Designing optimal networks for multicommodity transport problem

Alessandro Lonardi [1,*] Enrico Facca,[2] Mario Putti,[3] and Caterina De Bacco [1,†]

[1]*Max Planck Institute for Intelligent Systems, Cyber Valley, Tübingen 72076, Germany*
[2]*Centro di Ricerca Matematica Ennio De Giorgi, Scuola Normale Superiore, Piazza dei Cavalieri 3, Pisa, Italy*
[3]*Department of Mathematics "Tullio Levi-Civita", University of Padua, Via Trieste 63, Padua, Italy*

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

I. INTRODUCTION

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

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

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

---

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

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

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

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

II. MODEL

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

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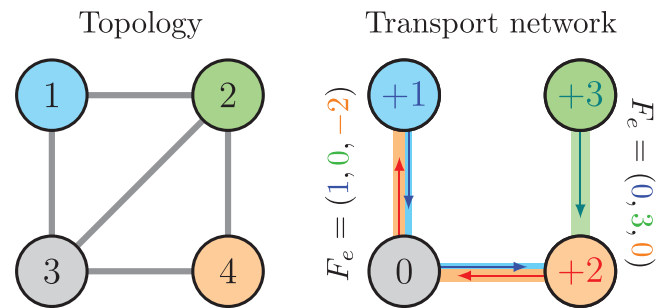

FIG. 1. Multicommodity problem illustration for $M = 3$. Left: topology of the graph; numbers inside nodes correspond to their indices v. Right: an admissible configuration of fluxes. Here, numbers inside nodes correspond to their mass inflow, and we assume each commodity to have its mass concentrated in a single vertex; in the gray node, no mass is entering or exiting, i.e., $S_3^i = 0$ for every i. Widths of edges are drawn proportional to F_e^i, and in the case where all $F_e^i = 0$, links are not drawn; arrows denote the direction of the colored fluxes F_e^i. Notice how blue mass and orange mass share the same edges, thus creating possible traffic congestion. The inflow-outflow rates of each color are $S^1 = (1, 0, 0, -1)$ (blue), $S^2 = (0, 3, 0, -3)$ (green), and $S^3 = (-2, 0, 0, 2)$ (orange).

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

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

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

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

A. Introducing a shared conductivity

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

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

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

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

B. The dynamics

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

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

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

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

In the presence of only one commodity, our model is similar to the dynamics used to solve the basis pursuit problem on networks [15] and as a principled mechanism for filtering networks from redundancies [16]. However, both cases are limited to one-commodity scenarios. A similar dynamics is also proposed in Ref. [5], where the authors focus on the average time evolution of a stochastic model with fluctuating loads. Analogously to these one-commodity cases, one can efficiently solve the system in Eqs. (1)–(4) using optimized numerical methods; however, in our case the complexity increases with the number of colors (see Appendix E 2 for more details). Our model also bears a close mathematical relationship to a recent work, where similar ideas have been studied in a multicommodity setup [17]. Beyond the fact that this work focuses on the case $\beta = 1$ (i.e., shortest-path-like) and thus on a convex optimization scenario, there is one other main conceptual difference compared with our model. Notice

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

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

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

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

We formally show that Eq. (5) defines a well-defined Lyapunov functional for the dynamics of Eq. (4) in Appendix A 1, following similar arguments as in Ref. [17]. This extends the work of Bonifaci *et al.* [11], where a similar functional has been proposed to complete the characterization of the dynamics regulating slime molds' evolutionary feedback mechanism.

C. Mapping to standard optimization setups

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

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

such that

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

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

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

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

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

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

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

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

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

$$F_e^i = \frac{\hat{\mu}_e}{\ell_e} \left(p_u^i - p_v^i \right), \qquad (12)$$

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

III. ANALYSIS OF THE OPTIMAL TOPOLOGIES

A. Optimal topologies may have loops

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

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

B. Phase diagram tree-loops (*S* fixed)

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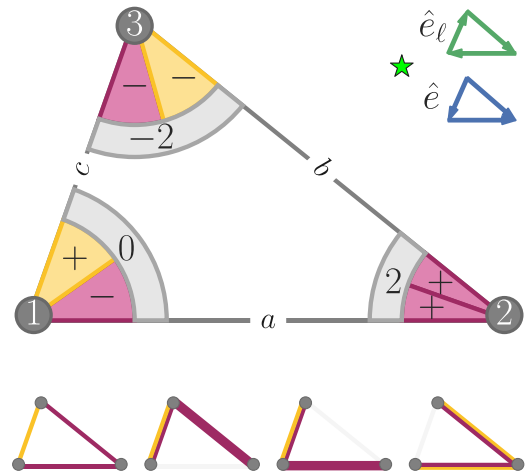

FIG. 2. Toy model where loops are optimal. Here, $M = 2$; hence only two colors move though the network. The triangle network has source vectors $S^1 = (+1, -1, 0)$, $S^2 = (-1, +2, -1)$, $\ell_a = \ell_b = 1.5$, $\ell_c = 1$. The gray patches denote the net loads of each node when ignoring the colors. At the bottom we show one loopy solution on the left and three trees on the right. The green and blue arrows denote the orientation of the loop defined in Appendix D and of the edges, respectively. In detail, the loop has fluxes $F_a = (0, -1)$, $F_b = (0, -1)$, $F_c = (-1, 0)$, the leftmost tree has $F_a = (0, 0)$, $F_b = (0, -2)$, $F_c = (-1, +1)$ (similarly for the other two). The green star refers to the topology of the toy model used in Fig. 3.
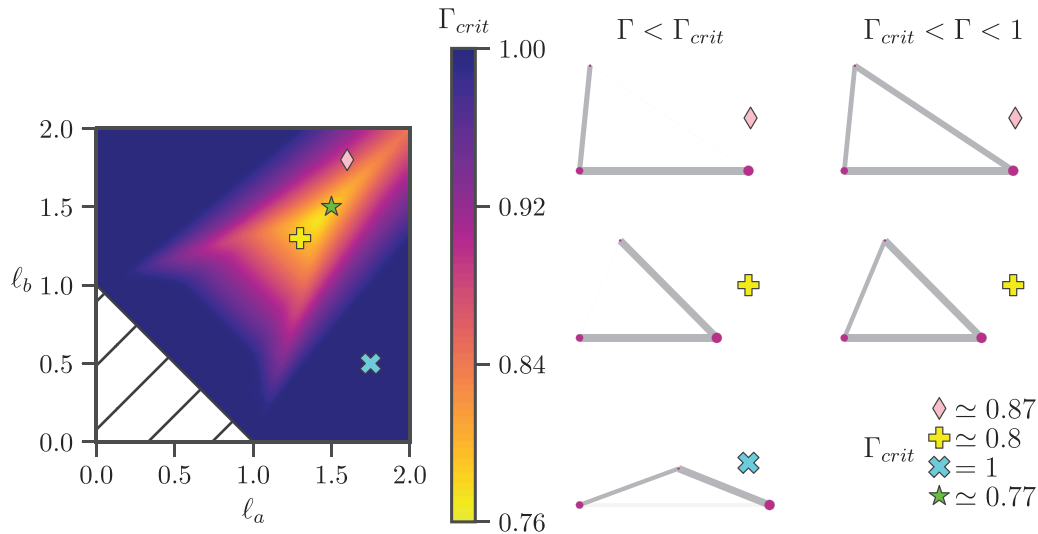
FIG. 3. Phase diagram in ℓ. Γ_{crit} denotes the minimum value of Γ above which loops are optimal. The setup is the same as for the toy model in Fig. 2. Values of Γ_{crit} are found by solving Eq. (13); $\ell_c = 1$. The area under the triangular surface with white background is not allowed as the triangular geometry is not defined there. One can notice that there is an entire region where $\Gamma_{\text{crit}} \leqslant 1$; inside it, loops are optimal. On the right-hand side of the figure we show three different topologies, i.e., choices of ℓ_a, ℓ_b for which optimal solutions can be loopy; these are associated with the markers drawn on the heat map, and the green star is the configuration of the toy model in Fig. 2. In particular, running our dynamics on the pink-diamond graph (respectively, the yellow-plus graph) leads to a loopy configuration for $\Gamma > \Gamma_{\text{crit}}$ or to a tree if $\Gamma < \Gamma_{\text{crit}}$. Running our dynamics on the blue-cross graph returns always since its Γ_{crit} is equal to 1. Widths of the edges are proportional to the final $||F_e||_2$, and edges that are not visible have negligible fluxes; nodes' dimensions are proportional to their inflowing mass. In the bottom right portion of the panel we report the values of Γ_{crit} obtained solving Eq. (13) fixing ℓ_a, ℓ_b as given by the markers.

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

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

(13)

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

C. Phase diagram tree-loops (lengths fixed)

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

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

Additionally, the way the topology changes between these two configurations depends on the exponent Γ. In this simple scenario, we can have either a tree or a loop at $\Gamma = 0.6$ ($\Gamma < \Gamma_{\text{crit}} \simeq 0.77$) or $\Gamma = 0.8$ ($\Gamma > \Gamma_{\text{crit}}$), respectively; see
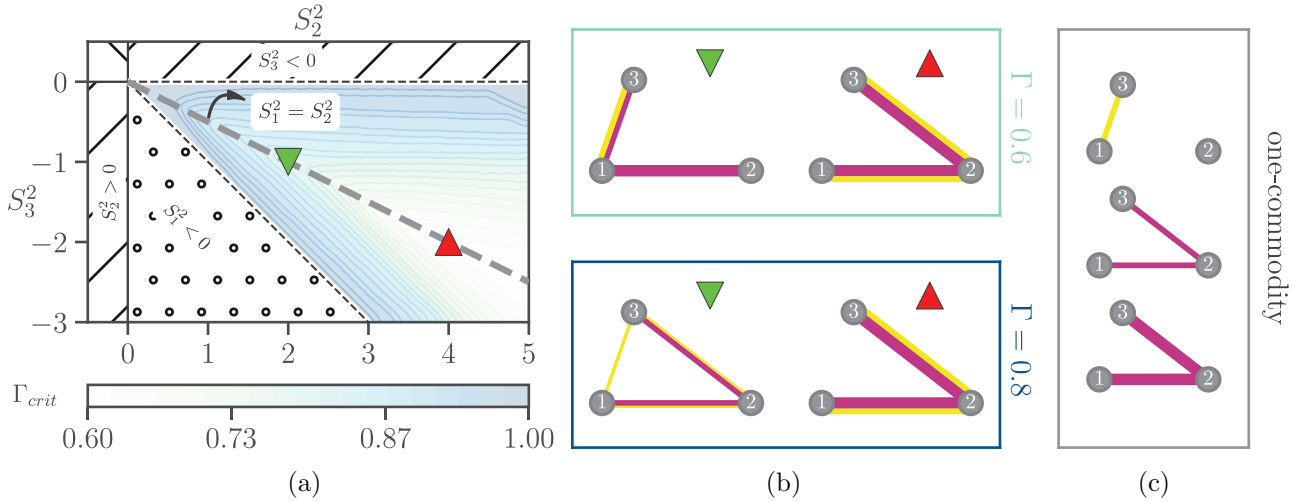
FIG. 4. Interaction between commodities. The figure shows how changing the load of one commodity influences the path taken by others. (a) We plot the heat map of Γ_{crit} obtained using Eq. (13), with $S_1^1 = -S_3^1 = +1$ and for different configurations of the purple commodity ($i = 2$): (i) $S_2^2 = +2$, $S_3^2 = S_1^2 = -1$; (ii) $S_2^2 = +4$, $S_3^2 = S_1^2 = -2$. The areas under the white surface correspond to regions where node 1 is a sink, respectively, nodes 2 and 3 are sources, for the purple mass. The green and the red triangular markers denote the configurations we discuss in Sec. III C and in the rest of the panel. (b) Optimal graphs for the green and the red triangle, fixing $\Gamma = 0.6$ and $\Gamma = 0.8$. The width of each edge e is proportional to $|F_e^i|$ for each color. (c) One-commodity solutions obtained injecting only the yellow or the purple mass in the triangle network.

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

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

IV. CONCLUSIONS

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

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

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

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

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

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

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

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

ACKNOWLEDGMENTS

APPENDIX A: THE LYAPUNOV FUNCTIONAL

1. The Lyapunov functional is well defined

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

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

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

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

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

Now, we claim that for each edge

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

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

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

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

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

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

obtain

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

which is the equality we wanted to show.

APPENDIX B: MAPPING THE DYNAMICS TO AN OPTIMIZATION PROBLEM

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

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

$$J_\beta(\{\hat\mu_e\}, \{F_e\}) = \frac{1}{2} \sum_e \frac{\ell_e}{\hat\mu_e} ||F_e||_2^2$$

$$+ \frac{\lambda}{2(2-\beta)} \left(\sum_e \ell_e \hat\mu_e^{2-\beta} - K^{2-\beta} \right)$$

$$+ \sum_{v,i} \chi_v^i \left(\sum_e B_{ve} F_e^i - S_v^i \right). \quad \text{(B1)}$$

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

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

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

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

yielding

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

Substituting back into Eq. (B2), we obtain

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

APPENDIX C: PHASE DIAGRAM FOR A TOY MODEL

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

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

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

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

$$J_\Gamma(\mathcal{G}_L) = \left[(F_a^1)^2 + (F_a^2)^2 \right]^{\Gamma/2} \ell_a$$

$$+ \left[(F_a^1)^2 + (F_a^2 + 2)^2 \right]^{\Gamma/2} \ell_b$$

$$+ \left[(F_a^1 - 1)^2 + (F_a^2 + 1)^2 \right]^{\Gamma/2} \ell_c. \quad \text{(C4)}$$

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

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

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

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

Analytically from Eqs. (C5)–(C7) or numerically solving Eq. (13) (fixing F_a^* as proposed for $\mathcal{G}_L$) one can show that $\Gamma^* \simeq 0.83$. Notice that such Γ^* is not optimal, in the sense that for other choices of F_a we may find lower values of the exponent Γ enabling loopy networks to be optimal; we denote the minimum of these values as Γ_{crit}. Numerically solving Eq. (13) for the toy model just discussed returns $\Gamma_{\text{crit}} \simeq 0.77$,
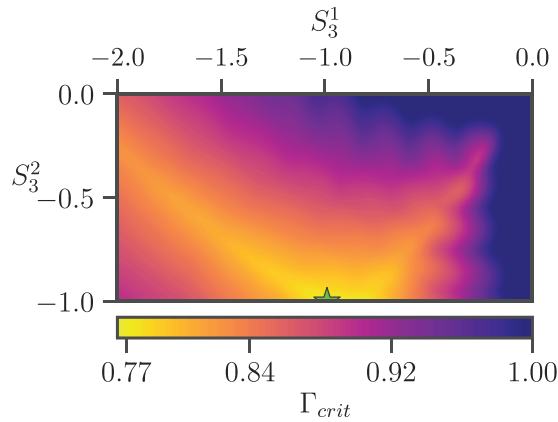
FIG. 5. Phase diagram in S. This can be numerically found by fixing $\{\ell_e\}$ as in Fig. 2, $S_2^1 = -1 - S_3^1$, $S_1^2 = -2 - S_3^2$, while varying S_3^1 and S_3^2. The color bar denotes the value of Γ_{crit} above which optimal solutions can be loopy. The green star denotes the configuration of S used in Fig. 2.

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

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

APPENDIX D: TRIMMING LOOPS TO OBTAIN TREES

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

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

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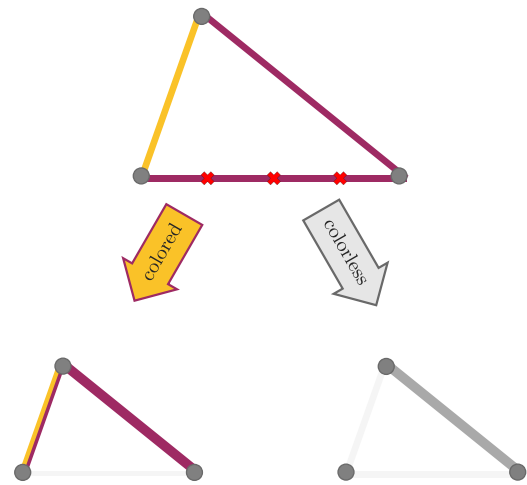

FIG. 6. Sketch of the trimming procedure described in Appendix D, on the toy model of Fig. 2. The tree obtained in the colored case is not optimal, while the one obtained in the colorless case has lower cost but violates Kirchhoff's law.

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

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

APPENDIX E: NUMERICAL IMPLEMENTATION

1. Implementation details and gradient descent

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

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

Algorithm 1. Dynamics.

1: Input: $\mathcal{G}(\mathcal{V}, \mathcal{E})$ = adjacency list, nodes coordinates; M;
   inflows; $0 < \beta < 2$

2: Initialize: (i) S and (ii) $\{\hat{\mu}_e\}$ [e.g., sampling as i.i.d. $\hat{\mu}_e \sim U(0, 1)$]

3: **while** convergence not achieved **do**

4:     solve Kirchhoff's law as in Eq. (1) $\rightarrow \{P_v^i\}$

5:     update conductivities with a finite difference
       discretization of Eq. (4): $\{\hat{\mu}_e^t\} \rightarrow \{\hat{\mu}_e^{t+1}\}$

6: **end while**

7: Return: fluxes $\{F_e^i\}$ at convergence, computed using
   $F_e^i = \hat{\mu}_e(p_u^i - p_v^i)/\ell_e, \; e = (u, v)$

---

to Δt^2, with Δt being the difference between two consecutive time steps, which can be arbitrarily set in the input. Solutions of Kirchhoff's law have been computed using a sparse direct solver.

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

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

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

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

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

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

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

Algorithm 2. Optimization.

1: Input: $\mathcal{G}(\mathcal{V}, \mathcal{E})$ = adjacency list, nodes coordinates; M;
   inflows; $0 < \beta < 2$

2: Initialize: (i) S and (ii) $\{\hat{\mu}_e\}$ [e.g., sampling as i.i.d. $\hat{\mu}_e \sim U(0, 1)$]

3: **while** convergence not achieved **do**

4:     solve Kirchhoff's law as in Eq. (1) $\rightarrow \{P_v^i\}$

5:     update fluxes using Eq. (12)

6:     compute $\hat{\mu}_e(F_e)$ using Eq. (11)

7: **end while**

8: Return: fluxes $\{F_e^i\}$ at convergence

---

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

2. Computational complexity

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

We test the efficiency of our schemes by measuring the total running time (in seconds) to reach convergence for different values of β, M, N. Results are shown in Fig. 7. We notice that the two algorithms, Dynamics and Optimization, have similar computational complexity. Their small running-time differences are negligible and only due to how convergence is precisely defined, i.e., how the corresponding parameters τ_{dyn} and τ_{opt} are set. The running time is shorter for $\beta < 1$ (traffic optimization, loopy) than in the opposite scenario of $\beta > 1$ (minimization of infrastructural cost, treelike). The case $\beta = 1$ is more nuanced, as the cost transitions between two opposite situations. In this case, Optimization fails to converge for $M/N < 1$, if convergence is defined in terms of variations of $||F_e||_2$ between iteration steps. This is because the algorithm gets lost in degenerate local minima, configurations with the same cost but different sets of fluxes. This lack of convergence suggests that, for $\beta = 1$, the energy landscape around these minima is flat, i.e., there are many configurations with the same cost but non-negligible differences in their fluxes. The Optimization routine keeps switching between these different states. In this case, one can simply pick one of these many possible solutions as an example local optimum. The dynamics does instead converge. This suggests that Dynamics is biased towards one of these degenerate solutions. For $M/N = 1$, Optimization converges with the same running time as Dynamics, suggesting that as we enlarge M, the landscape becomes less flat. A possible cause is that by increasing M the system has more constraints to be satisfied via Kirchhoff's law, which reduces the number of possible degenerate solutions. This claim is also supported by the
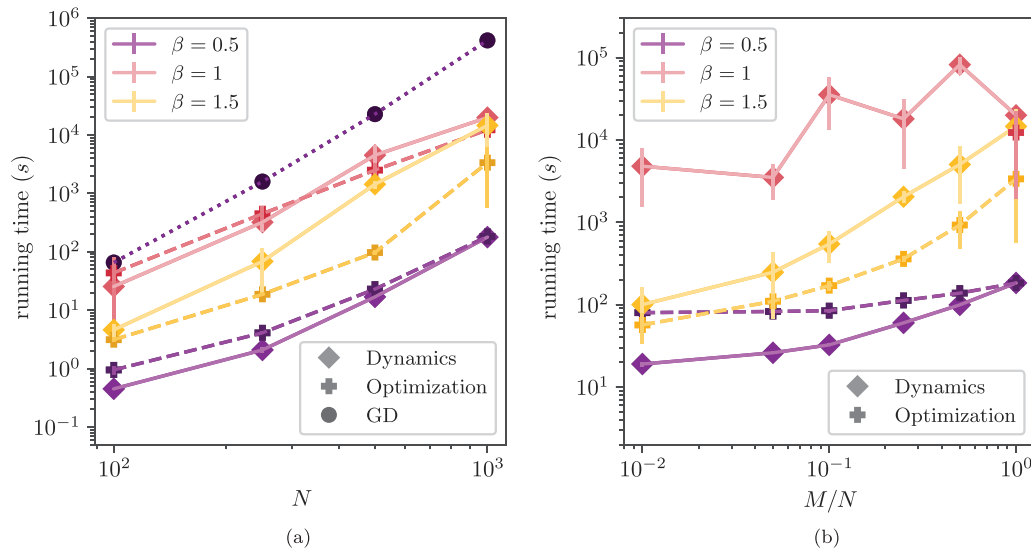
FIG. 7. Computational complexity. (a) Running time (in seconds) as a function of system size N. (b) Running time as a function of the ratio M/N between the number of commodities and system size. GD denotes gradient descent, implemented with Eqs. (E3) and (E4); we only show it for $\beta = 0.5$ as for the other values it fails to converge within a reasonable time. Similarly, for $\beta = 1$ and $M/N < 1$, Optimization fails to converge, and hence we only report Dynamics.

behavior of Dynamics' running time, which does not monotonically increase with M/N in this case, as shown in Fig. 7(b). These behaviors highlight relevant differences between the two implementations. Finally, we note that the computational complexity could in principle be further reduced to $O(MN)$ using multigrid methods [29]; we do not explore this here.

[1] J. R. Banavar, F. Colaiori, A. Flammini, A. Maritan, and A. Rinaldo, Topology of the Fittest Transportation Network, Phys. Rev. Lett. **84**, 4745 (2000).

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

[3] F. Corson, Fluctuations and Redundancy in Optimal Transport Networks, Phys. Rev. Lett. **104**, 048703 (2010).

[4] K. Sinclair and R. C. Ball, Mechanism for Global Optimization of River Networks from Local Erosion Rules, Phys. Rev. Lett. **76**, 3360 (1996).

[5] D. Hu and D. Cai, Adaptation and Optimization of Biological Transport Networks, Phys. Rev. Lett. **111**, 138701 (2013).

[6] H. Ronellenfitsch and E. Katifori, Global Optimization, Local Adaptation, and the Role of Growth in Distribution Networks, Phys. Rev. Lett. **117**, 138301 (2016).

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

[8] H. Ronellenfitsch and E. Katifori, Phenotypes of Vascular Flow Networks, Phys. Rev. Lett. **123**, 248101 (2019).

[9] F. Kaiser, H. Ronellenfitsch, and D. Witthaut, Discontinuous transition to loop formation in optimal supply networks, Nat. Commun. **11**, 5796 (2020).

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

[11] V. Bonifaci, K. Mehlhorn, and G. Varma, *Physarum* can compute shortest paths, J. Theor. Biol. **309**, 121 (2012).

[12] F. Santambrogio, Optimal channel networks, landscape function and branched transport, Interfaces Free Boundaries **9**, 149 (2007).

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

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

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

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

[17] V. Bonifaci, E. Facca, F. Folz, A. Karrenbauer, P. Kolev, K. Mehlhorn, G. Morigi, G. Shahkarami, and Q. Vermande, *Physarum* multi-commodity flow dynamics, arXiv:2009.01498.

[18] K. Salimifard and S. Bigharaz, The multicommodity network flow problem: State of the art classification, applications, and solution methods, Oper. Res. Int. J. (2020).

[19] C. H. Yeung, D. Saad, and K. Y. M. Wong, From the physics of interacting polymers to optimizing routes on the London Underground, Proc. Nat. Acad. Sci. **110**, 13717 (2013).

[20] C. H. Yeung and D. Saad, Networking - A statistical physics perspective, J. Phys. A: Math. Theor. **46**, 103001 (2011).

[21] T. Nelson and N. Dengler, Leaf vascular pattern formation, Plant Cell **9**, 1121 (1997).

[22] C. B. Schaffer, B. Friedman, N. Nishimura, L. F. Schroeder, P. S. Tsai, F. F. Ebner, P. D. Lyden, and D. Kleinfeld,

Two-photon imaging of cortical surface microvessels reveals a robust redistribution in blood flow after vascular occlusion, PLoS Biol. **4**, e22 (2006).

[23] A. Nardini, G. Pedà, and N. L. Rocca, Trade-offs between leaf hydraulic capacity and drought vulnerability: Morphoanatomical bases, carbon costs and ecological consequences, New Phytol. **196**, 788 (2012).

[24] J. R. Banavar, A. Maritan, and A. Rinaldo, Size and form in efficient transportation networks, Nature (London) **399**, 130 (1999).

[25] Q. Xia, Optimal paths related to transport problems, Commun. Contemp. Math. **5**, 251 (2003).

[26] A. A. Ibrahim, A. Lonardi, and C. De Bacco, Optimal transport in multilayer networks for traffic flow optimization, Algorithms **14**, 189 (2021).

[27] https://github.com/aleable/McOpt.

[28] V. Bonifaci, A Laplacian approach to ℓ_1-norm minimization, Comput. Optim. Appl., **79**, 441 (2021).

[29] W. L. Briggs, V. E. Henson, and S. F. McCormick, *A Multigrid Tutorial* (SIAM, Philadelphia, 2000).