with optimal transport, BROT) [47].

Experimental setup.—We analyze BROT's optimal networks against two baselines. The first, referred to as OT, consists of finding passengers' shortest paths without any intervention from the network manager. We assume a unitary cost per unit of length fare, i.e., we set $w = \ell$ with ℓ the Euclidean lengths of the edges, and numerically integrate Eq. (8). The second, referred to as PSGD, reflects the scenario of a network manager that tunes w only relying on the shortest paths taken when $w = \ell$ and that disregards how fluxes redistribute while updating w. In practice, this corresponds to running PSGD only, with initial conditions being $w(0) = \ell + \xi$ and $c_e^i \simeq |F^i_{\mathrm{Dij},e}|$ [39], and then to integrating Eq. (8), with $w = w^\star_{\mathrm{PSGD}}$ being the optimal weights returned by the network manager. Here, ξ is a small zero-sum uniform noise, F_{Dij} are the shortest path fluxes computed with Dijkstra's algorithm, and the approximation arises because, to avoid numerical instabilities, a small

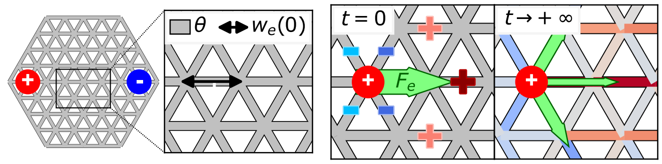

FIG. 1. Bilevel optimization scheme on a lattice. Entry and exit inflows are the red and blue nodes, respectively. Initially, (green) fluxes distribute minimizing the travel cost $w_e(t=0) = \ell_e$, being the length of an edge. If they exceed θ they get penalized; hence, the network manager tunes the weights to encourage rerouting over more expensive (red) or cheaper (blue) edges (for a companion figure, see Supplemental Material [39]).

nonzero c_e^i is allocated to all edges. We fix BROT's initial conditions to $w(0) = \ell + \xi$ and $c_e^i(0) = S^i_{O^i}$.

Synthetic experiments.—First, we study a network of size $|V| = 300$, $|E| = 864$, with nodes placed uniformly at random in the unitary disk and edges extracted from their Delaunay triangulation. Entry and exit inflows are $S^i_{O^i} = +1$ on an origin node at the center, and $S^i_{D^i} = -1/D$, on $D = 4$, 8 destinations D^i on the disk edge. Since $M = 1$, there is only a single index i. Here we discuss results for $D = 8$, for experiments with varying q for $D = 4$, 8; see Supplemental Material [39].

We evaluate J and Ω at convergence for all methods with different q and ranging θ from $\theta = 0$ to a large value $\theta^\star$ where few edges are congested. Results are in Fig. 2(a).

Since for OT the network manager does not intervene, J is constant for all θ, and it is the origin-destination shortest length. Its profile changes when the network manager influences passengers' routes by tuning the weights. Specifically, for PSGD J drops when reducing θ, making it cheaper for the passengers to move. On the contrary, lower θ corresponds to a larger J for BROT. This behavior seemingly favors an uninformed network manager (PSGD) over an informed one (BROT). However, the profile of Ω shows that, even though the traveling cost of PSGD is cheaper, all transport networks at convergence are highly congested (large Ω). BROT successfully trades off the cost of traveling against traffic, outputting low values of Ω for all θ, with only a mild increase as θ approaches zero. This is clarified in Fig. 2(c), where BROT generates ramified loopy networks.

The dropout parameter q allows us to explore the minimization landscape of Eqs. (6) and (7). By decreasing q, i.e., setting more gradients to zero, BROT returns lower Js at all θ, whereas PSGD gives higher ones, conversely for Ω. This impacts the network topologies, which are less ramified and akin to OT trees [48], when q is lower and for the same θ [39]. The trade-off between J and Ω is further laid out in Fig. 2(b) where we show $J - J_{\mathrm{OT}}$ against $\Omega - \Omega_0$, $\Omega_0 = 0$. We highlight in red the nondominated points (also referred to as maximal points) at four values of θ, computed over all q [as in Fig. 2(a)] and 25 random

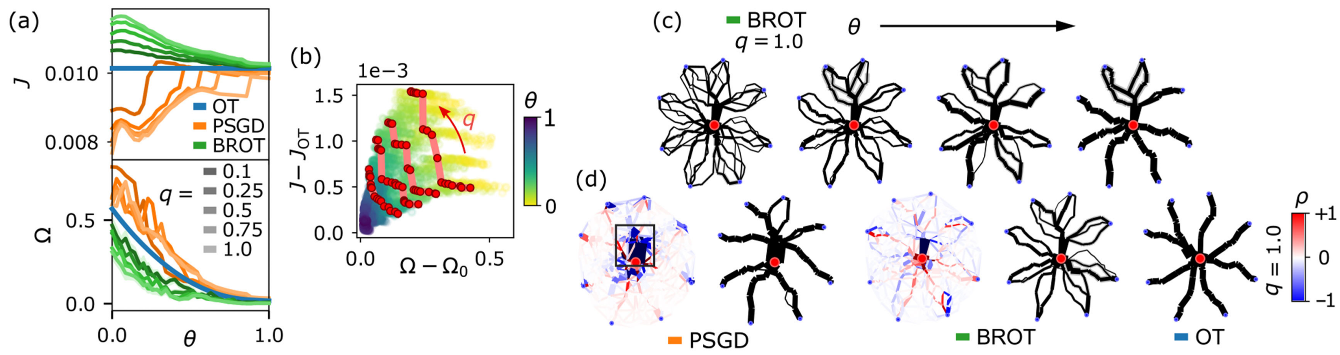

FIG. 2. Overview of the routing schemes. (a) J and Ω against θ. (b) Trade-off $J - J_{\mathrm{OT}}$ vs $\Omega - \Omega_0$ with varying (θ, q, ξ). Nondominated points for $\theta/\theta^\star \simeq \{0.06, 0.2, 0.3, 0.4\}$ are in red. (c) BROT's networks at different θ. Edge widths are proportional to the average fluxes in 50 runs of the algorithm. Gray edge contours are fluxes' standard deviations. (d) Cost (left) and flux (right) networks for all methods and $\theta/\theta^\star = 0.4$. Flux networks are as in (c), whereas edges in the cost networks are colored with ρ and their widths are proportional to the fluxes. The black rectangle frames a region where the network manager triggers high congestion. We conveniently normalize $\theta^\star$ and ρ.

initializations of BROT. Such points are the best J-Ω trade-off attained by the experimental runs [39].

For all q and sufficiently low θ, the price of anarchy (PoA) [49] is greater for PSGD than for OT, i.e., the network manager's intervention increases traffic congestion, having the opposite effect to that intended. We illustrate exemplary networks at convergence in Fig. 2(c). The parameter $\rho = w_X^\star - \ell$ ($X =$ BROT, PSGD), expressing the variation of cost, indicates that the uninformed network manager naively—and significantly—decreases the cost of a small fraction of edges [square in Fig. 2(d)]. This encourages fluxes to largely concentrate on them, thus creating congestion.

To further discern the nature of congestion, we propose two additional metrics. First, the Gini coefficient of the fluxes, $\mathrm{Gini} = \sum_{mn} |x_m - x_n|/2|E|^2\bar{x}$, where $\bar{x} = \sum_e x_e/|E|$ and $x_e = \sum_i |F_e^i|$. $\mathrm{Gini} = 0$ corresponds to uniformly distributed fluxes and larger Gini corresponds to high congestion. Second, the total travel time $T_\theta(s) = \sum_{ei} t_{\theta,e}(s)|F_e^i|$, computed with an affine latency function for overtrafficked edges [38,50], namely, $t_{\theta,e}(s) = \ell_e(1 + s\Delta_e/\theta)/v_\infty$ if $\sum_i |F_e^i| \geq \theta$, and $t_{\theta,e}(s) = \ell_e/v_\infty$ otherwise. Here $v_\infty = 1$ is a (conventionally fixed) free-flow velocity, and s is a sensitivity coefficient to penalize traffic. Results are in Fig. 3.

The Gini coefficient of PSGD fluctuates slightly around the high values attained by the congested shortest path network of OT. For BROT, as θ decreases—more flux gets penalized—Gini sharply drops, yielding progressively distributed networks. The total travel time reveals once again that the uncoordinated action of passengers and the network manager may be detrimental compared to having no tuning of w. In fact, times for PSGD are higher than those for OT. BROT keeps $T_\theta(s)$ small for any value of θ and for both low and high sensitivity. Finally, as θ increases, traffic gradually mitigates, with $\lim_{\theta\to+\infty} T_\theta(s) = T_\infty$ ($T_\infty = J_{\mathrm{OT}}$) being the travel time for infinite capacities, when all passengers flow freely.

The E-road network.—We study the methods on a graph extracted from the international European highways (E-road) [51,52], of size $|V| = 541$ and $|E| = 712$. Entry inflows of passengers are populations of 15 large cities. We assume that all passengers travel from one city to another. Thus, we set for O^i and $v \in D^i$ (being also origin nodes O^j) the exiting number of passengers $\tilde{S}_v$ to be proportional to the product $r_v = \tilde{S}_{O^i}\tilde{S}_{O^j}$, properly normalized to ensure conservation of mass. In this way, cities with high inflows have large outflows, and vice versa for small ones. The total number of passengers to be routed is $\sum_i \tilde{S}_{O^i} \simeq 3 \times 10^7$. We fix $\tilde{\theta}$ (dimensionalized by S_c) so that 43% of the passengers reroute from their congested shortest path, found with Dijkstra's and $w = \ell$.

Results are in Fig. 4. We observe that, in the shortest path configuration of OT, a large volume of passengers travels between the two most populous cities, Madrid and Berlin, on the southernmost region of the network. The uninformed network (PSGD) heavily increases the price of the connections to Milan [39]. This causes a heavy rerouting from Madrid to the north and congests the roads connecting Madrid to Paris and then from Paris to Berlin. In contrast, BROT distributes traffic over a ramified road network.

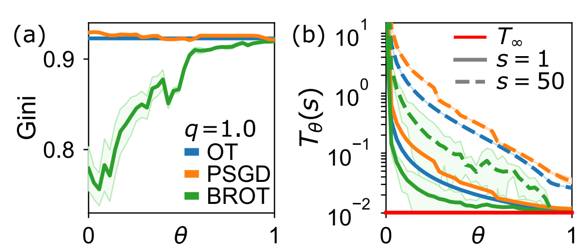

FIG. 3. Measuring traffic congestion, $D = 8$. (a) Gini coefficient against θ. (b) $T_\theta(s)$ against θ. Solid lines correspond to low sensitivity $s = 1$ and dashed ones to $s = 50$; in red we draw T_∞ (free flow). Shades are standard deviations over 50 realizations of the algorithms.

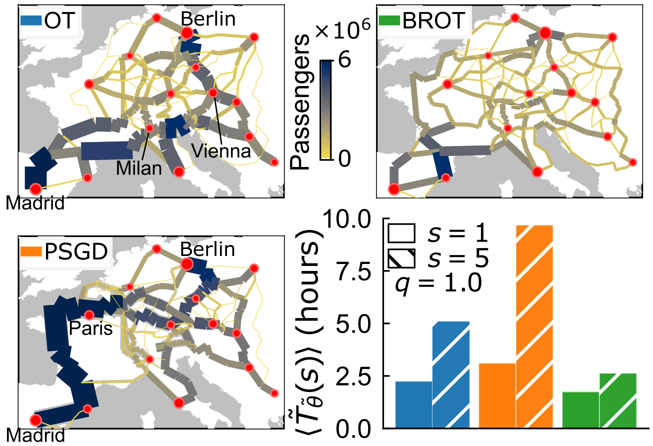

FIG. 4. E-road transport networks. Nodes in red are 15 main cities taken as passenger inflows, their size is proportional to the entry inflows. Edge widths are the total number of passengers $\sum_i |\tilde{F}_e^i|$; gray shades are standard deviations over 50 realizations of the algorithms.

We study the average travel time for all routing schemes. This is $\langle \tilde{T}_{\tilde{\theta}}(s)\rangle = \sum_{ei} \tilde{t}_{e,\tilde{\theta}}(s)|\tilde{F}_e^i| / \sum_{ei} |\tilde{F}_e^i|$, where $\tilde{t}_{\tilde{\theta}}$ is a dimensionalized latency function computed using $\widetilde{\ell}$, the Euclidean distance between cities, and $v_\infty = 100$ (km/h).

Results for $s = 1, 5$ in Fig. 4 show that the average travel time of BROT is substantially lower than that of OT and PSGD. Particularly, for low sensitivity BROT's $\langle \tilde{T}_{\tilde{\theta}}(s)\rangle$ is approximately 1.7 h), while OT's and PSGD's are 2.3 and 3.1 h. Here, BROT leads to a reduction in traveled time of approximately 26% and 45% compared to OT and PSGD. This result becomes starker if the sensitivity increases, here BROT reduces $\langle \tilde{T}_{\tilde{\theta}}(s)\rangle$ of 48% compared to OT—from 5 to 2.6 h—and of 74% compared to PSGD—whose heavy congestion gives $\langle \tilde{T}_{\tilde{\theta}}(s)\rangle \simeq 10$ h. Once again, the PoA (the travel time) is higher if the network manager's intervention is uncoordinated with the passengers (PSGD), as opposed to when there is no intervention (OT).

Experiments on the E-road network for $q = 0.25$, 0.5, and 0.75 are in the Supplemental Material [39].

Conclusion.—BROT relies on theoretical assumptions that can be challenging to meet in real-world traffic control [53], e.g., passengers rerouting more unpredictably than expected by theoretical models. Nevertheless, our analysis on the E-road network demonstrates how an informed tuning of road tolls—where the network manager factors in passengers' rerouting—can be beneficial for reducing the carbon footprint of roads, since traffic jams, and hence longer travels, critically impact greenhouse gas emissions of vehicles [54–56].

To facilitate practitioners using our algorithms, we open source our code [47].

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

*alessandro.lonardi@tuebingen.mpg.de
†caterina.debacco@tuebingen.mpg.de

[1] L. Koçillari *et al.*, The widened pipe model of plant hydraulic evolution, Proc. Natl. Acad. Sci. U.S.A. **118**, e2100314118 (2021).
[2] E. Katifori, G. J. Szöllősi, and M. O. Magnasco, Damage and fluctuations induce loops in optimal transport networks, Phys. Rev. Lett. **104**, 048704 (2010).
[3] H. Ronellenfitsch and E. Katifori, Global optimization, local adaptation, and the role of growth in distribution networks, Phys. Rev. Lett. **117**, 138301 (2016).
[4] H. Ronellenfitsch and E. Katifori, Phenotypes of vascular flow networks, Phys. Rev. Lett. **123**, 248101 (2019).
[5] A. Tero, S. Takagi, T. Saigusa, K. Ito, D. P. Bebber, M. D. Fricker, K. Yumiki, R. Kobayashi, and T. Nakagaki, Rules for biologically inspired adaptive network design, Science **327**, 439 (2010).
[6] A. Lonardi, E. Facca, M. Putti, and C. De Bacco, Designing optimal networks for multicommodity transport problem, Phys. Rev. Res. **3**, 043010 (2021).
[7] A. Lonardi, M. Putti, and C. De Bacco, Multicommodity routing optimization for engineering networks, Sci. Rep. **12**, 7474 (2022).
[8] A. Lonardi, E. Facca, M. Putti, and C. De Bacco, Infrastructure adaptation and emergence of loops in network routing with time-dependent loads, Phys. Rev. E **107**, 024302 (2023).
[9] A. A. Ibrahim, A. Lonardi, and C. De Bacco, Optimal transport in multilayer networks for traffic flow optimization, Algorithms Mol. Biol. **14**, 189 (2021).
[10] A. A. Ibrahim, D. Leite, and C. De Bacco, Sustainable optimal transport in multilayer networks, Phys. Rev. E **105**, 064302 (2022).
[11] V. Bonifaci, E. Facca, F. Folz, A. Karrenbauer, P. Kolev, K. Mehlhorn, G. Morigi, G. Shahkarami, and Q. Vermande, Physarum-inspired multi-commodity flow dynamics, Theor. Comput. Sci. **920**, 1 (2022).
[12] H. Youn, M. T. Gastner, and H. Jeong, Price of anarchy in transportation networks: Efficiency and optimality control, Phys. Rev. Lett. **101**, 128701 (2008).
[13] M. T. Gastner and M. E. J. Newman, Optimal design of spatial distribution networks, Phys. Rev. E **74**, 016117 (2006).
[14] R. Selten, T. Chmura, T. Pitz, S. Kube, and M. Schreckenberg, Commuters route choice behaviour, Games Econ. Behav. **58**, 394 (2007).
[15] M. Beckmann, C. B. McGuire, and C. B. Winsten, *Studies in the Economics of Transportation* (Cowles Commission for Research in Economics, 1956).
[16] V. Bonifaci, K. Mehlhorn, and G. Varma, Physarum can compute shortest paths, J. Theor. Biol. **309**, 121 (2012).
[17] V. Bonifaci, Physarum can compute shortest paths: A short proof, Inf. Proc. Lett. **113**, 4 (2013).

[18] A. Tero, R. Kobayashi, and T. Nakagaki, A mathematical model for adaptive transport network in path finding by true slime mold, J. Theor. Biol. **244**, 553 (2007).

[19] F. Folz, K. Mehlhorn, and G. Morigi, Noise-induced network topologies, Phys. Rev. Lett. **130**, 267401 (2023).

[20] D. Hu and D. Cai, Adaptation and optimization of biological transport networks, Phys. Rev. Lett. **111**, 138701 (2013).

[21] J. B. Kirkegaard and K. Sneppen, Optimal transport flows for distributed production networks, Phys. Rev. Lett. **124**, 208101 (2020).

[22] D. Baptista, D. Leite, E. Facca, M. Putti, and C. De Bacco, Network extraction by routing optimization, Sci. Rep. **10**, 088702 (2020).

[23] E. Facca, F. Cardin, and M. Putti, Towards a stationary Monge-Kantorovich dynamics: The physarum polycephalum experience, SIAM J. Appl. Math. **78**, 651 (2016).

[24] E. Facca, S. Daneri, F. Cardin, and M. Putti, Numerical solution of Monge-Kantorovich equations via a dynamic formulation, J. Sci. Comput. **82**, 68 (2020).

[25] E. Facca, F. Cardin, and M. Putti, Branching structures emerging from a continuous optimal transport model, J. Comput. Phys. **447**, 110700 (2021).

[26] D. Leite and C. D. Bacco, Revealing the similarity between urban transportation networks and optimal transport-based infrastructures, arXiv:2209.06751.

[27] L. J. LeBlanc, E. K. Morlok, and W. P. Pierskalla, An efficient approach to solving the road network equilibrium traffic assignment problem, Transp. Res. **9**, 309 (1975).

[28] Y. Bao, Z. Gao, M. Xu, H. Sun, and H. Yang, Travel mental budgeting under road toll: An investigation based on user equilibrium, Transp. Res. Part A **73**, 1 (2015).

[29] A. Lonardi, D. Baptista, and C. De Bacco, Immiscible color flows in optimal transport networks for image classification, Front. Phys. **11**, 1089114 (2023).

[30] H. F. Po, C. H. Yeung, and D. Saad, Futility of being selfish in optimized traffic, Phys. Rev. E **103**, 022306 (2021).

[31] Y.-Z. Xu, H. F. Po, C. H. Yeung, and D. Saad, Scalable node-disjoint and edge-disjoint multiwavelength routing, Phys. Rev. E **105**, 044316 (2022).

[32] C. H. Yeung, Coordinating dynamical routes with statistical physics on space-time networks, Phys. Rev. E **99**, 042123 (2019).

[33] J. Jiang, X. Wang, and Y.-C. Lai, Optimizing biologically inspired transport networks by control, Phys. Rev. E **100**, 032309 (2019).

[34] V. Colizza, J. R. Banavar, A. Maritan, and A. Rinaldo, Network structures from selection principles, Phys. Rev. Lett. **92**, 198701 (2004).

[35] T. S. Tai and C. H. Yeung, Global benefit of randomness in individual routing on transportation networks, Phys. Rev. E **100**, 012311 (2019).

[36] T. S. Tai and C. H. Yeung, Adaptive strategies for route selection en-route in transportation networks, Chin. J. Phys. (Taipei) **77**, 712 (2022).

[37] G. Yan, T. Zhou, B. Hu, Z.-Q. Fu, and B.-H. Wang, Efficient routing on complex networks, Phys. Rev. E **73**, 046108 (2006).

[38] B. Li, D. Saad, and C. H. Yeung, Bilevel optimization in flow networks: A message-passing approach, Phys. Rev. E **106**, L042301 (2022).

[39] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.131.267401 for more information on the problem formulation, closed-form expressions of the gradients, and additional experiments, which includes [4,40–45].

[40] A. A. Ibrahim, M. Muehlebach, and C. D. Bacco, Optimal transport with constraints: From mirror descent to classical mechanics, arXiv:2309.04727.

[41] M. Muehlebach and M. I. Jordan, On constraints in first-order optimization: A view from non-smooth dynamical systems, J. Mach. Learn. Res. **23**, 1 (2022).

[42] G. H. Golub and V. Pereyra, The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate, SIAM J. Numer. Anal. **10**, 413 (1973).

[43] L. Kantorovich, Mathematical methods of organizing and planning production, Manage. Sci. **6**, 366 (1960).

[44] C. Villani, *Optimal Transport: Old and New* (Springer, Berlin, Heidelberg, 2009), Vol. 338.

[45] A.-G. Rolland-Lagan and P. Prusinkiewicz, Reviewing models of auxin canalization in the context of leaf vein pattern formation in Arabidopsis, Plant J. **44**, 854 (2005).

[46] J. W. Rocks, H. Ronellenfitsch, A. J. Liu, S. R. Nagel, and E. Katifori, Limits of multifunctionality in tunable networks, Proc. Natl. Acad. Sci. U.S.A. **116**, 2506 (2019).

[47] Bilevel routing on network with optimal transport: BROT, https://github.com/aleable/BROT.

[48] J. R. Banavar, F. Colaiori, A. Flammini, A. Maritan, and A. Rinaldo, Topology of the fittest transportation network, Phys. Rev. Lett. **84**, 4745 (2000).

[49] The price of anarchy measures how much a system degrades due to the greedy behavior of its agents. In our case, the global efficiency of the system is measured by Ω, which increases as the network gets more congested by greedy passengers.

[50] T. Roughgarden and E. Tardos, How bad is selfish routing?, J. ACM **49**, 236 (2002).

[51] J. Kunegis, KONECT: The Koblenz network collection, in *Proceedings of the 22nd International Conference on World Wide Web*, WWW ’13 Companion (Association for Computing Machinery, New York, 2013), pp. 1343–1350.

[52] L. Šubelj and M. Bajec, Robust network community detection using balanced propagation, Eur. Phys. J. B **81**, 353 (2011).

[53] M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos, and Y. Wang, Review of road traffic control strategies, Proc. IEEE **91**, 2043 (2003).

[54] T. Van Woensel, R. Creten, and N. Vandaele, Managing the environmental externalities of traffic logistics: The issue of emissions, Prod. Oper. Manag. **10**, 207 (2001).

[55] F. Kellner, Exploring the impact of traffic congestion on CO_2 emissions in freight distribution networks, Logist. Res. **9**, 21 (2016).

[56] M. Barth and K. Boriboonsomsin, Traffic congestion and greenhouse gases, ACCESS Mag. **1**, 2 (2009).