Optimal transport in multilayer networks

Abdullahi Adinoyi Ibrahim,[1, *] Alessandro Lonardi,[1, †] and Caterina De Bacco[1, ‡]

[1] *Max Planck Institute for Intelligent Systems, Cyber Valley, Tübingen 72076, Germany*

Modeling traffic distribution and extracting optimal flows in multilayer networks is of utmost importance to design efficient multi-modal network infrastructures. Recent results based on optimal transport theory provide powerful and computationally efficient methods to address this problem, but they are mainly focused on modeling single-layer networks. Here we adapt these results to study how optimal flows distribute on multilayer networks. We propose a model where optimal flows on different layers contribute differently to the total cost to be minimized. This is done by means of a parameter that varies with layers, which allows to flexibly tune the sensitivity to traffic congestion of the various layers.

As an application, we consider transportation networks, where each layer is associated to a different transportation system and show how the traffic distribution varies as we tune this parameter across layers. We show an example of this result on the real 2-layer network of the city of Bordeaux with bus and tram, where we find that in certain regimes the presence of the tram network significantly unburdens the traffic on the road network. Our model paves the way to further analysis of optimal flows and navigability strategies in real multilayer networks.

I. Introduction

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

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

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

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

---

* abdullahi.ibrahim@tuebingen.mpg.de

† alessandro.lonardi@tuebingen.mpg.de

‡ caterina.debacco@tuebingen.mpg.de

What makes multilayer networks different than single-layer in transportation

Having given the broader context for our work, we now highlight the main features of transport on multilayer networks. The presence of edges between layers (inter-layer edges) makes a multilayer network fundamentally distinct from a standard single-layer one, as these edges allow passengers to switch between transportation modes. However, this is not the only difference. In fact, in a multilayer network, the various layers have different characteristics. The main one is that the type of transportation cost varies across layers. For example, the cost to build and maintain the infrastructure differs depending on the transportation mode, with subway or rail tracks costing more than a road network. Moreover, the cost assigned to traffic congestion is also different, as road networks are more sensitive to traffic bottlenecks than rail ones. In addition, also the power dissipated differs depending on the means of transportation, as running a tram generally produces less CO_2 emissions than running a bus. All these different features impact the results of an optimal transport problem, as the network features contributing to the cost function to be optimized vary with layers, and thus also the optimal solution.
Finally, the network topologies themselves vary with layers [41], as a bus network has many edges with short lengths, while a rail network tends to have fewer edges but longer. In addition, the weights assigned to each edge differ based on the layer, which can induce a coupling between layers [42].

II. Materials and Methods

A. Multilayer transportation networks

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

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

In case of multilayer networks, we need to be careful on how stations connecting multiple transportation modes are represented. In fact, if an entry station connects more than one layer, we may not be able to distinguish in what layer a passenger enters. In other words, if a node u belongs to more than one layer, i.e. a node u_α exists for more than one value of α, we may not be able to tell whether the passengers entering u entered from u_α, u_γ or from any of the other instances of node u in the various layers. To alleviate this problem, we build auxiliary *super* nodes u, which do not belong to any layer in particular but instead connect the various instances of the same node in the various layers together. Specifically, we remove all the inter-layer edges (u_α, u_γ) and replace them with auxiliary *inter-super* edges (u_α, u) connecting all the instances u_α of node u with the super node u, as in a star graph, so that the original edge (u_α, u_γ) has been replaced by a 2-edge path $\{(u_\alpha, u), (u_\gamma, u)\}$.

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

Moreover, the introduction of super nodes and edges facilitates how we represent the multilayer network. In fact, by adding these auxiliary super nodes and inter-super edges, we only need to consider an individual network adjacency matrix A, instead of two separate tensors. This matrix has entries $A_{uv} = 1$ if an edge exists between nodes u and v and 0 otherwise, where a node u can be a node u_α in layer α or a super node u. The set of nodes will then be
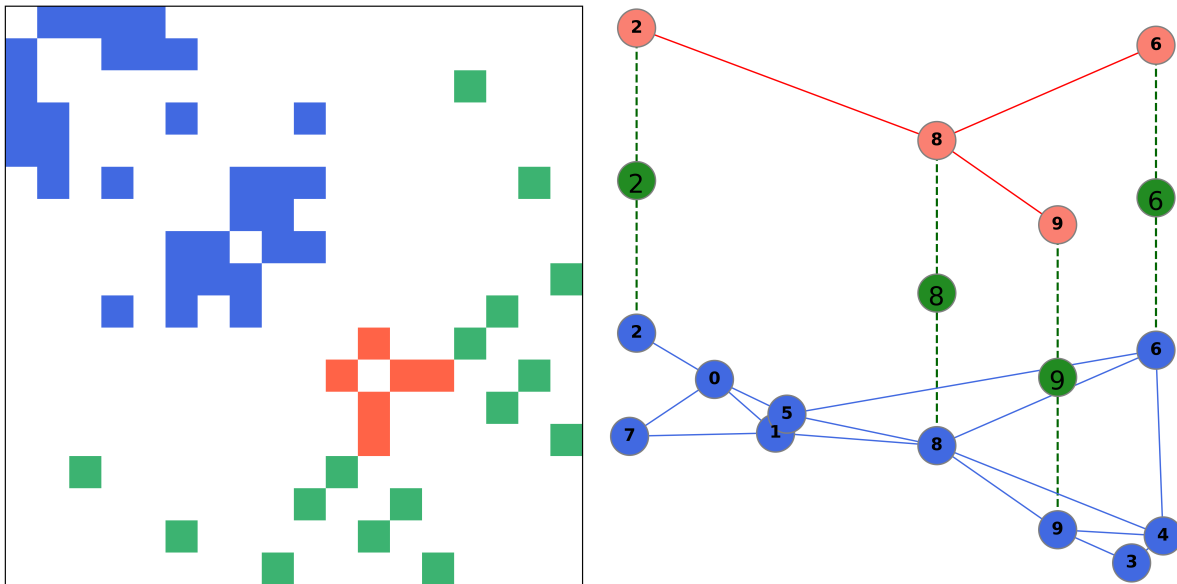
FIG. 1: Example of multilayer structure. We show an example of a 2-layer network with $N = 18$ ($N_1 = 10$, $N_2 = 4$ and $N_{super} = 4$). (Left) adjacency matrix A, colors denote the layer type: blue is layer 1, red is layer 2 and green is the super layer. (Right) the 2-layer network with layer 1 on the bottom, layer 2 on top, and the super nodes in between.

$\mathcal{V} = \mathcal{V}_0 \cup \mathcal{V}_{super}$, where $\mathcal{V}_{super}$ is the set of super nodes, and $|\mathcal{V}_{super}| = N_{super}$ is their number, which corresponds to the number of nodes that belong to more than one layer. Similarly, the new set of edges is $\mathcal{E} = (\cup_\alpha \mathcal{E}_\alpha) \cup \mathcal{E}_{super}$ where $\mathcal{E}_{super}$ is the set of inter-super edges. The final numbers of nodes and edges are $N = |\mathcal{V}| = N_0 + N_{super}$ and $E = |\mathcal{E}| \geq E_0$. Notice that this construction is equivalent to assume that the network has $L + 1$ layers, where the extra layer is made of inter-super edges $\mathcal{E}_{super}$ and all nodes incident to them (without loss of generality, we assume that all the inter-super edges are treated equally). We denote it as *super* layer and this corresponds to $\alpha = L + 1$, so that $\mathcal{E}_{L+1} \equiv \mathcal{E}_{super}$. We show an example of this structure in Figure 1.

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

B. The model

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

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

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

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

With this in mind, we enforce mass conservation by imposing Kirchhoff's law on nodes. To properly enforce this constraint we need to consider all the edges, both intra-layer and inter-layer edges. This can be compactly written by considering the multilayer network signed incidence matrix B with entries $B_{ve} = 1, -1, 0$ if node $v \in \mathcal{V}$ is the start, end of edge $e \in \mathcal{E}$, or none of them, respectively. With this in mind, Kirchhoff's law can be written as:

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

Finally, we assume that the conductivities follow the dynamics:

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

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

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

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

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

The parameter β_{q_e} (taking value β_α on layer α) regulates how the fluxes should distributes in each of the layers. In fact, according to Eq. (4), when $\beta_\alpha > 1$, the fluxes are encouraged to consolidate into few edges of layer α, being $\Gamma(\beta_\alpha) < 1$ and thus the cost in Eq. (4) sub-linear. In the opposite scenario, when $0 < \beta_\alpha < 1$, we have that the fluxes are encouraged to distribute over more edges and with lower values, in order to keep traffic congestion low. Finally, when $\beta_\alpha = 1$ we obtain shortest path-like minimization. The consequence of having different β_α in different layers is that the optimal trajectories will have different topologies in each of the layers. At the same times, layers are coupled together, thus the final trajectories will be a complex combination of the weights w_α and the β_α. We give an example of optimal flows for various combinations of these parameters in Figure 2.
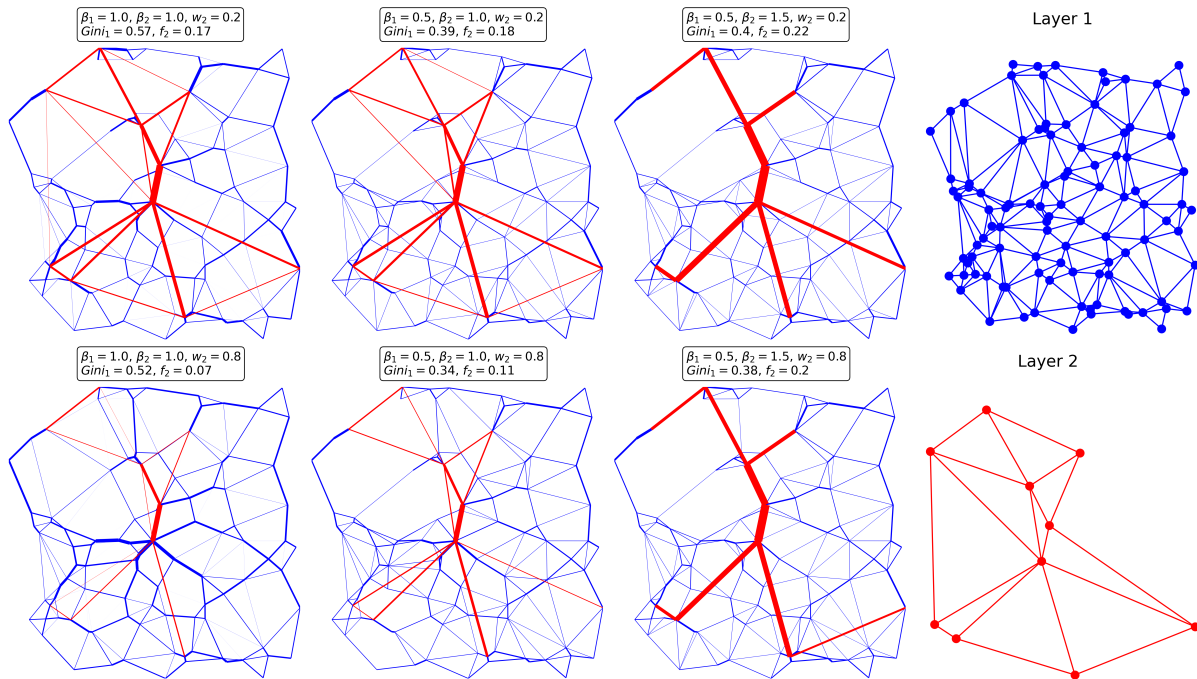
FIG. 2: Example of optimal paths. We show an example of optimal paths obtained with: $p = 0.2$ and (top) $w_1 = 0.2$, (bottom) $w_1 = 0.8$. Values of β_1, β_2 are those reported on top of each network. The statistics $Gini_1$ and f_2 are those defined in Section III A. The width of edges is proportional to the optimal $||F_e||_2$. Blue and red edges are for layer 1 and 2, respectively. The two layers are plotted individually on the rightmost column.

C. The algorithmic implementation

The numerical implementation consists in initializing the $\mu_e > 0$ at random. Then one iterates between i) extracting the pressure potentials (or the fluxes) using Eqs. (1) and (2); ii) use these to recompute the μ_e by means of Eq. (3), this can be solved numerically with a finite difference discretization. The iteration is repeated until convergence. In our experiments, we terminate a run of the algorithm when the difference $J_\beta^{(t+1)} - J_\beta^{(t)}$ between two successive updates is lower than a threshold (the superscript (t) is the iteration step). The cost J_β in Eq. (4) is not strictly convex in general, hence the solution of Algorithm 1 may converge to a local optima. One should then run the algorithm several times, each time initializing to a different random initial realization of $\mu_e > 0$. A possible choice for a final optimal solution would be the one that has lower J_β. We give the pseudocode for this in Algorithm 1. Most of the computational effort required by Algorithm 1 is in the solution of M linear systems as in Eq. (2). In our implementation, this has been performed by a sparse direct solver (UMFPACK), performing a LU decomposition of each column of the right hand side of Eq. (2), and having complexity scaling as $\mathcal{O}(M\,N^2)$.

Algorithm 1 Multilayer Optimal Transport

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

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

III. Results

A. Results on synthetic data

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

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

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

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

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

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

Looking at Figure 3, we see that $Gini$ increases with β_2 and thus the network usage becomes more hierarchical, as expected in this case (we report here results for $Gini$ w.r.t. the flux, but similar results are observed for FBC, see Figure S1). The exact value of $Gini$ depends on the travel demand, as for $p = 0.2$, i.e. when the central node is a destination in 80% of the journeys, $Gini$ is higher than when $p = 0.8$. This is because with fewer destinations there are also fewer possible path trajectories, and thus more passengers use the same part of the network. We can also see how $Gini$ decreases for higher w_2, i.e. when traveling by tram is not much faster than traveling on the road network. Finally, we can notice the drop in $Gini$ compared to the shortest path-like scenario $\beta_1 = \beta_2 = 1$. In this case, the traffic distribution is the most hierarchical, suggesting that possible traffic congestions can be avoided by setting lower values of β_1.
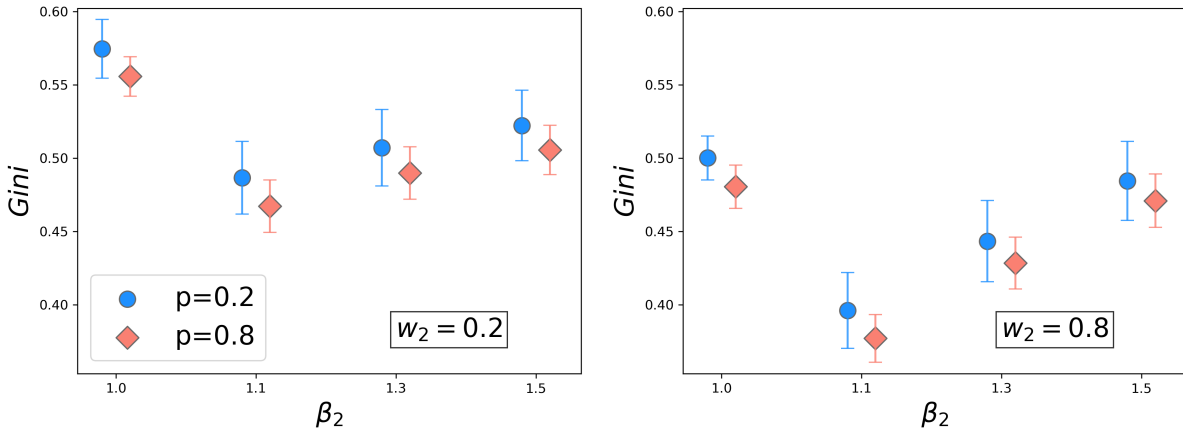
FIG. 3: Results on synthetic data. We show the $Gini$ w.r.t. the optimal $||F_e||_2$ (y axis) vs β_2 (x axis) for synthetic 2-layer networks generated as in Section III A. Blue and red markers denote $p = 0.2, 0.8$, respectively, $w_1 = 1$ in all cases, while $w_2 = 0.2$ (left) and $w_2 = 0.8$ (right); $\beta_1 = 0.5$ in all cases, except for the case where $\beta_2 = 1$, for which $\beta_1 = 1$. This case is a shortest path-like baseline. Markers are averages over 20 network samples and 50 source matrix samples (for a total of 1000 individual samples).

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

B. Results on real data

We illustrate our model on a real 2-layer network of the city of Bordeaux, where the two layers are the bus and tram, respectively. Data are taken from [49]. We simulated a monocentric source matrix S, i.e. $p = 0.0$, to asses the scenario where all the passengers travel to the city center, however results are similar for other values of p (not reported here). Optimal paths are extracted using our model for $\beta_1 = 0.5$, $\beta_2 = 1.5$, $w_2 = 0.2$ and compared against the case where the tram network is absent. This can be simulated by setting a high value of w_2, so that the cost on the tram edges makes it extremely unlikely to use any tram connection (here we used $w_2 = 100$). We measure the total percentage flux $f_2 = \sum_{e \in \mathcal{E}_2} ||F_e||_2 / (\sum_{e \in \mathcal{E}_1} ||F_e||_1 + \sum_{e \in \mathcal{E}_2} ||F_e||_2)$ passing through layer 2. Remarkably, in this scenario the tram network absorbs $f_2 = 17\%$ of the total flow of passengers, even though the tram network contains only $E_2 = 112$ edges, compared to $E_1 = 2347$ bus edges. This allows to reduce significantly the traffic along the road network, as can be seen in Figure 4: the road edges, and in particular those parallel to the tram line and close to the city center, get thinner, as more passengers use the tram. This also results in a higher $Gini_1 = 0.26$ (calculated on edges in layer 1 w.r.t. $||F_e||_2$), compared to the $Gini_1 = 0.23$ when the tram is absent: as the passengers use the tram, they decrease traffic on many road edges. While the traffic distribution on layer 1 gets more hierarchical (higher $Gini_1$), this does not necessarily lead to more traffic congestion. In fact, the total percentage flow f_1 decreases, as we saw above. Additional plots can be seen in Figure S2.
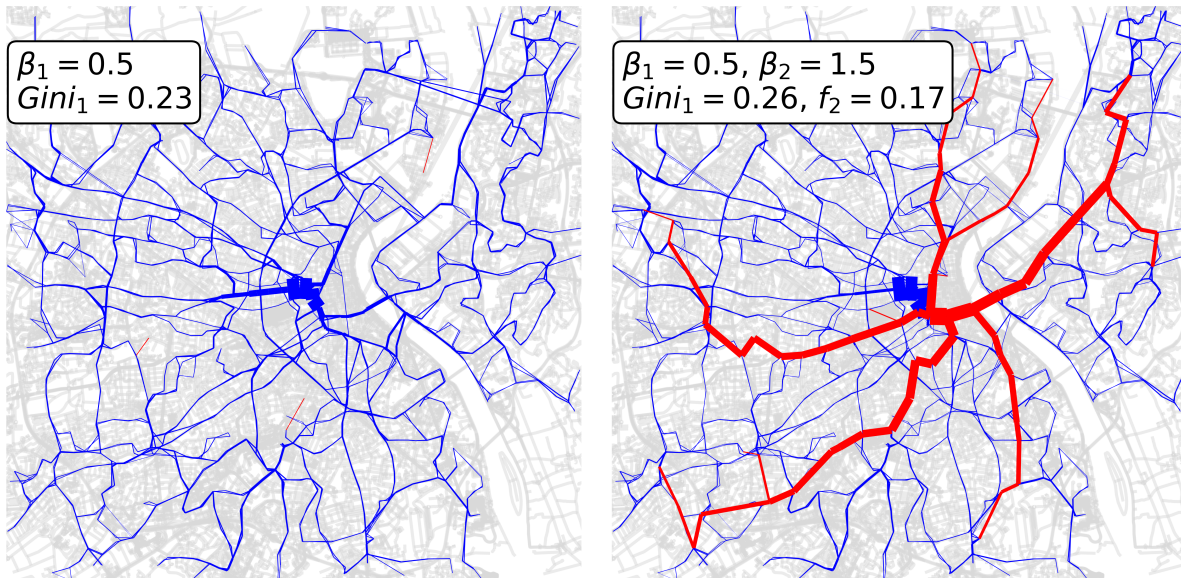
FIG. 4: Example of optimal paths in the city of Bordeaux for a bus and tram network. The paths are obtained with (left) and without (right) the tram layer. Here $\beta_1 = 0.5$ in both cases, while $\beta_2 = 1.5$ in the second case. The width of the edges is proportional to the optimal $||F_e||_2$. The reported $Gini_1$ coefficient for the bus network (layer 1) is calculated using $||F_e||_2$. The total percentage flux $f_2 = \sum_{e \in \mathcal{E}_2} ||F_e||_2 / (\sum_{e \in \mathcal{E}_1} ||F_e||_1 + \sum_{e \in \mathcal{E}_2} ||F_e||_2) = 0.17$, distributed over $E_2 = 112$ tram edges, compared to $E_1 = 2347$ bus edges.

IV. Discussion

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

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

[1] S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin, Physics reports **544**, 1 (2014).

[2] J. Gómez-Gardenes, I. Reinares, A. Arenas, and L. M. Floría, Scientific reports **2**, 1 (2012).

[3] J. F. Donges, H. C. Schultz, N. Marwan, Y. Zou, and J. Kurths, The European Physical Journal B **84**, 635 (2011).

[4] A. Saumell-Mendiola, M. Á. Serrano, and M. Boguná, Physical Review E **86**, 026106 (2012).

[5] M. Dickison, S. Havlin, and H. E. Stanley, Physical Review E **85**, 066109 (2012).

[6] J. Chen, M.-B. Hu, and M. Li, Physical Review E **101**, 012301 (2020).

[7] M. Kurant and P. Thiran, Physical review letters **96**, 138701 (2006).

[8] J. Wu, C. Pu, L. Li, and G. Cao, Digital Communications and Networks **6**, 58 (2020).

[9] Y. Zhuo, Y. Peng, C. Liu, Y. Liu, and K. Long, Physica A: Statistical Mechanics and its Applications **390**, 2401 (2011).

[10] A. Aleta and Y. Moreno, Annual Review of Condensed Matter Physics **10**, 45 (2019).

[11] G. Bianconi, *Multilayer networks: structure and function* (Oxford university press, 2018).

[12] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, Journal of complex networks **2**, 203 (2014).

[13] E. Strano, S. Shai, S. Dobson, and M. Barthelemy, Journal of The Royal Society Interface **12**, 20150651 (2015).

[14] A. Solé-Ribalta, S. Gómez, and A. Arenas, Physical review letters **116**, 108701 (2016).

[15] A. Aleta, S. Meloni, and Y. Moreno, Scientific reports **7**, 1 (2017).

[16] S. Gomez, A. Diaz-Guilera, J. Gomez-Gardenes, C. J. Perez-Vicente, Y. Moreno, and A. Arenas, Physical review letters **110**, 028701 (2013).

[17] M. De Domenico, C. Granell, M. A. Porter, and A. Arenas, Nature Physics **12**, 901 (2016).

[18] M. De Domenico, A. Solé-Ribalta, S. Gómez, and A. Arenas, Proceedings of the National Academy of Sciences **111**, 8351 (2014).

[19] M. Barthélemy, Physics Reports **499**, 1 (2011).

[20] A. Solé-Ribalta, A. Arenas, and S. Gómez, New Journal of Physics **21**, 035003 (2019).

[21] A. Lampo, J. Borge-Holthoefer, S. Gómez, and A. Solé-Ribalta, Physical Review Research **3**, 013267 (2021).

[22] D. Quercia, R. Schifanella, and L. M. Aiello, in *Proceedings of the 25th ACM conference on Hypertext and social media* (2014) pp. 116–125.

[23] L. Kantorovich, "On the transfer of masses (in russian)," (1942).

[24] V. Bonifaci, K. Mehlhorn, and G. Varma, Journal of Theoretical Biology **309**, 121 (2012).

[25] F. Santambrogio, Interfaces and Free Boundaries **9**, 149 (2007).

[26] E. Facca, F. Cardin, and M. Putti, SIAM Journal on Applied Mathematics **78**, 651 (2016).

[27] E. Facca, S. Daneri, F. Cardin, and M. Putti, Journal of Scientific Computing **82**, 68 (2020).

[28] E. Facca, F. Cardin, and M. Putti, "Branching structures emerging from a continuous optimal transport model," arXiv:1811.12691.

[29] D. Baptista, D. Leite, E. Facca, M. Putti, and C. De Bacco, Scientific Reports **10**, 088702 (2020).

[30] 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.

[31] J. B. Kirkegaard and K. Sneppen, Phys. Rev. Lett. **124**, 208101 (2020).

[32] S. Bohn and M. O. Magnasco, Phys. Rev. Lett. **98**, 088702 (2007).

[33] J. R. Banavar, A. Maritan, and A. Rinaldo, Nature **399**, 130 (1999).

[34] D. Hu and D. Cai, Physical review letters **111**, 138701 (2013).

[35] H. Ronellenfitsch and E. Katifori, Physical review letters **117**, 138301 (2016).

[36] E. Katifori, G. J. Szöllősi, and M. O. Magnasco, Phys. Rev. Lett. **104**, 048704 (2010).

[37] H. Ronellenfitsch and E. Katifori, Phys. Rev. Lett. **123**, 248101 (2019).

[38] D. Baptista and C. De Bacco, arXiv preprint arXiv:2012.12758 (2020), arXiv:2012.12758 [cs.CV].

[39] F. Kaiser, H. Ronellenfitsch, and D. Witthaut, Nature communications **11**, 1 (2020).

[40] A. Lonardi, E. Facca, M. Putti, and C. De Bacco, arXiv preprint arXiv:2010.14377 (2020), arXiv:2010.14377 [physics.soc-ph].

[41] A. Halu, S. Mukherjee, and G. Bianconi, Physical Review E **89**, 012806 (2014).

[42] R. G. Morris and M. Barthelemy, Physical review letters **109**, 128703 (2012).

[43] M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. A. Porter, S. Gómez, and A. Arenas, Physical Review X **3**, 041022 (2013).

[44] A. Tero, S. Takagi, T. Saigusa, K. Ito, D. P. Bebber, M. D. Fricker, K. Yumiki, R. Kobayashi, and T. Nakagaki, Science **327**, 439 (2010).

[45] L. Guibas and J. Stolfi, ACM transactions on graphics (TOG) **4**, 74 (1985).

[46] M. E. Newman, Social networks **27**, 39 (2005).

[47] U. Brandes and D. Fleischer, in *Annual symposium on theoretical aspects of computer science* (Springer, 2005) pp. 533–544.

[48] P. M. Dixon, J. Weiner, T. Mitchell-Olds, and R. Woodley, Ecology **68**, 1548 (1987).

[49] R. Kujala, C. Weckström, R. K. Darst, M. N. Mladenović, and J. Saramäki, Scientific data **5**, 1 (2018).

[50] L. Alexander, S. Jiang, M. Murga, and M. C. González, Transportation research part c: emerging technologies **58**, 240 (2015).

[51] P. Wang, T. Hunter, A. M. Bayen, K. Schechtner, and M. C. González, Scientific reports **2**, 1 (2012).

[52] L. Gao, P. Shu, M. Tang, W. Wang, and H. Gao, Physical Review E **100**, 012310 (2019).

[53] L. G. N. Orozco, F. Battiston, G. Iniguez, and M. Szell, Transport Findings **13171** (2020).
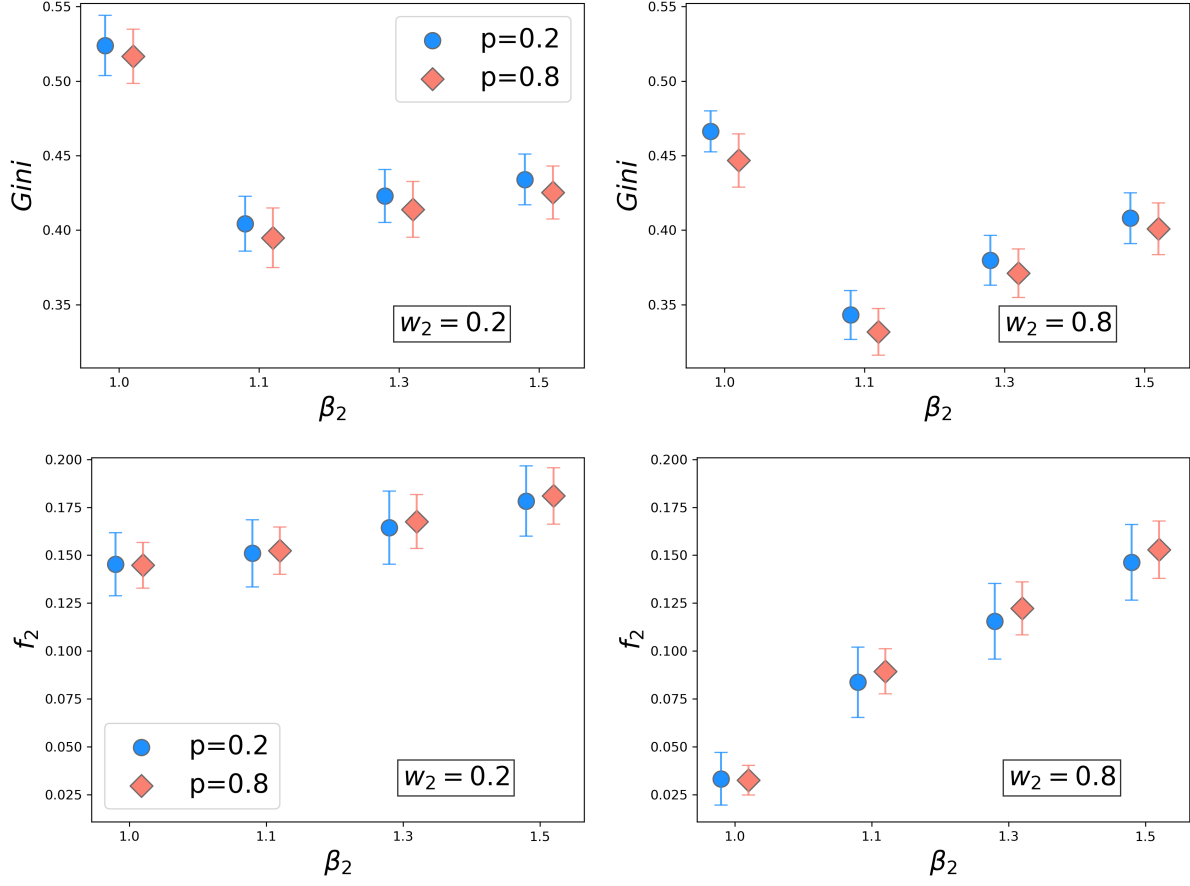
Supporting Information (SI)

FIG. S1: Additional results on synthetic data. We show the $Gini$ w.r.t. the optimal FBC (top) and the total percentage flux f_2 on layer 2 (bottom) vs β_2 (x axis), for synthetic 2-layer networks generated as in Section III A; $w_2 = 0.2, 0.8$ (left,right), $\beta_1 = 0.5$ in all cases, except for the case where $\beta_2 = 1$, for which $\beta_1 = 1$. This cases is a shortest path-like baseline. Markers are averages over 20 network samples and 50 source matrix samples.
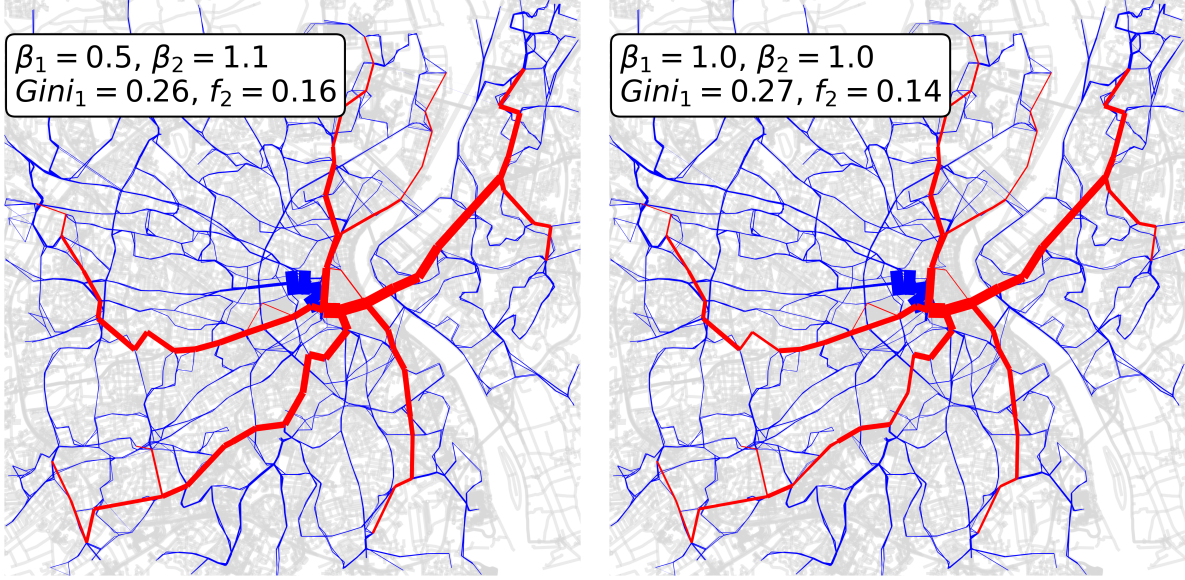
FIG. S2: Additional example of optimal paths in the city of Bordeaux for a bus and tram network. Here $p = 0.0$, $w_2 = 0.2$, $\beta_1, \beta_2 = (0.5, 1.1), (1.0, 1.0)$ (left and right). The width of the edges is proportional to the optimal $||F_e||_2$. $Gini_1$ is calculated w.r.t. to the flux on layer 1; $f_2 = \sum_{e \in \mathcal{E}_2} ||F_e||_2 / (\sum_{e \in \mathcal{E}_1} ||F_e||_1 + \sum_{e \in \mathcal{E}_2} ||F_e||_2)$.