Unsupervised learning for Optimal Transport plan prediction between unbalanced graphs

Sonia Mazelet, Rémi Flamary, Bertrand Thirion

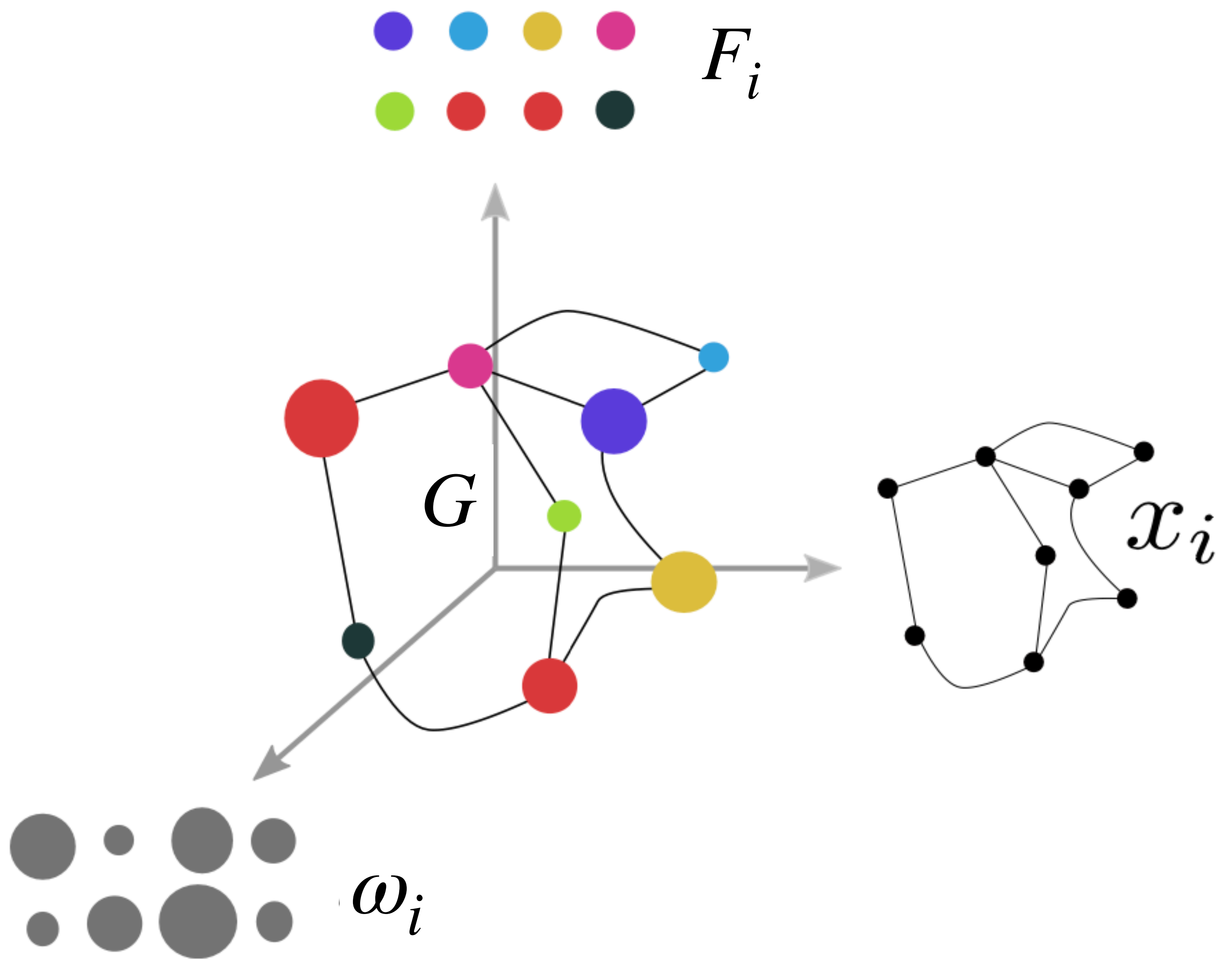
20/12/2025

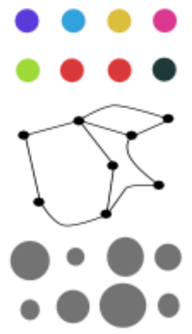
ÉCOLE POLYTECHNIQUE
INSTITUT POLYTECHNIQUE DE PARIS
Inria informatics mathematics

Optimal transport on graphs

Graphs modeled as probability measures [Vayer et al., 2019] characterized by:

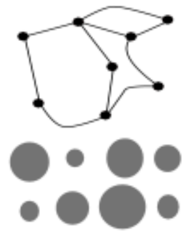
- geometry (adjacency matrix, shortest path distance matrix…): D
- node features: F
- node weights: ω





$$\mu = \sum_i \omega_i \delta_{(x_i, F_i)}$$

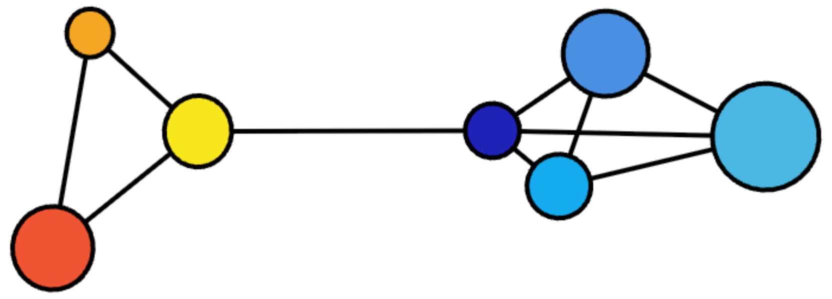
$$\mu_F = \sum_i \omega_i \delta_{F_i}$$



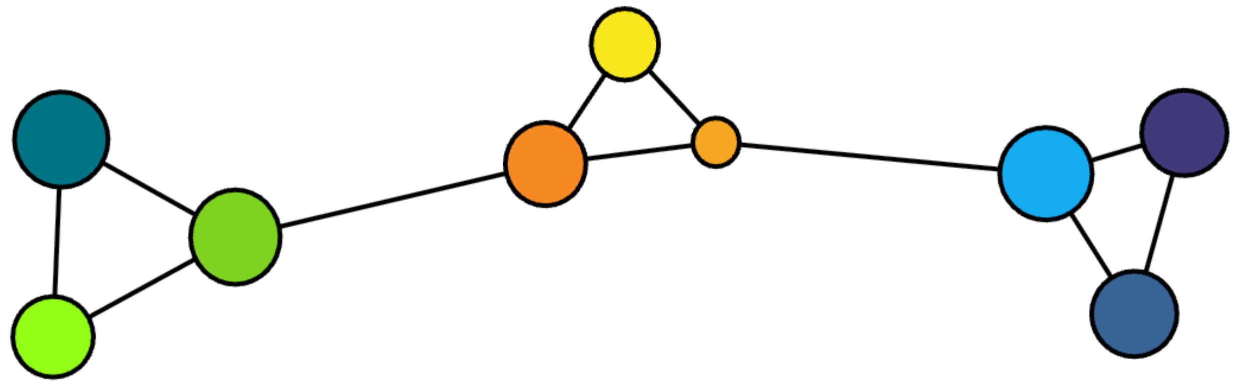
$$\mu_X = \sum_i \omega_i \delta_{x_i}$$

Graph matching

Goal: given a pair of graphs, find a matching between the nodes that preserves the graph geometry, node features and discards nodes that do not have a good matching.



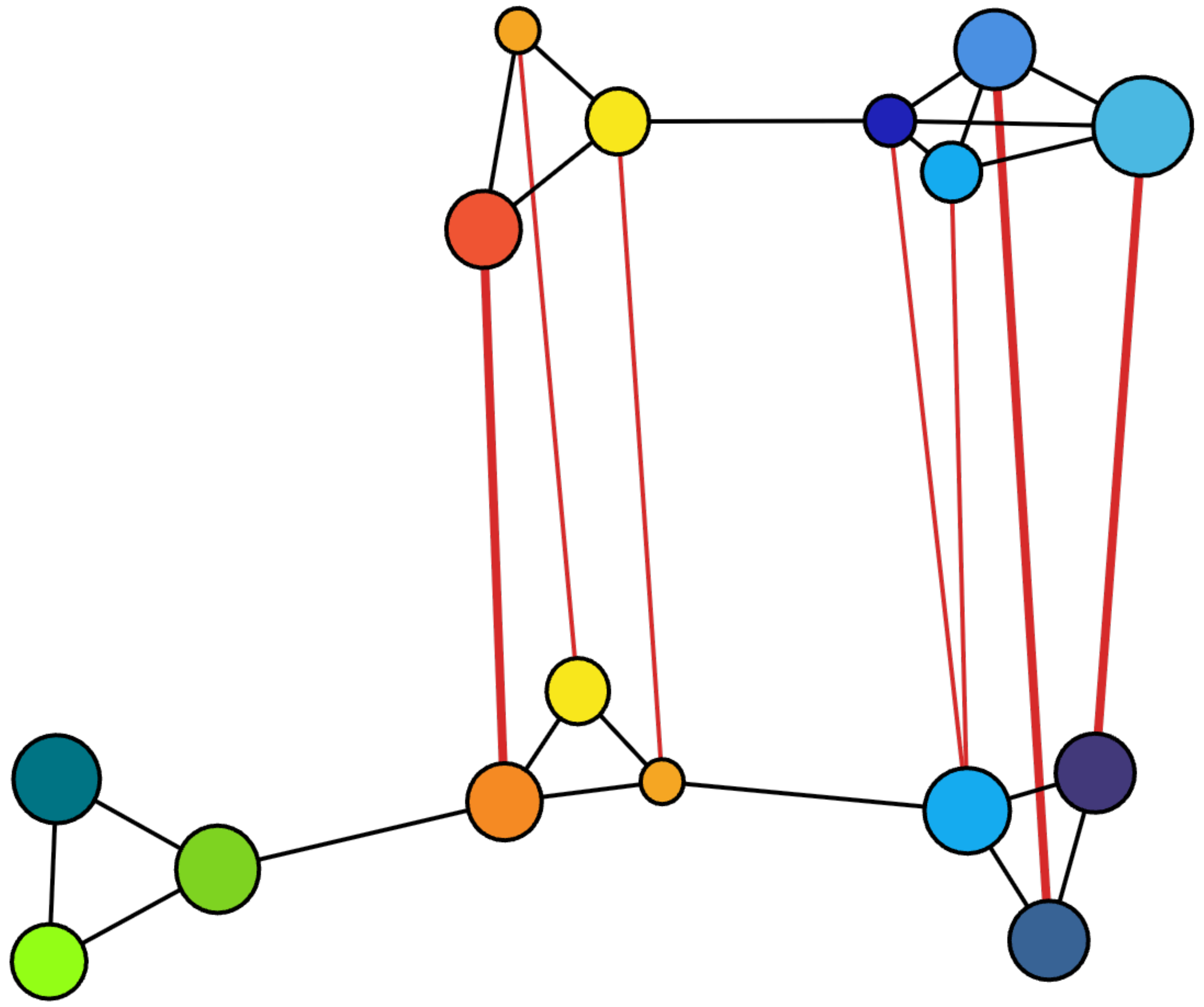
$$G_1 = (F_1, D_1, \omega_1)$$



$$G_2 = (F_2, D_2, \omega_2)$$

Graph matching

Goal: given a pair of graphs, find a matching between the nodes that preserves the graph geometry, node features and discards nodes that do not have a good matching.

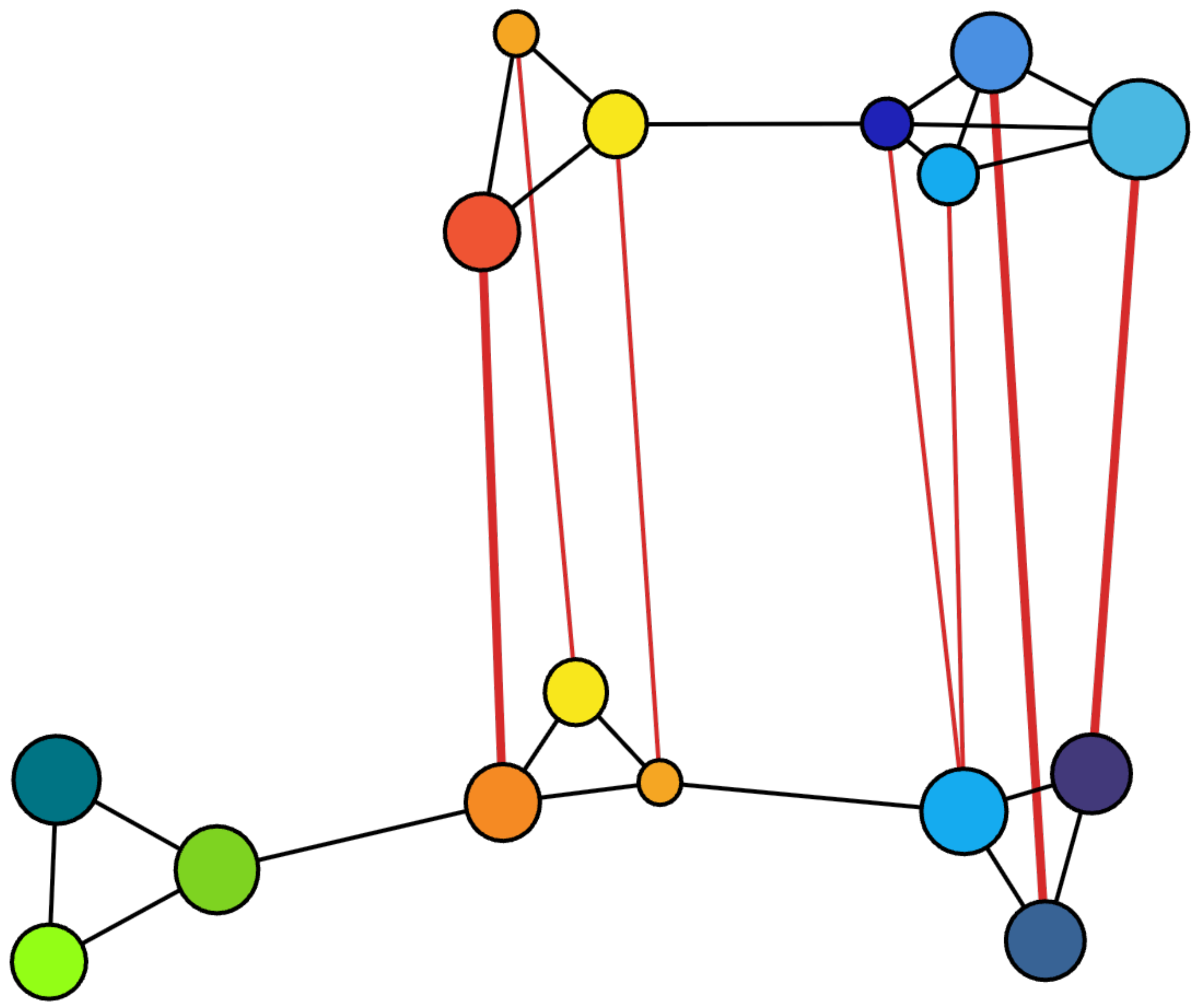


$$G_1 = (F_1, D_1, \omega_1)$$

$$G_2 = (F_2, D_2, \omega_2)$$

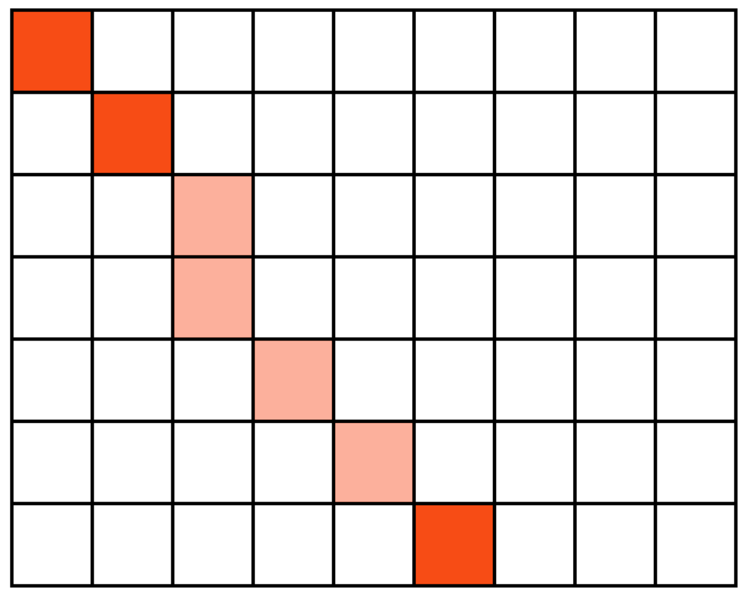
Graph matching

Goal: given a pair of graphs, find a matching between the nodes that preserves the graph geometry, node features and discards nodes that do not have a good matching.



optimal transport plan P:

$P_{i,j}$ = mass transported from $n_1(i)$ to $n_2(j)$

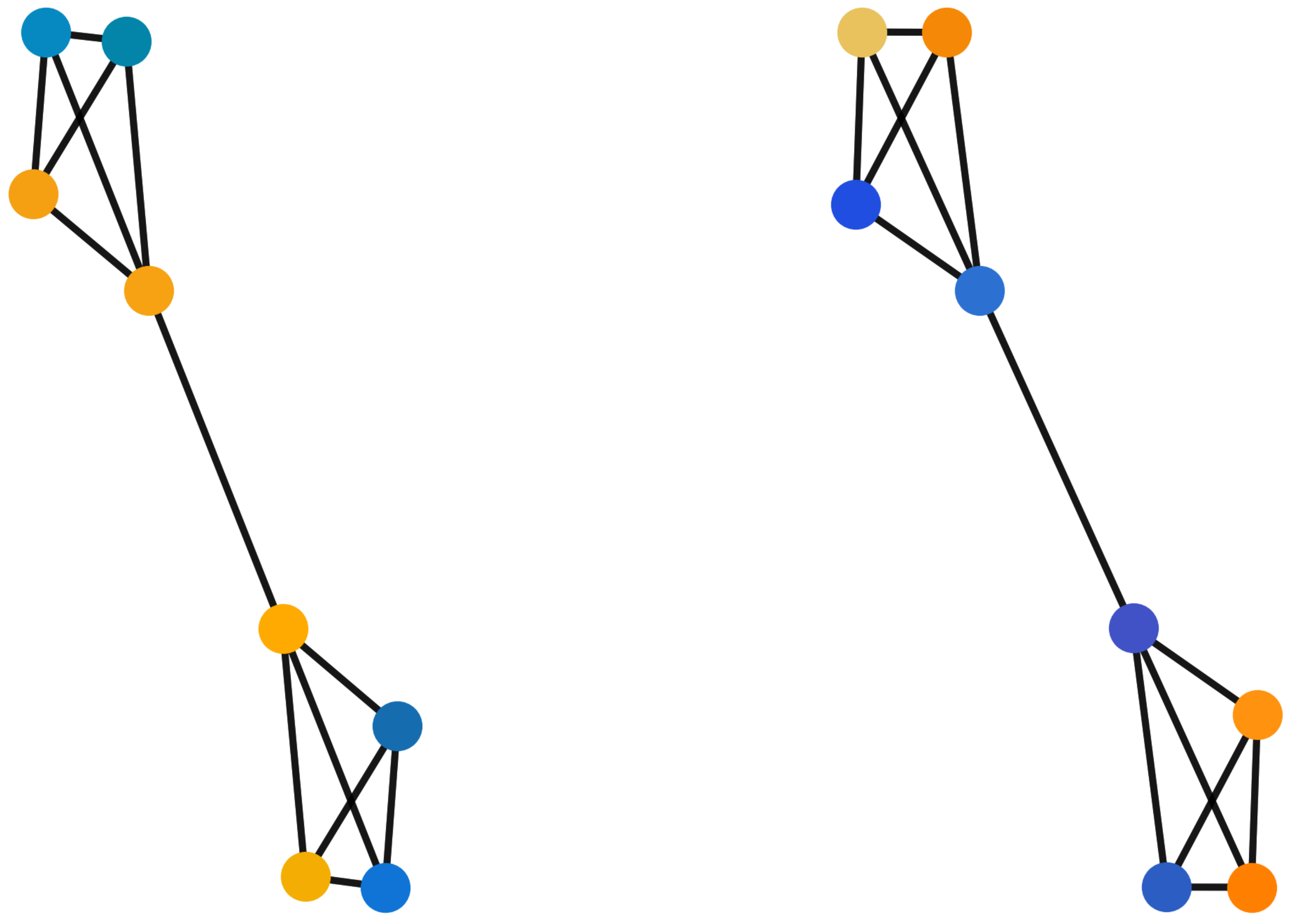


Optimal transport distance between graphs

Fused Unbalanced Gromov Wasserstein (FUGW) optimal transport loss [Thual et al., 2022]

$$\mathrm{L}^{\alpha,\rho}(G_1, G_2, \mathbf{P}) = (\mathbf{1}-\alpha)\sum_{i,j=1}^{\mathbf{n_1,n_2}} \left\|(\mathbf{F_1})_{\mathbf{i}} - (\mathbf{F_2})_{\mathbf{j}}\right\|_{\mathbf{2}}^{\mathbf{2}} \mathbf{P_{i,j}} + \alpha \sum_{i,j,k,l=1}^{\mathbf{n_1,n_2,n_1,n_2}} \left|(\mathbf{D_1})_{\mathbf{i,k}} - (\mathbf{D_2})_{\mathbf{j,l}}\right|^{\mathbf{2}} \mathbf{P_{i,j}P_{k,l}} + \rho\left(\mathbf{KL}(\mathbf{P_{\#1}} \otimes \mathbf{P_{\#1}} | \omega_1 \otimes \omega_1) + \mathbf{KL}(\mathbf{P_{\#2}} \otimes \mathbf{P_{\#2}} | \omega_2 \otimes \omega_2)\right)$$

match nodes with similar node features

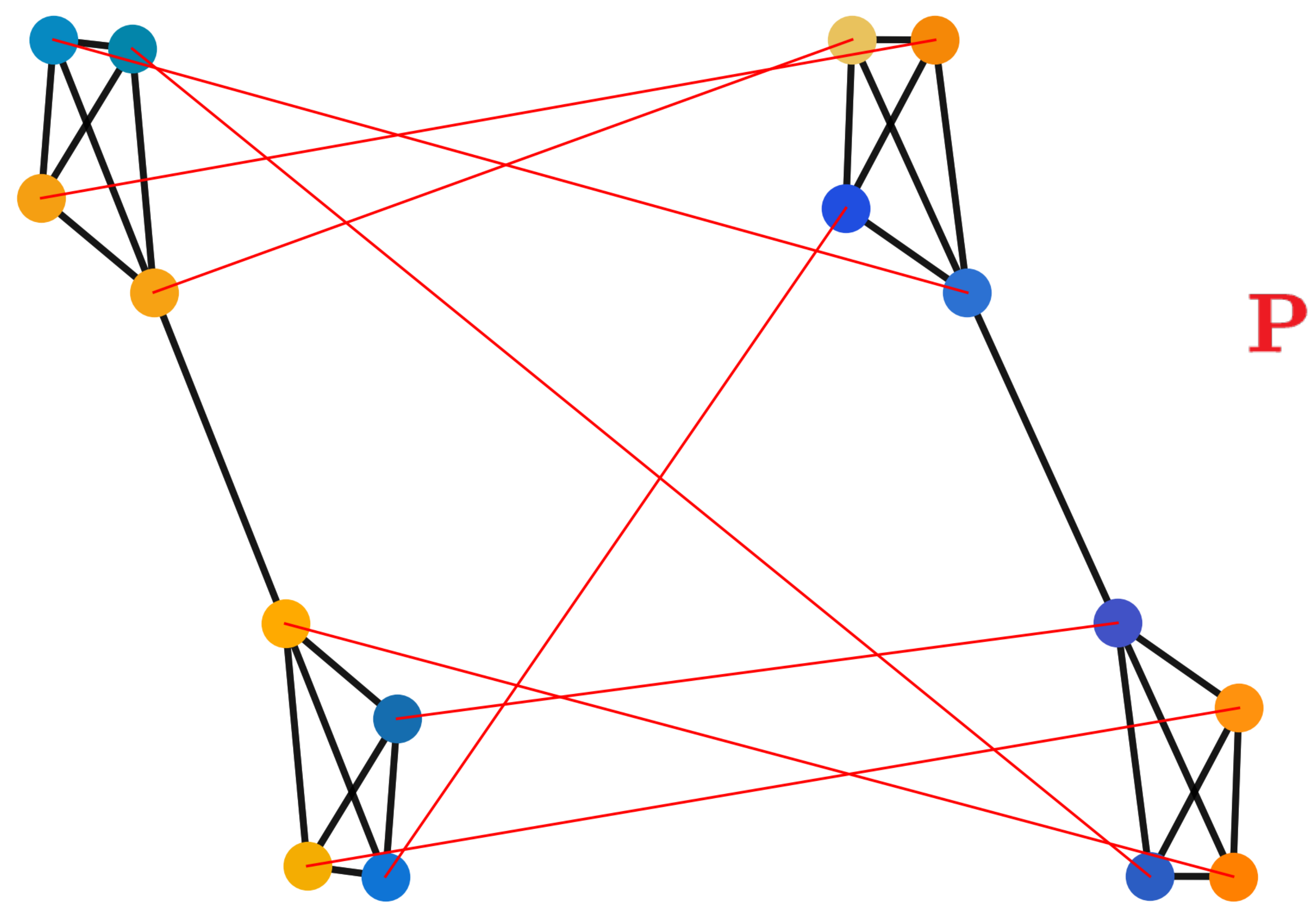


Optimal transport distance between graphs

Fused Unbalanced Gromov Wasserstein (FUGW) optimal transport loss [Thual et al., 2022]

$$\mathrm{L}^{\alpha,\rho}(G_1, G_2, \mathbf{P}) = (\mathbf{1}-\alpha)\sum_{i,j=1}^{\mathbf{n_1},\mathbf{n_2}} \left\|(\mathbf{F_1})_{\mathbf{i}} - (\mathbf{F_2})_{\mathbf{j}}\right\|_{\mathbf{2}}^{\mathbf{2}} \mathbf{P_{i,j}} + \alpha \sum_{i,j,k,l=1}^{\mathbf{n_1},\mathbf{n_2},\mathbf{n_1},\mathbf{n_2}} \left|(\mathbf{D_1})_{\mathbf{i,k}} - (\mathbf{D_2})_{\mathbf{j,l}}\right|^{\mathbf{2}} \mathbf{P_{i,j}}\mathbf{P_{k,l}} + \rho\left(\mathbf{KL}(\mathbf{P}_{\#1} \otimes \mathbf{P}_{\#1} | \omega_1 \otimes \omega_1) + \mathbf{KL}(\mathbf{P}_{\#2} \otimes \mathbf{P}_{\#2} | \omega_2 \otimes \omega_2)\right)$$

match nodes with similar node features



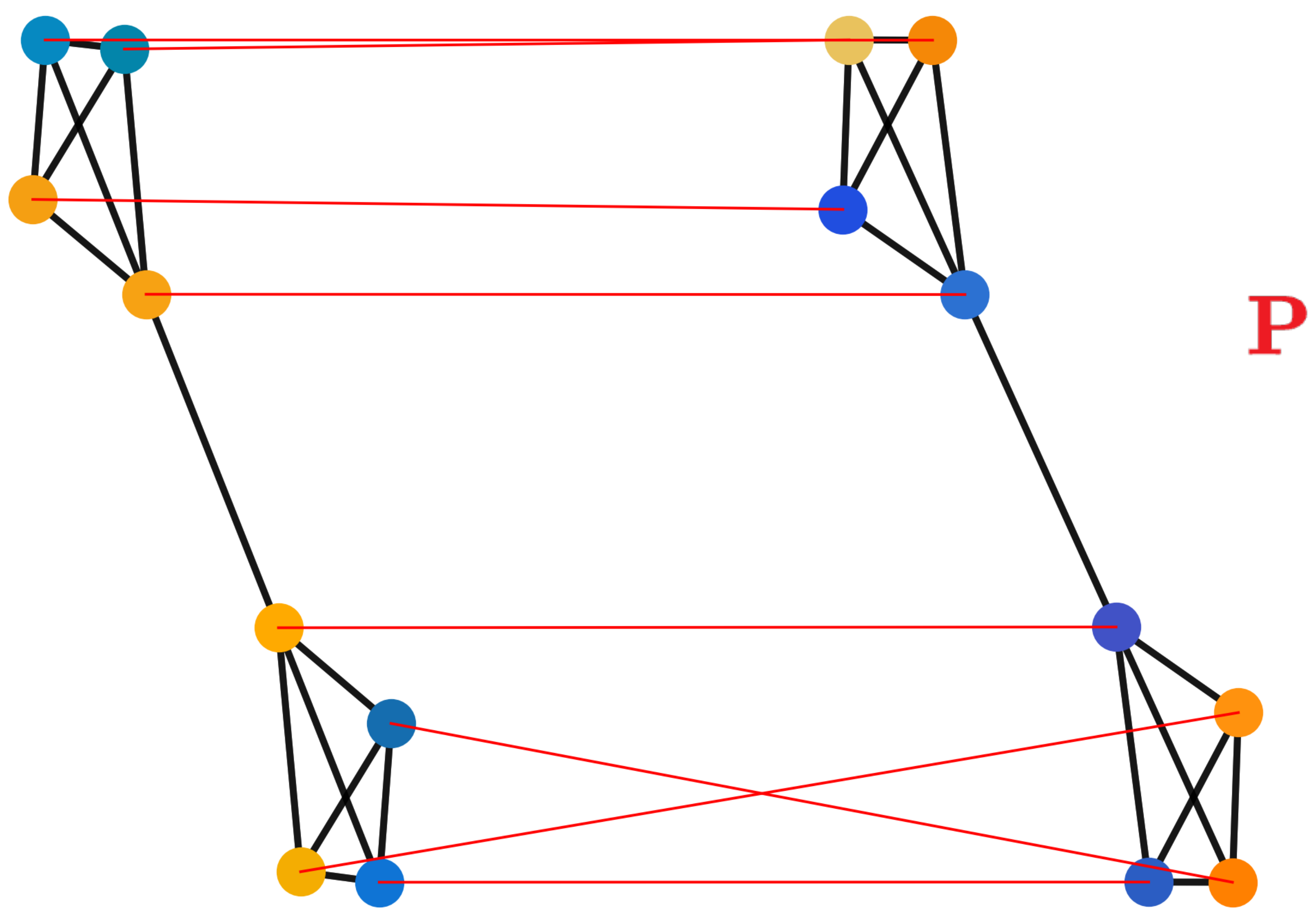
$$\mathbf{P} = \arg\min_{P \geq 0} \mathrm{L}^{\alpha,\rho}(G_1, G_2, \mathbf{P})$$

Optimal transport distance between graphs

Fused Unbalanced Gromov Wasserstein (FUGW) optimal transport loss [Thual et al., 2022]

$$\mathrm{L}^{\alpha,\rho}(G_1, G_2, \mathbf{P}) = (1-\alpha)\sum_{i,j=1}^{\mathbf{n_1},\mathbf{n_2}} \left\|(\mathbf{F}_1)_\mathbf{i} - (\mathbf{F}_2)_\mathbf{j}\right\|_2^2 \mathbf{P}_{\mathbf{i},\mathbf{j}} + \alpha \boxed{\sum_{i,j,k,l=1}^{\mathbf{n_1},\mathbf{n_2},\mathbf{n_1},\mathbf{n_2}} \left|(\mathbf{D}_1)_{\mathbf{i},\mathbf{k}} - (\mathbf{D}_2)_{\mathbf{j},\mathbf{l}}\right|^2 \mathbf{P}_{\mathbf{i},\mathbf{j}}\mathbf{P}_{\mathbf{k},\mathbf{l}}} + \rho\left(\mathbf{KL}(\mathbf{P}_{\#1} \otimes \mathbf{P}_{\#1} | \omega_1 \otimes \omega_1) + \mathbf{KL}(\mathbf{P}_{\#2} \otimes \mathbf{P}_{\#2} | \omega_2 \otimes \omega_2)\right)$$

preserve local geometry



$$\mathbf{P} = \arg\min_{P \geq 0} \mathrm{L}^{\alpha,\rho}(G_1, G_2, \mathbf{P})$$

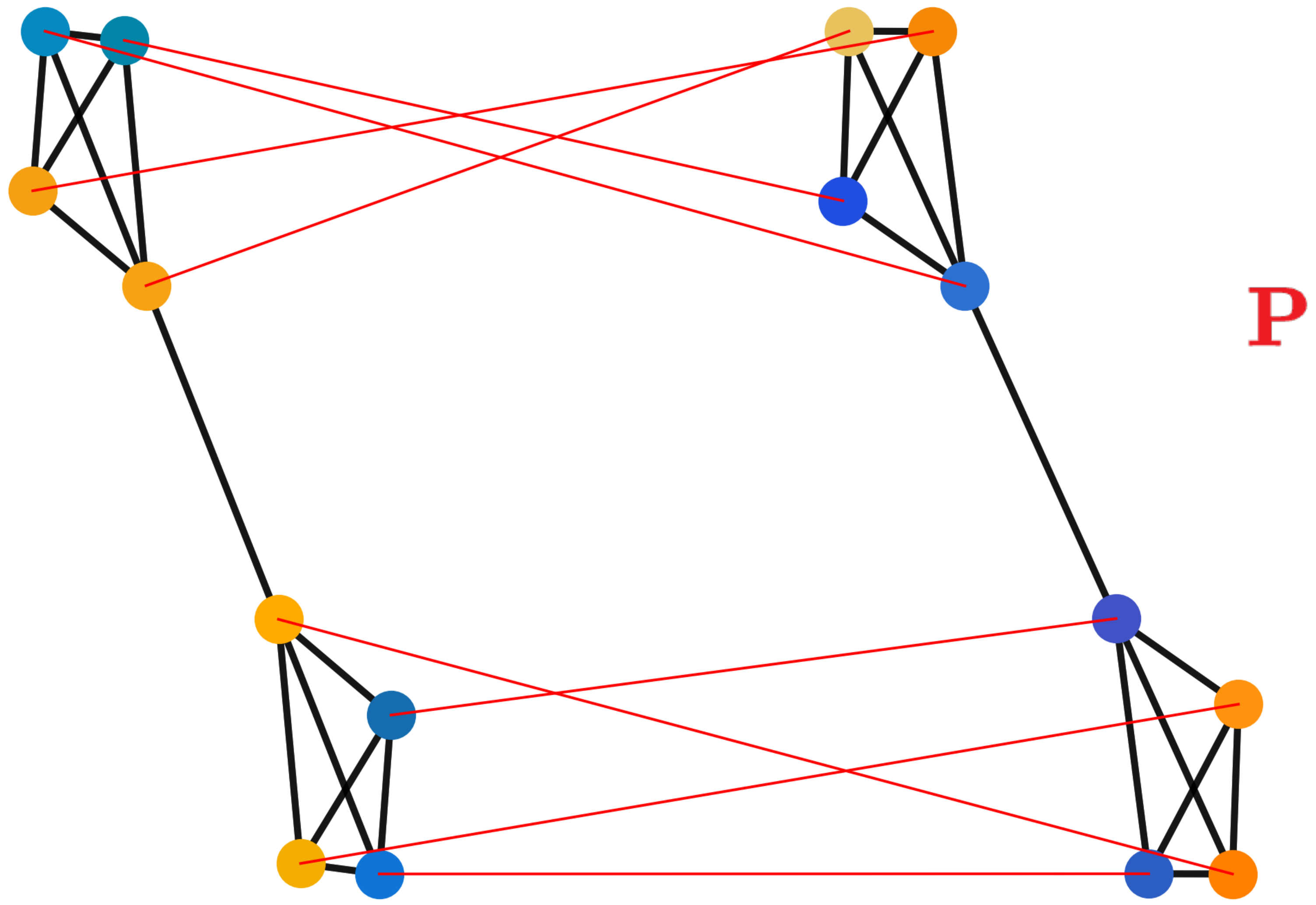
Optimal transport distance between graphs

Fused Unbalanced Gromov Wasserstein (FUGW) optimal transport loss [Thual et al., 2022]

$$\mathrm{L}^{\alpha,\rho}(G_1, G_2, \mathbf{P}) = (1-\alpha)\sum_{i,j=1}^{n_1,n_2} \left\|(\mathbf{F}_1)_i - (\mathbf{F}_2)_j\right\|_2^2 \mathbf{P}_{i,j} + \alpha \sum_{i,j,k,l=1}^{n_1,n_2,n_1,n_2} \left|(\mathbf{D}_1)_{i,k} - (\mathbf{D}_2)_{j,l}\right|^2 \mathbf{P}_{i,j}\mathbf{P}_{k,l} + \rho\left(\mathbf{KL}(\mathbf{P}_{\#1} \otimes \mathbf{P}_{\#1} | \omega_1 \otimes \omega_1) + \mathbf{KL}(\mathbf{P}_{\#2} \otimes \mathbf{P}_{\#2} | \omega_2 \otimes \omega_2)\right)$$

match nodes with similar node features

preserve local geometry

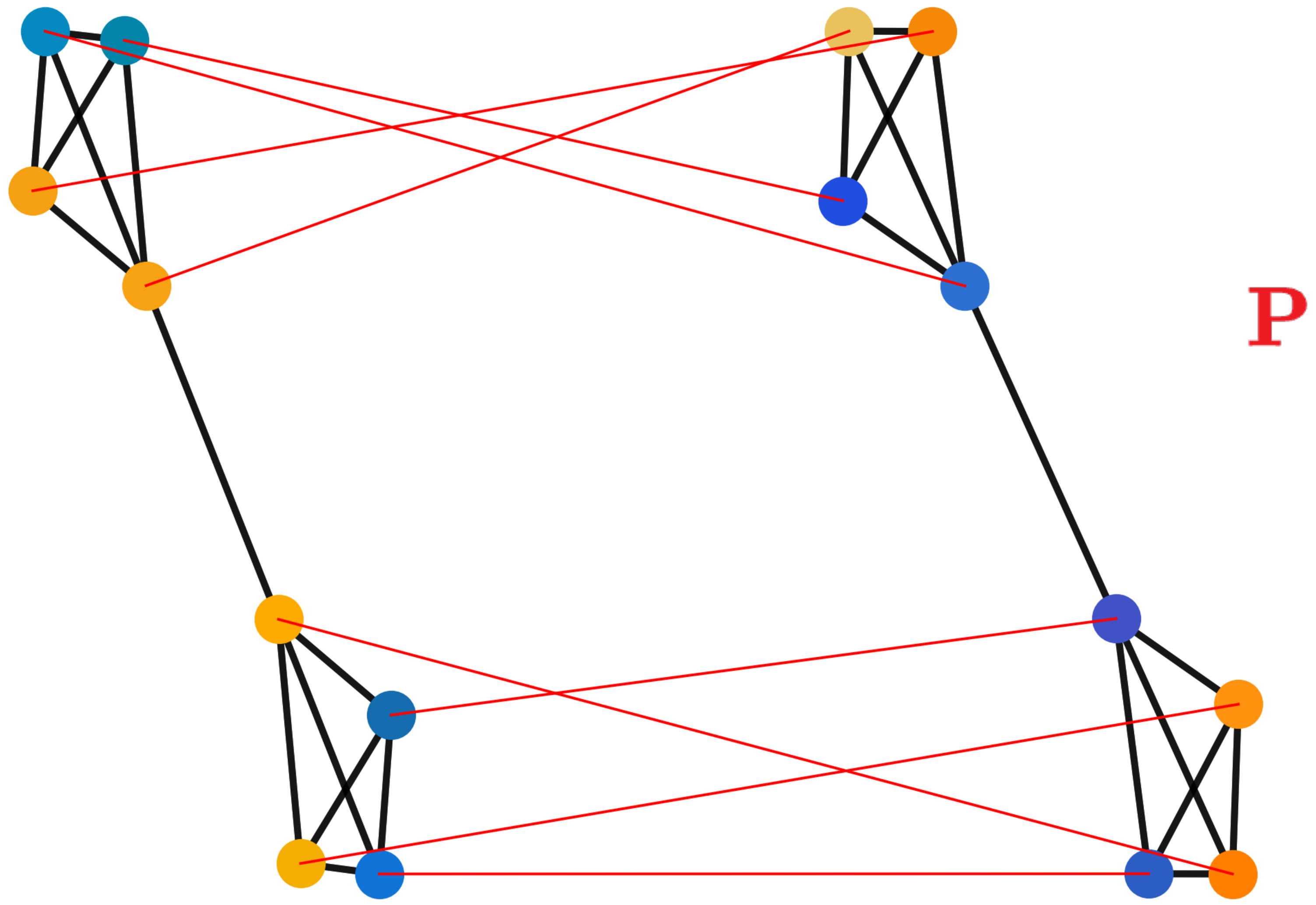


$$\mathbf{P} = \arg\min_{P \geq 0} \mathrm{L}^{\alpha,\rho}(G_1, G_2, \mathbf{P})$$

Optimal transport distance between graphs

Fused Unbalanced Gromov Wasserstein (FUGW) optimal transport loss [Thual et al., 2022]

$$\mathrm{L}^{\alpha,\rho}(G_1,G_2,\mathbf{P}) = (1-\alpha)\underbrace{\sum_{i,j=1}^{n_1,n_2}\left\|(\mathbf{F}_1)_i-(\mathbf{F}_2)_j\right\|_2^2\mathbf{P}_{i,j}}_{\text{match nodes with similar node features}} + \alpha\underbrace{\sum_{\mathbf{i,j,k,l=1}}^{\mathbf{n_1,n_2,n_1,n_2}}\left|(\mathbf{D_1})_{\mathbf{i,k}}-(\mathbf{D_2})_{\mathbf{j,l}}\right|^{\mathbf{2}}\mathbf{P_{i,j}P_{k,l}}}_{\text{preserve local geometry}} + \rho\left(\mathbf{KL}(\mathbf{P}_{\#1}\otimes\mathbf{P}_{\#1}|\omega_1\otimes\omega_1)+\mathbf{KL}(\mathbf{P}_{\#2}\otimes\mathbf{P}_{\#2}|\omega_2\otimes\omega_2)\right)$$



$$\mathbf{P} = \arg\min_{P\geq 0}\mathrm{L}^{\alpha,\rho}(G_1,G_2,\mathbf{P})$$

Optimal transport distance between graphs

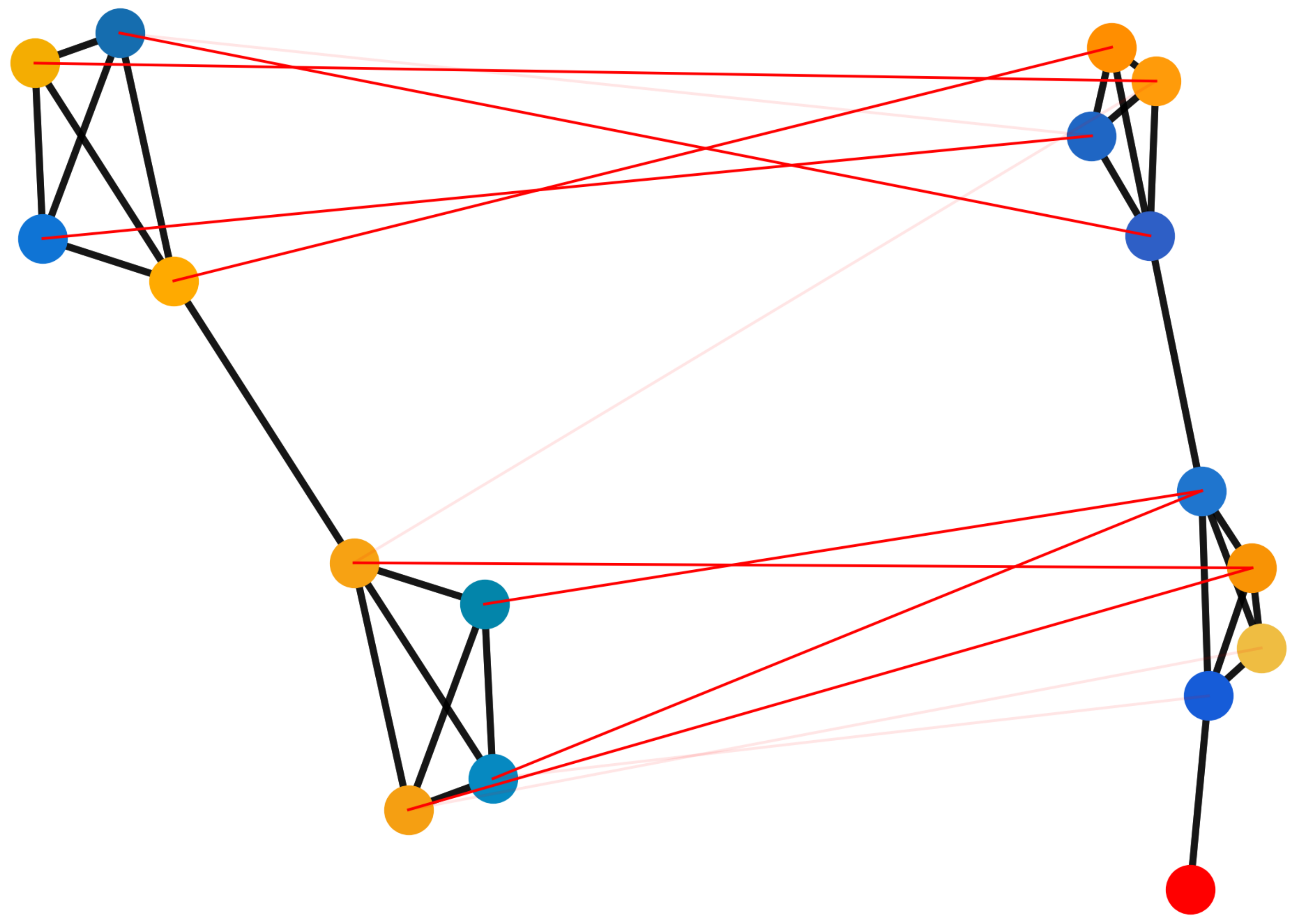
Fused Unbalanced Gromov Wasserstein (FUGW) optimal transport loss [Thual et al., 2022]

$$\mathrm{L}^{\alpha,\rho}(G_1, G_2, \mathbf{P}) = (\mathbf{1}-\alpha)\sum_{i,j=1}^{\mathbf{n_1,n_2}} \left\|(\mathbf{F_1})_\mathbf{i} - (\mathbf{F_2})_\mathbf{j}\right\|_\mathbf{2}^\mathbf{2} \mathbf{P_{i,j}} + \alpha \sum_{i,j,k,l=1}^{\mathbf{n_1,n_2,n_1,n_2}} \left|(\mathbf{D_1})_\mathbf{i,k} - (\mathbf{D_2})_\mathbf{j,l}\right|^\mathbf{2} \mathbf{P_{i,j}}\mathbf{P_{k,l}} + \rho\left(\mathbf{KL}(\mathbf{P}_{\#1} \otimes \mathbf{P}_{\#1} | \omega_\mathbf{1} \otimes \omega_\mathbf{1}) + \mathbf{KL}(\mathbf{P}_{\#2} \otimes \mathbf{P}_{\#2} | \omega_\mathbf{2} \otimes \omega_\mathbf{2})\right)$$

match nodes with similar node features

preserve local geometry

discard nodes that do not have a good match



$$\mathbf{P} = \arg\min_{P \geq 0} \mathrm{L}^{\alpha,\rho}(G_1, G_2, \mathbf{P})$$

Optimal transport distance between graphs

Fused Unbalanced Gromov Wasserstein (FUGW) optimal transport loss [Thual et al., 2022]

$$\mathrm{L}^{\alpha,\rho}(G_1, G_2, \mathbf{P}) = (\mathbf{1}-\alpha)\underbrace{\sum_{i,j=1}^{\mathbf{n_1},\mathbf{n_2}} \left\|(\mathbf{F_1})_\mathbf{i} - (\mathbf{F_2})_\mathbf{j}\right\|_\mathbf{2}^\mathbf{2} \mathbf{P_{i,j}}}_{\text{match nodes with similar node features}} + \alpha \underbrace{\sum_{i,j,k,l=1}^{\mathbf{n_1},\mathbf{n_2},\mathbf{n_1},\mathbf{n_2}} \left|(\mathbf{D_1})_\mathbf{i,k} - (\mathbf{D_2})_\mathbf{j,l}\right|^\mathbf{2} \mathbf{P_{i,j}}\mathbf{P_{k,l}}}_{\text{preserve local geometry}} + \rho \underbrace{(\mathbf{KL}(\mathbf{P_{\#1}} \otimes \mathbf{P_{\#1}} | \omega_\mathbf{1} \otimes \omega_\mathbf{1}) + \mathbf{KL}(\mathbf{P_{\#2}} \otimes \mathbf{P_{\#2}} | \omega_\mathbf{2} \otimes \omega_\mathbf{2}))}_{\text{discard nodes that do not have a good match}}$$

FUGW distance: $\mathrm{FUGW}^{\alpha,\rho}(G_1, G_2) = \min_{P \geq 0} \mathrm{L}^{\alpha,\rho}(G_1, G_2, \mathbf{P})$

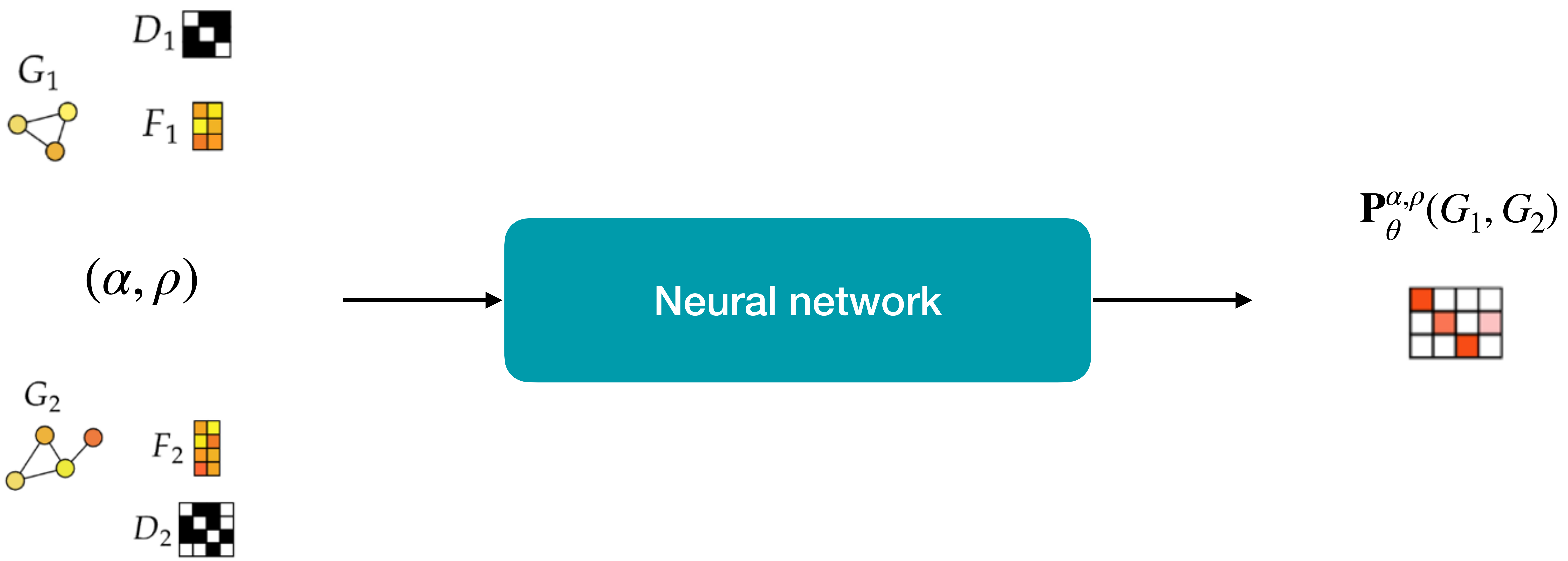
Solve the OT problem: batch coordinate descent with complexity $O(kn^3)$ for k the number of iterations and n the number of graph nodes.

→ unscalable for large graphs

Predicting FUGW plan

Goal: learn to predict FUGW plan $\mathbf{P}_\theta^{\alpha,\rho}(G_1, G_2)$ for all graph pairs $(G_1, G_2) \sim \mathscr{D}$ and parameters $(\alpha, \rho) \sim \mathscr{P}$.

Method: Neural Network based cross attention and Graph Convolutional Networks that predicts OT plans.



Training a graph matching neural network

- Most method are trained in a supervised way [Wang et al., 2019][Sarlin et al. 2020][Zanfir et al. 2020]: $\ell(\hat{M}, M)$ for ground truth matching M

- In practice, ground truth matching is hard to compute and most methods are only applicable to images.

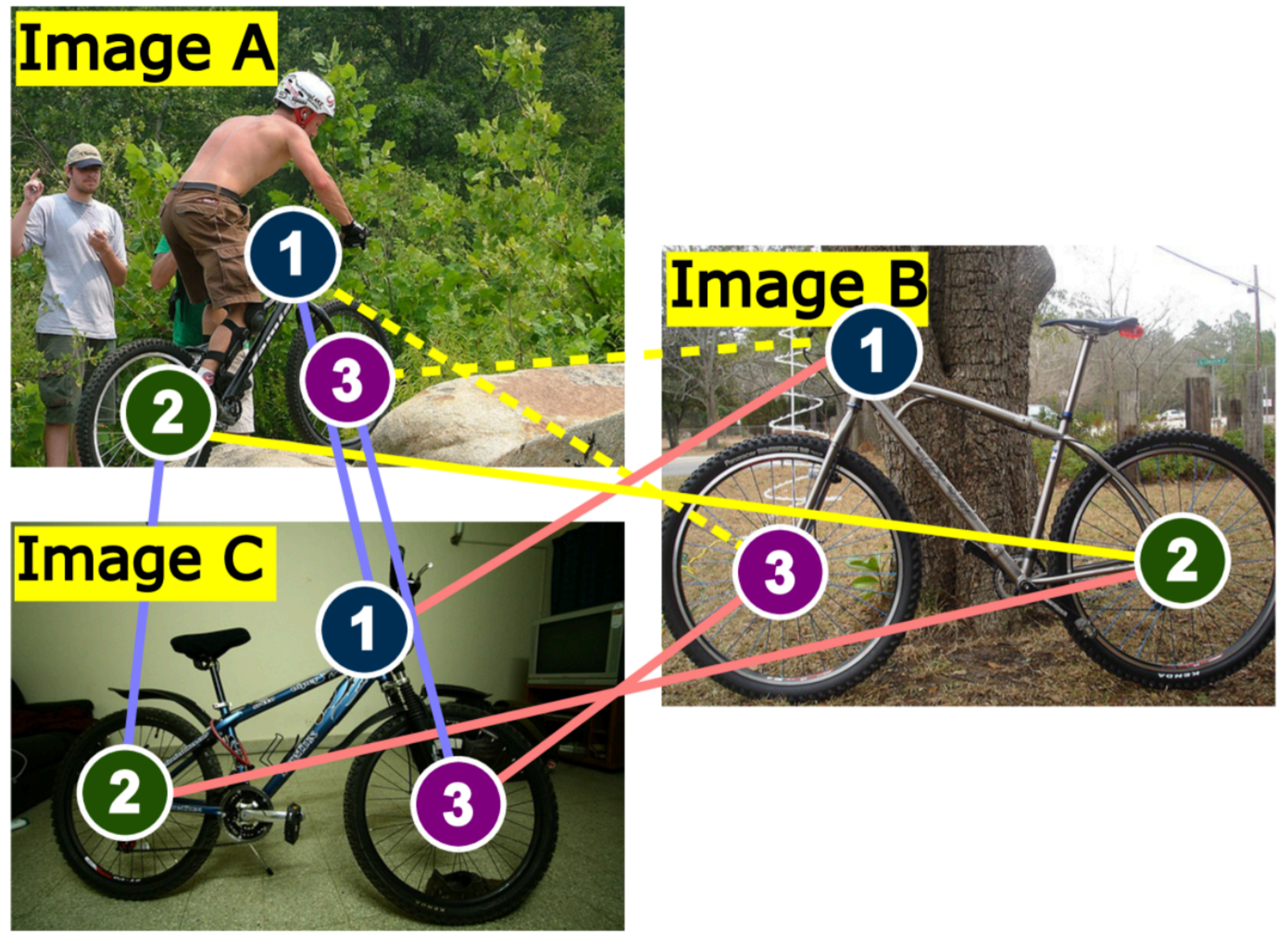
Training a graph matching neural network: a few unsupervised methods

- Learn to match copies of the same graph *[Liu et al. 2022]*

original image **augmented image**

- Minimize a criterion that is domain specific *[Tourani et al. 2024]*

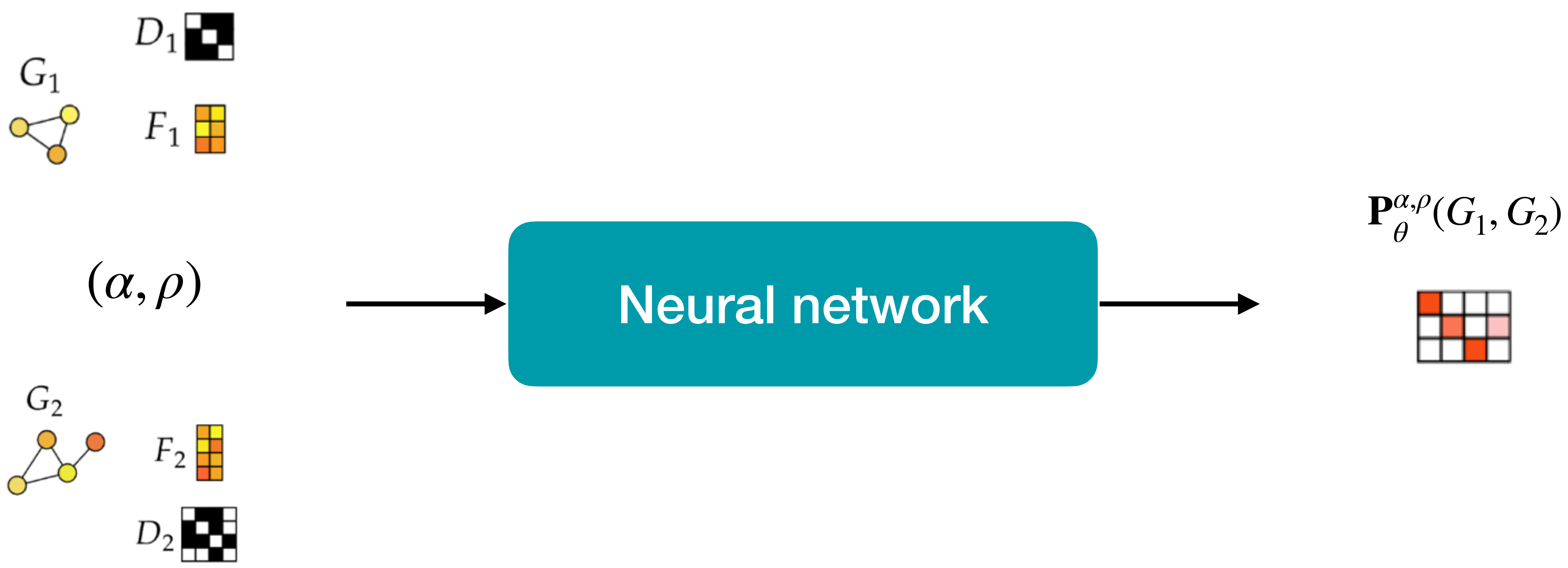
cycle consistency criterion



Predicting FUGW plan

Goal: learn to predict FUGW plan $\mathbf{P}_\theta^{\alpha,\rho}(G_1, G_2)$ for all graph pairs $(G_1, G_2) \sim \mathscr{D}$ and parameters $(\alpha, \rho) \sim \mathscr{P}$.

Method: Neural Network based cross attention and Graph Convolutional Networks that predicts OT plans.

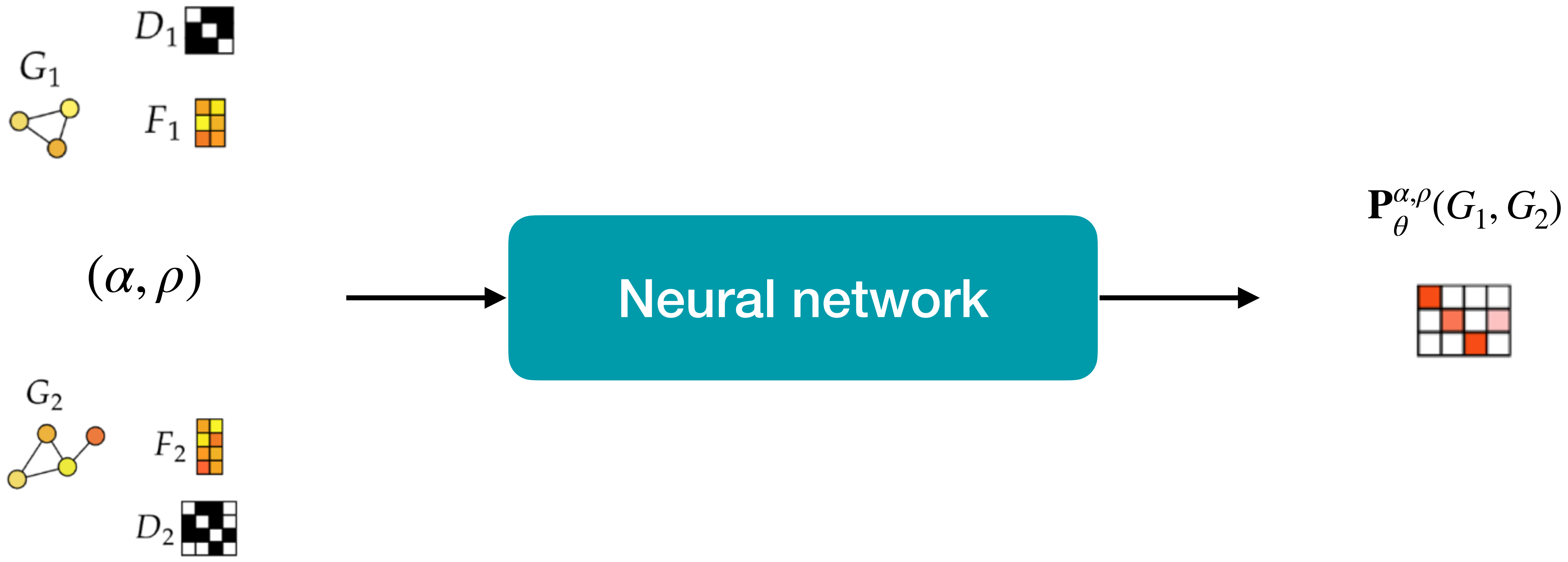


Supervised training? Unscalable

Predicting FUGW plan

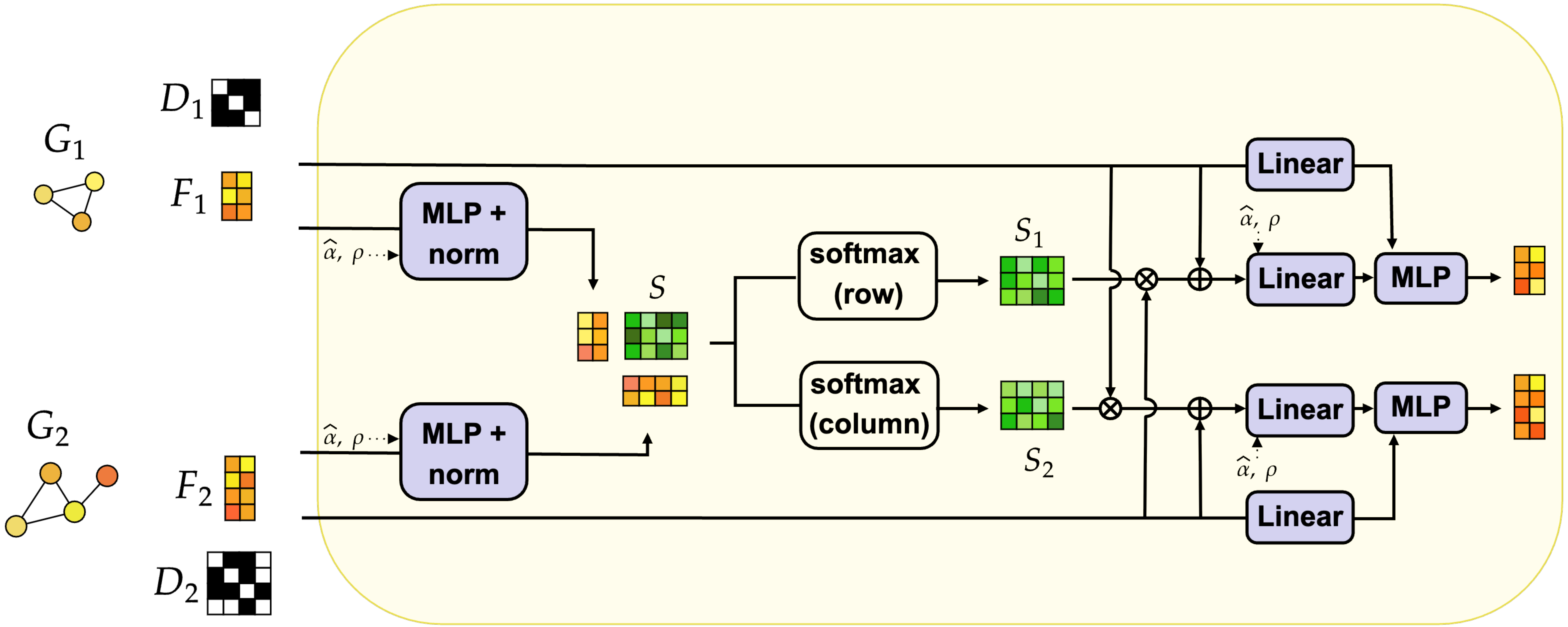
Goal: learn to predict FUGW plan $\mathbf{P}_\theta^{\alpha,\rho}(G_1, G_2)$ for all graph pairs $(G_1, G_2) \sim \mathscr{D}$ and parameters $(\alpha, \rho) \sim \mathscr{P}$.

Method: Neural Network based cross attention and Graph Convolutional Networks that predicts OT plans.

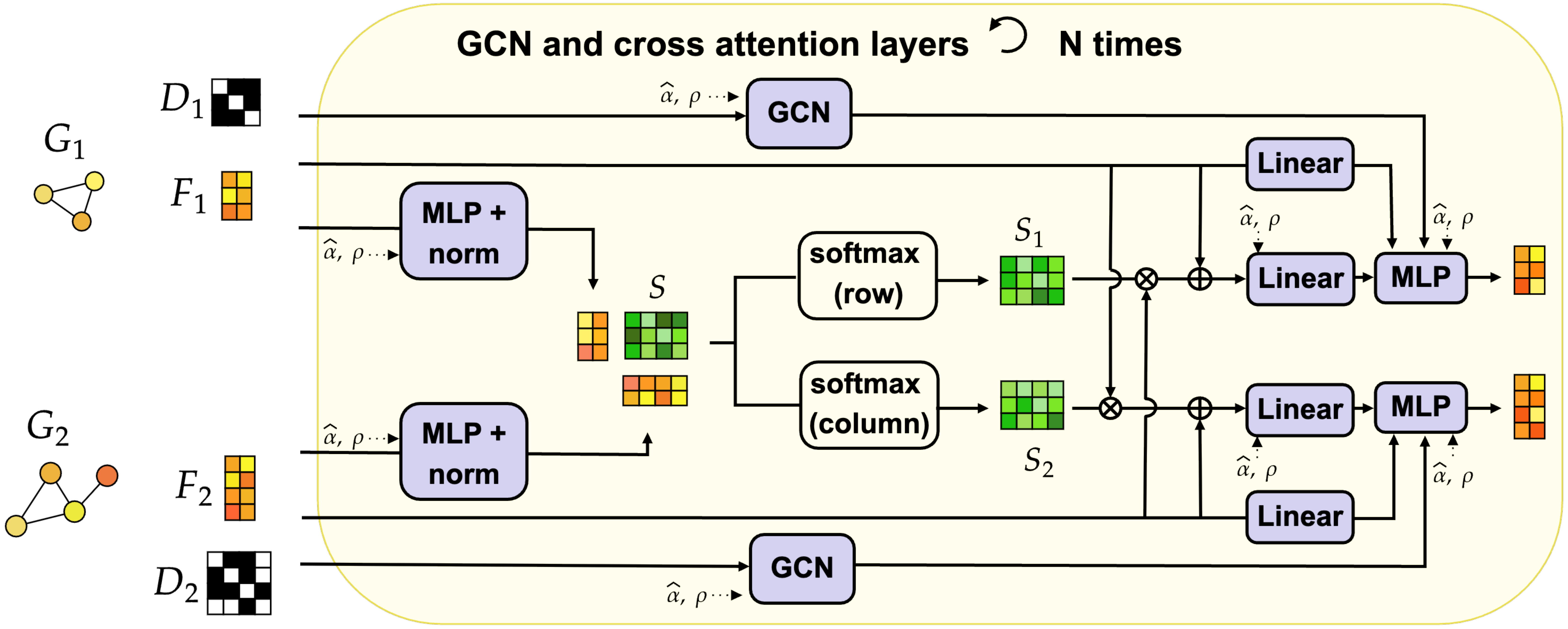


Amortized optimisation: $\min_\theta \mathbb{E}_{G_1, G_2 \sim \mathscr{D}, \alpha, \rho \sim \mathscr{P}} \left[\mathsf{L}^{\alpha,\rho}(G_1, G_2, \mathbf{P}_\theta^{\alpha,\rho}(G_1, G_2)) \right]$ $\longrightarrow$ unsupervised

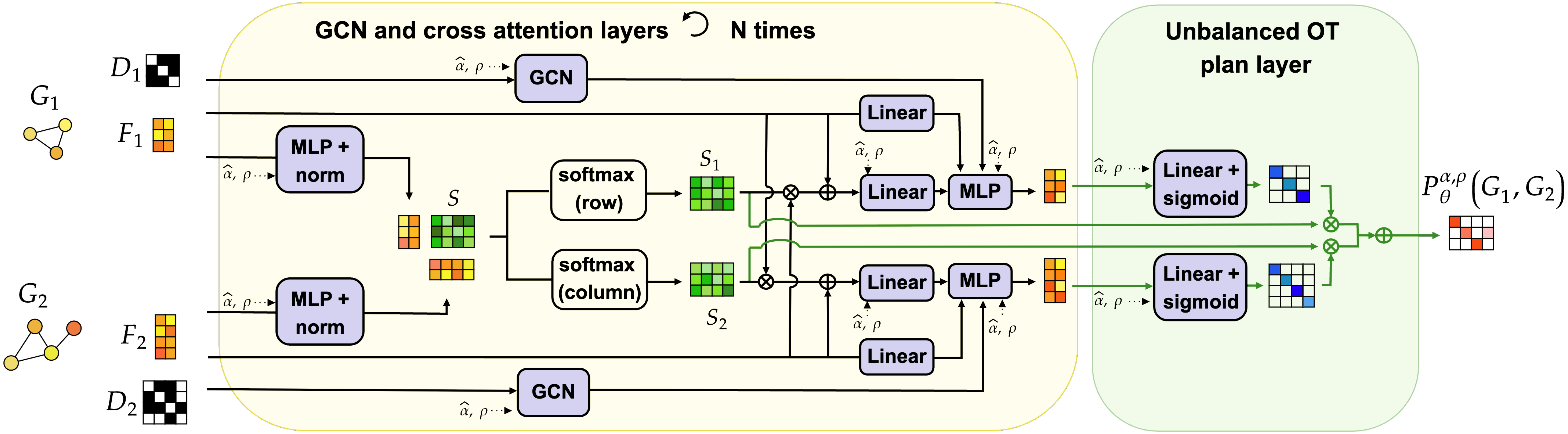
Unbalanced learning of Optimal Transport plans (ULOT)



Unbalanced learning of Optimal Transport plans (ULOT)



Unbalanced learning of Optimal Transport plans (ULOT)



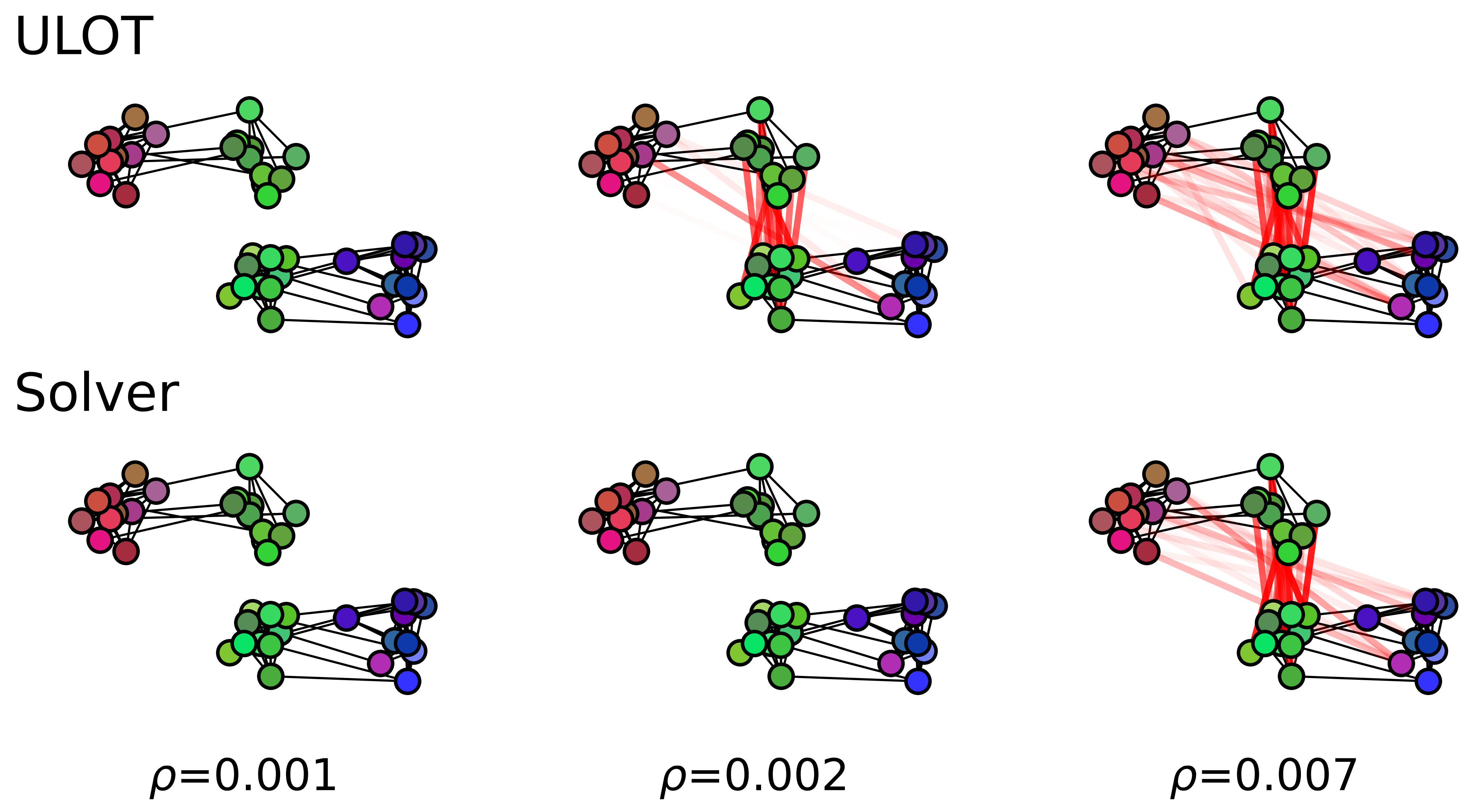
Complexity: $O(n^2)$ for n the number of graph nodes

Results on simulated graphs: effect of the ρ parameter

- ULOT trained on a dataset of Stochastic Block Model (SBM) graphs.
- Test on a pair of SBMs with one shared cluster
- ULOT plans similar to solver plans, sometimes better

$$\text{FUGW} = (1-\alpha)\underbrace{W}_{\text{node features}} + \alpha \underbrace{GW}_{\text{structure}} + \rho \underbrace{M}_{\text{marginals regularization}}$$

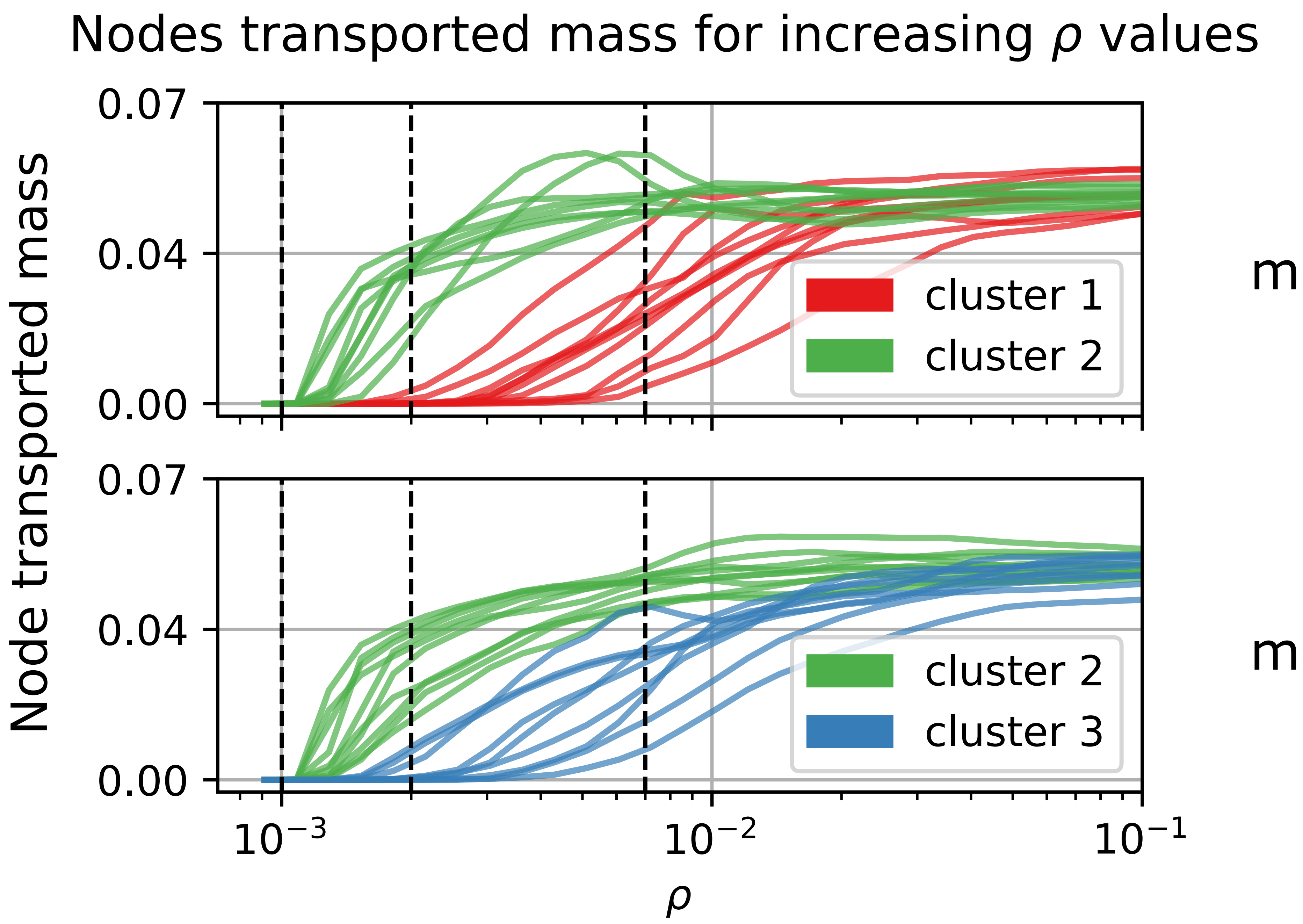
OT plan with respect to ρ for $(1,2) \rightarrow (2,3)$



Results on simulated graphs: effect of the ρ parameter

- ULOT trained on a dataset of Stochastic Block Model (SBM) graphs.
- Test on a pair of SBMs with one shared cluster
- ULOT plans similar to solver plans, sometimes better

$$\text{FUGW} = (1-\alpha)\underbrace{W}_{\text{node features}} + \alpha\underbrace{GW}_{\text{structure}} + \underbrace{\rho M}_{\text{marginals regularization}}$$



$$\text{mass}_i = \sum_j P_{i,j}$$

$$\text{mass}_j = \sum_i P_{i,j}$$

Results on simulated graphs: effect of the α parameter

- Test on a pair of SBMs with three shared clusters.
- ULOT plans similar to solver plans, sometimes better.

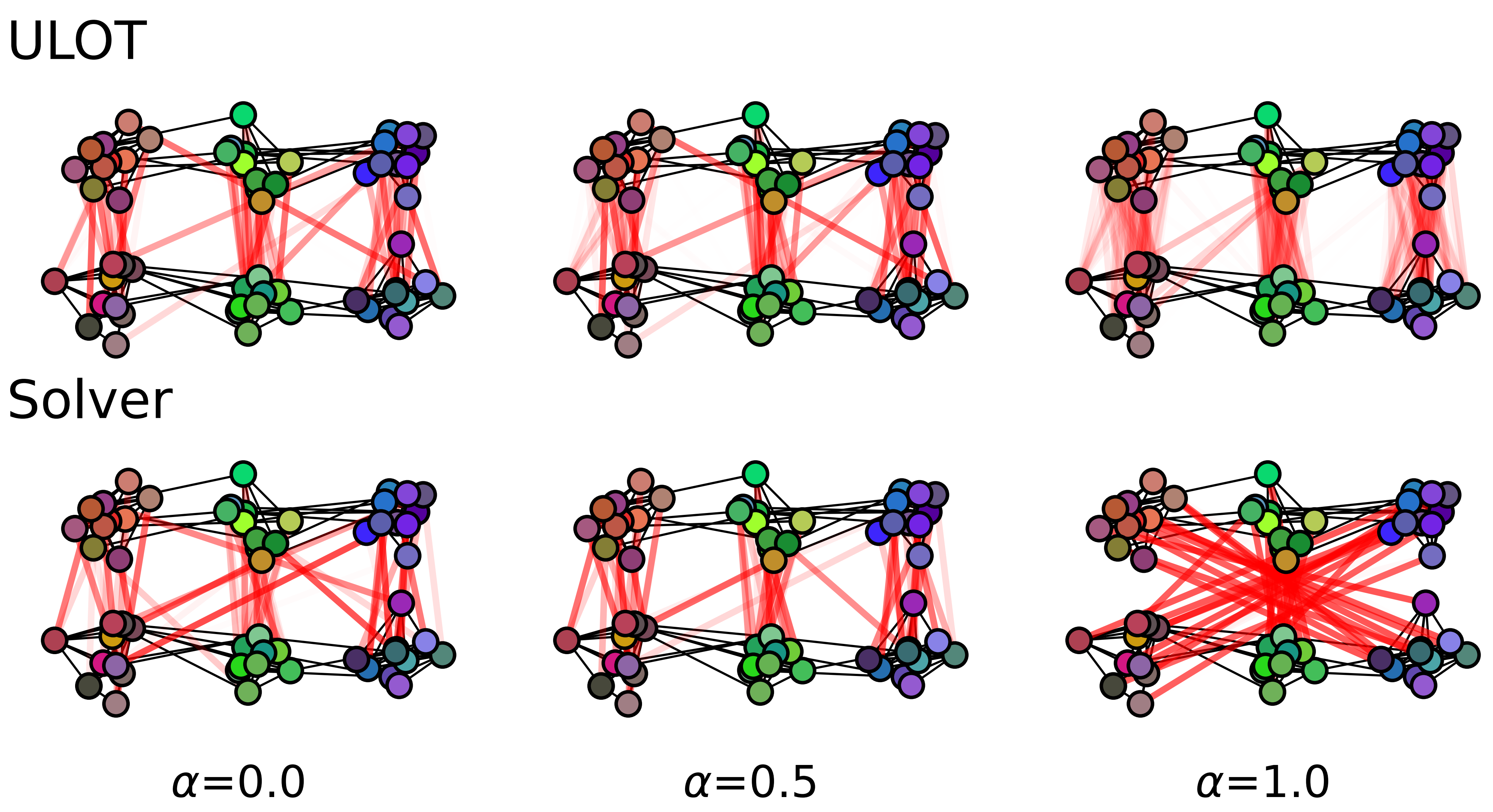
$$\text{FUGW} = (1-\alpha)W + \alpha GW + \rho M$$

node features

structure

marginals regularization

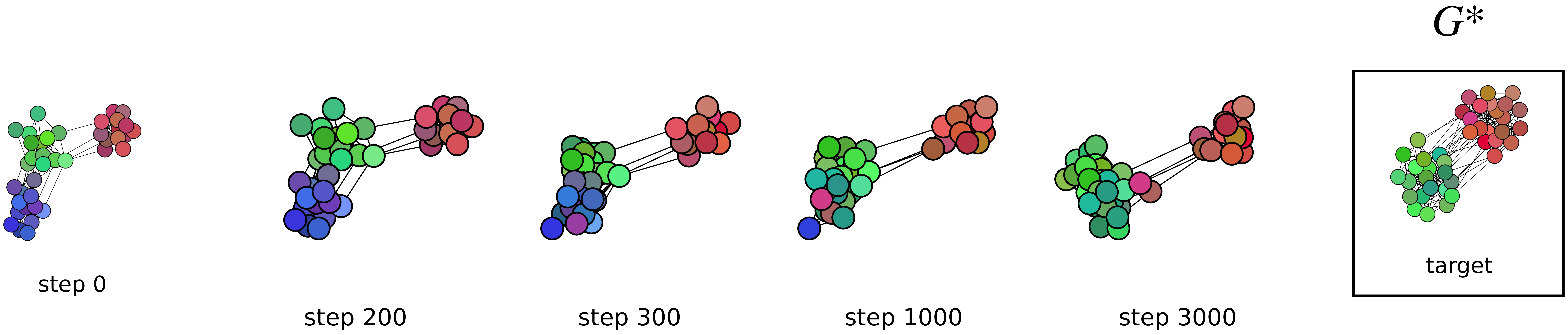
OT plan with respect to α for $(1, 2, 3) \rightarrow (1, 2, 3)$



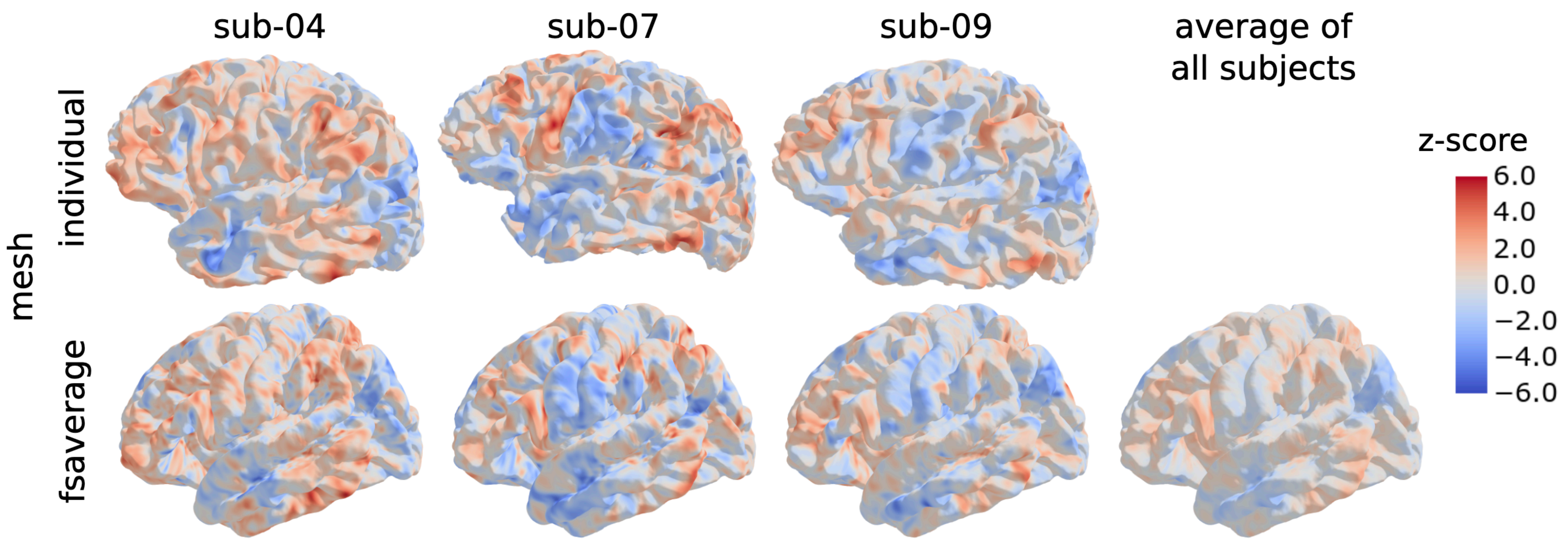
Application: minimizing functionals of the ULOT plan

ULOT transport plan is fully differentiable so we can minimize functionals of the plans and visualize the gradient descent steps

$$\min_{G} L^{\alpha,\rho}(G, G^*, P_{\theta}^{\alpha,\rho}(G, G^*))$$



Application to brain alignment



[Thual et al., 2022]

High inter subject variability (in terms of brain geometry or functional signatures) prevents generalization of observations made on a group of subjects.

Current methods: map the data to a common template, resulting in loss of detail.

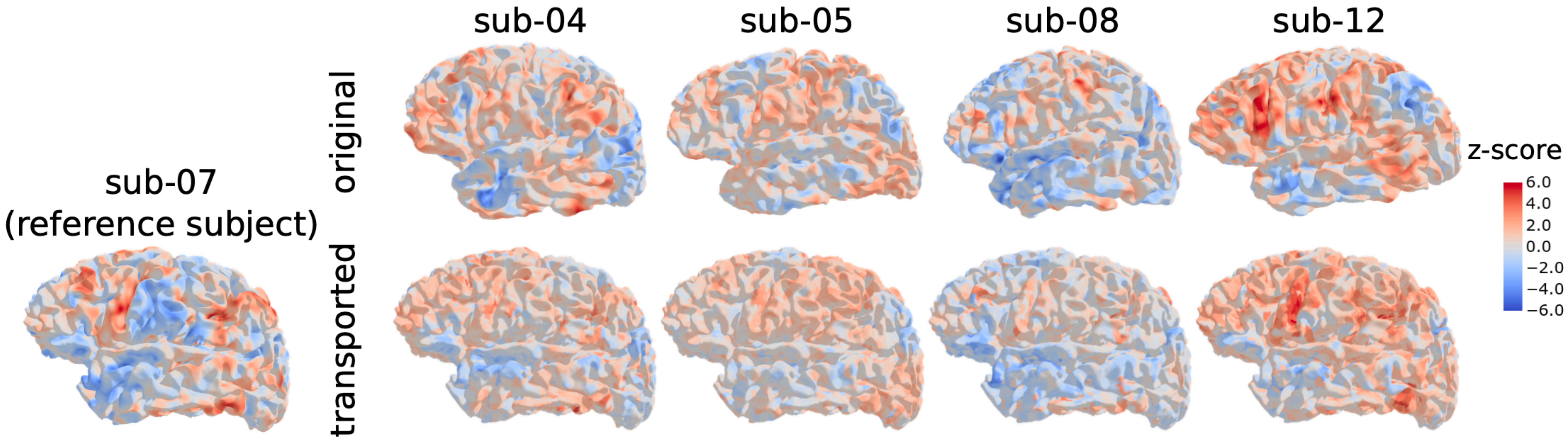
FUGW for brain alignment

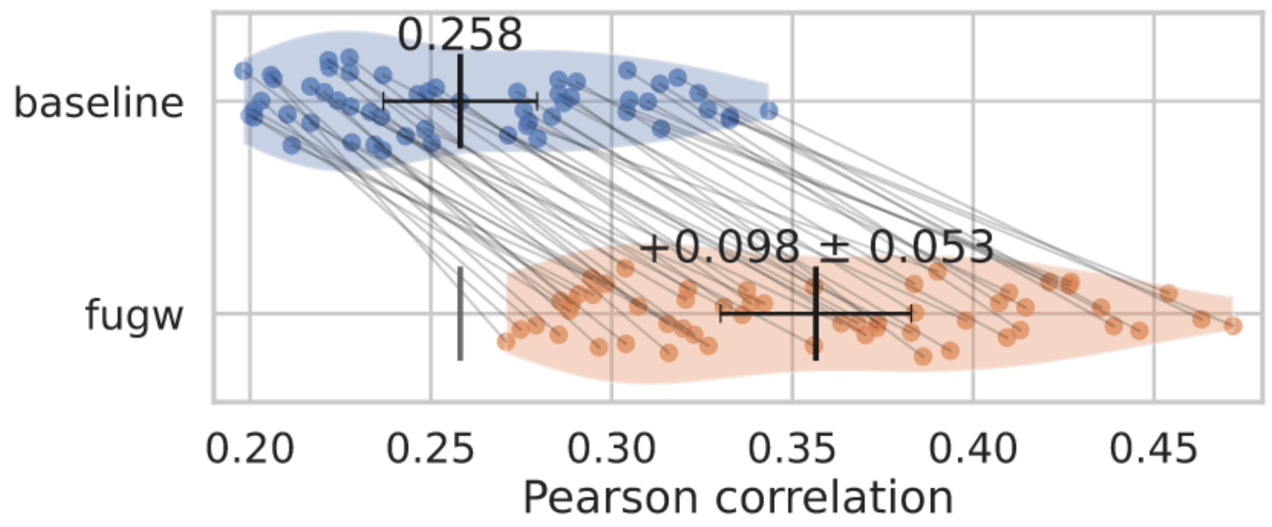
[Thual et al., 2022]: Brain alignment with FUGW transport plan.

Graphs constructed from the brain surface geometries and functional MRI activations for different tasks from the Individual Brain Charting dataset, 1000 nodes.

Results

- Transporting individual maps onto a reference subject.
- High correlation gains between the source and target contrasts after FUGW alignement.





Limitations

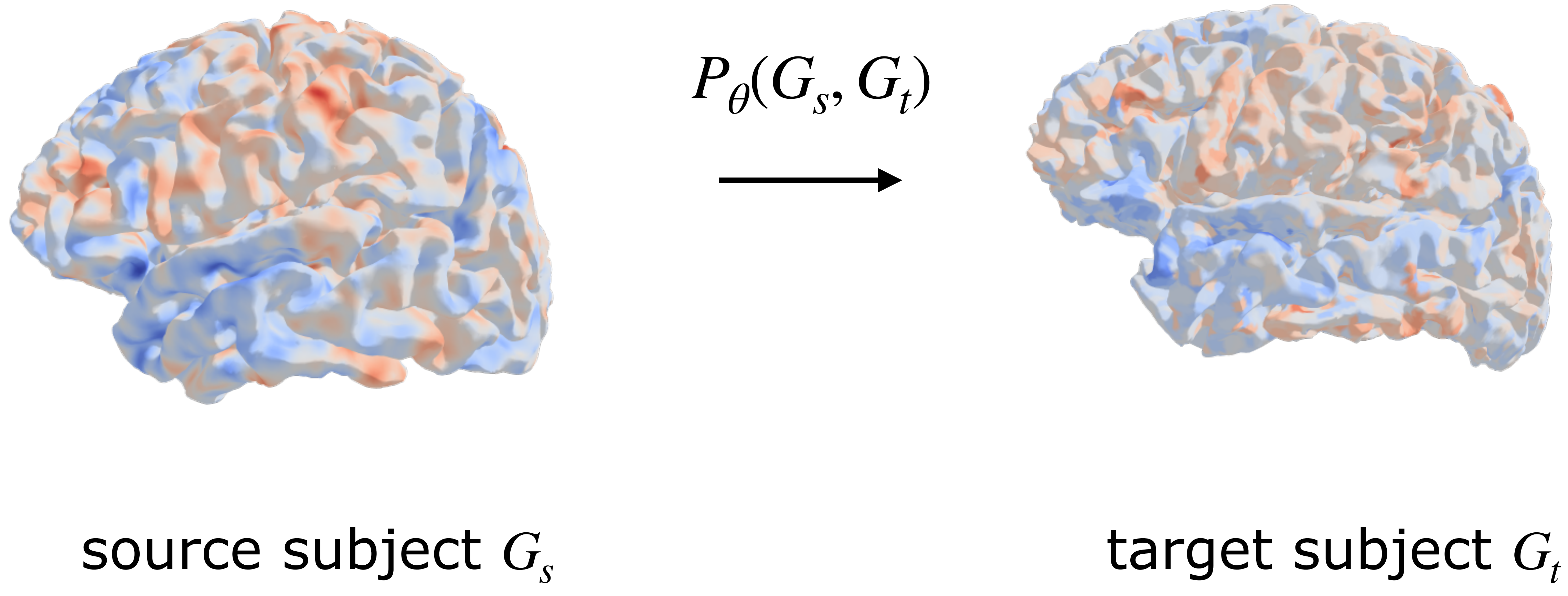
Computational time:

- 4 minutes for aligning one pair with 10k vertices on a single GPU.
- Limits applications for computing alignments on large populations

Choice of FUGW hyper parameters:

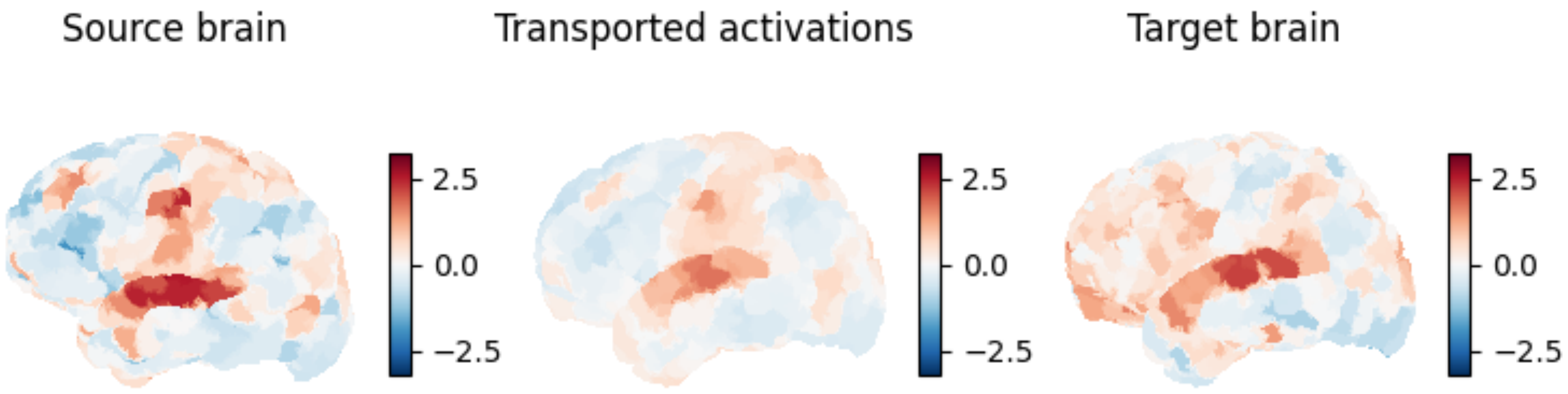
- Hyper parameters highly influence the transport plan
- Hyper parameters cannot be finely tuned because of high computational time.

Proposed method: use ULOT to align brains.

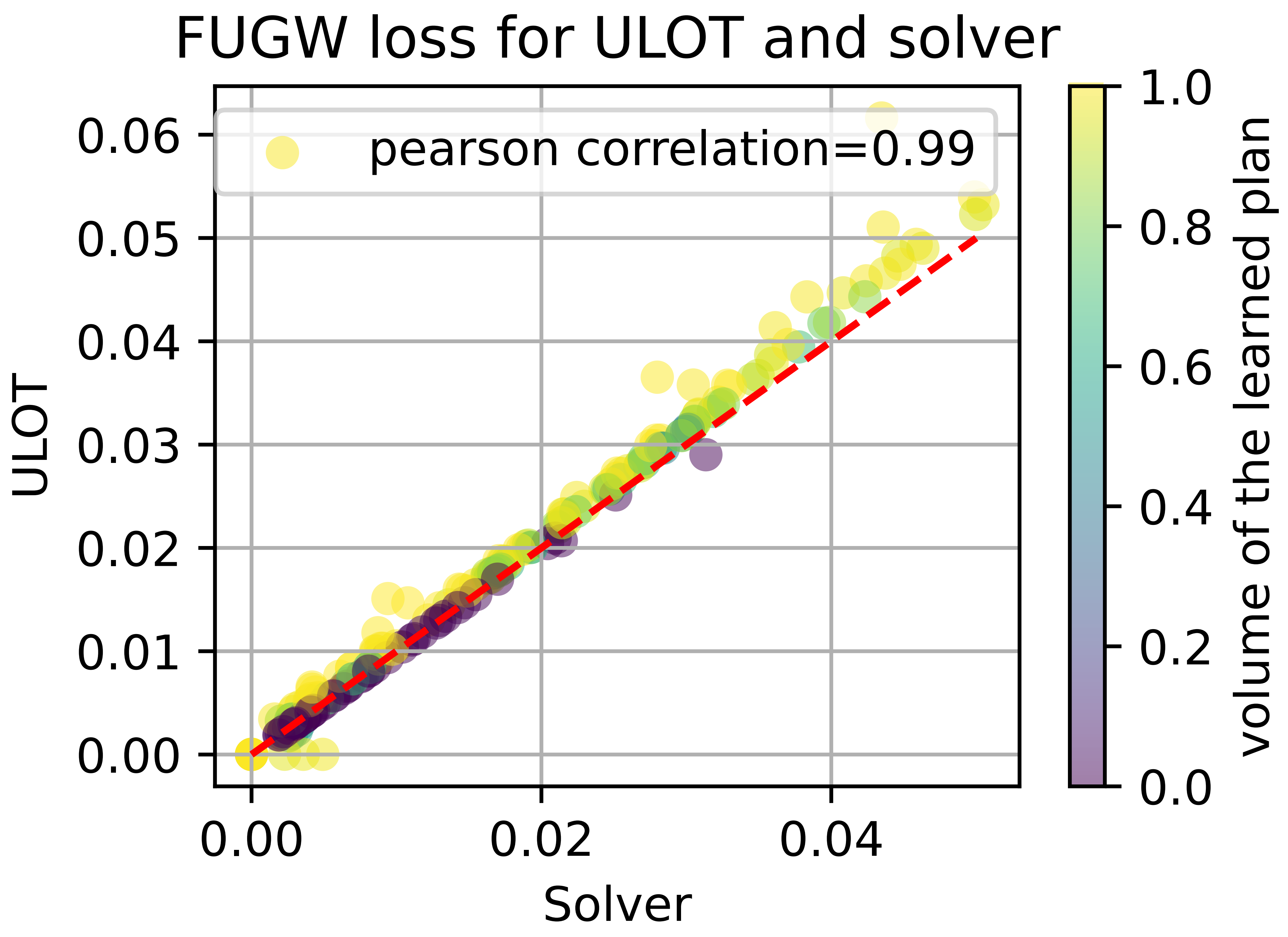


Visualisation of ULOT brain alignment

- Using the ULOT transport plan to transport functional MRI activations from a source subject to a target subject

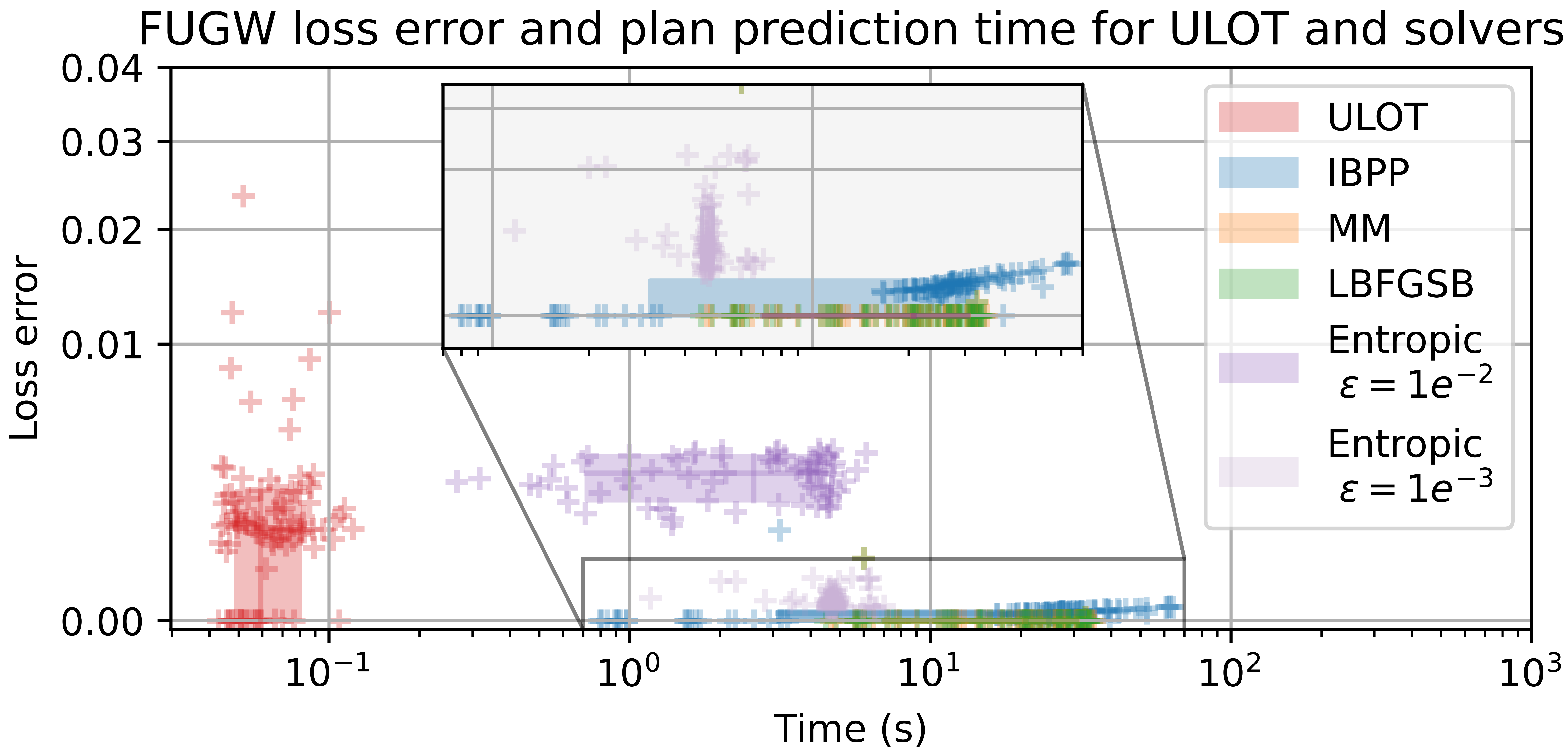


Applications to fMRI data



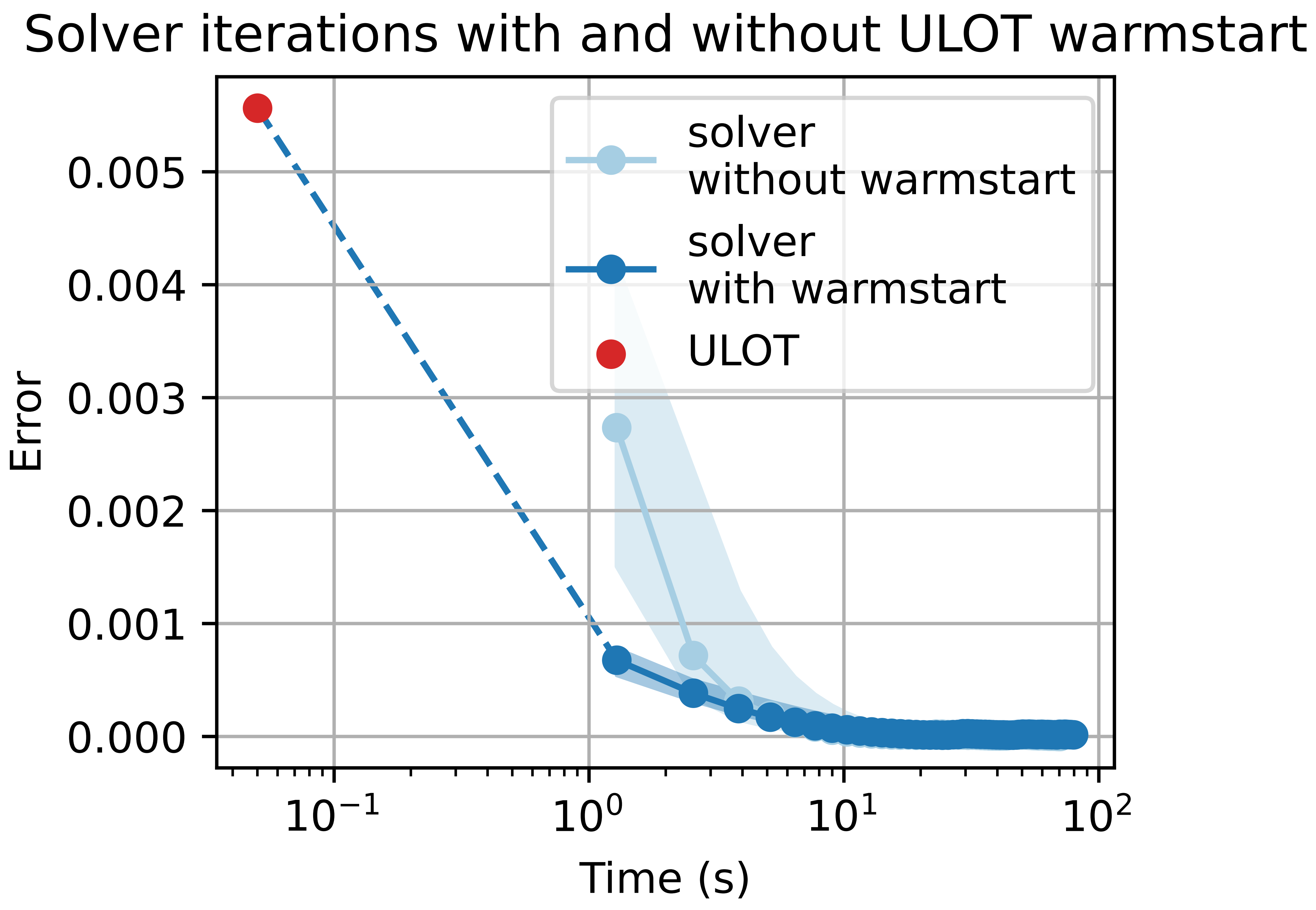
ULOT predicted plans with low error compared to solvers, and up to 100 times faster: allows extensive parameter selection and scalability to large graphs.

Applications to fMRI data



ULOT predicted plans with low error compared to solvers, and up to 100 times faster: allows extensive parameter selection and scalability to large graphs.

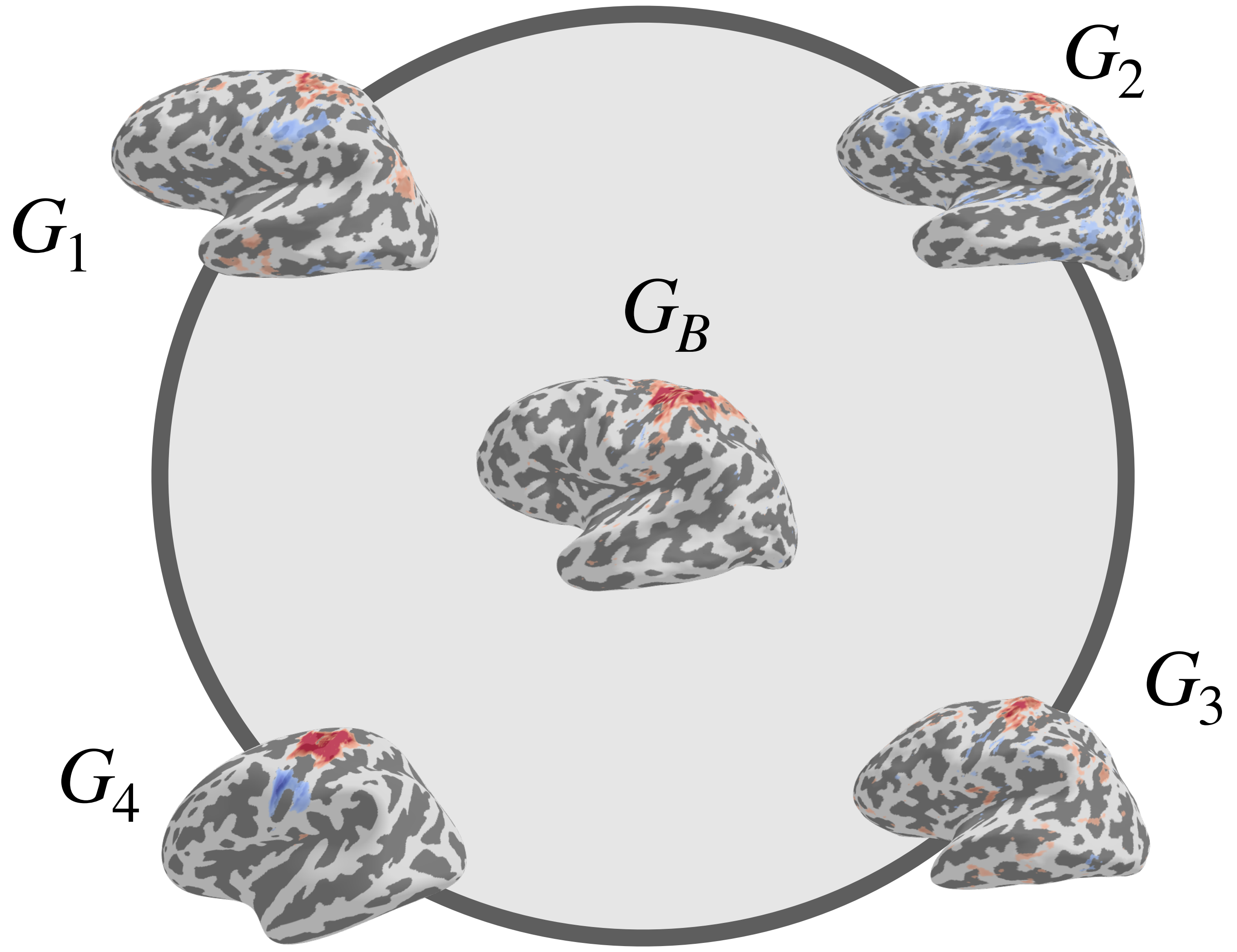
ULOT transport plan as warm start to solvers



In cases where high precision plans are needed ULOT can be used as a warm start to solvers for faster convergence.

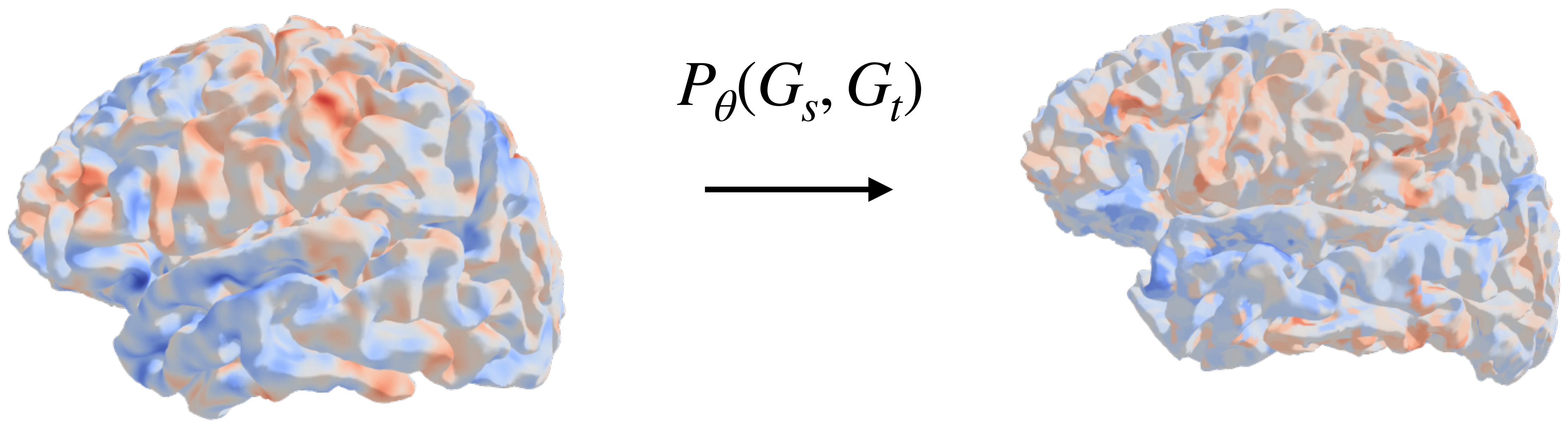
Future work on fMRI data

Fast and scalable barycenter computation: barycenter computation requires to solve many FUGW problems, which can be solved efficiently with ULOT



$$G_B = \arg\min_G \sum_{i=1}^{n} FUGW(G, G_i)$$

fMRI activation prediction: scale results from [Thual et al., 2022] to higher resolution graphs

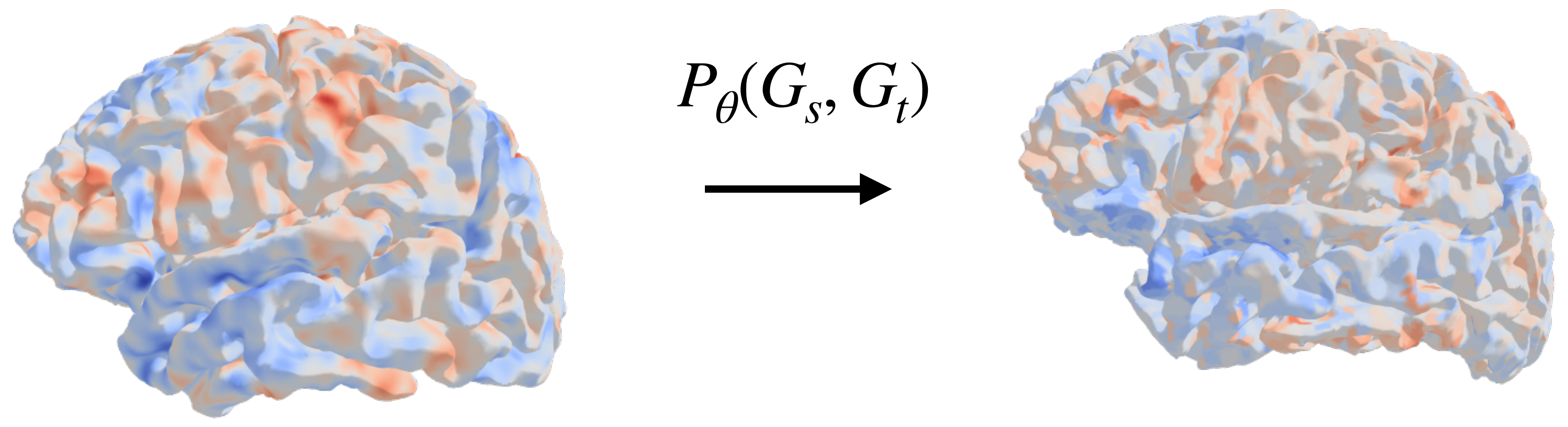


source subject G_s target subject G_t

Conclusion

- Efficient method for transport plan prediction between graphs with low error and up to 100 times faster than classical solvers.
- Enables FUGW hyper parameter selection, and applications that involve computing many transport plans (barycenters, minimization of functionals of the transport plan).
- Limitations and future work:
  - transport plan error can still be a problem in some applications where high precision is needed.
  - applications to neural dataset is limited because of the small size of datasets: need to investigate further data augmentation techniques.

paper



source subject G_s target subject G_t