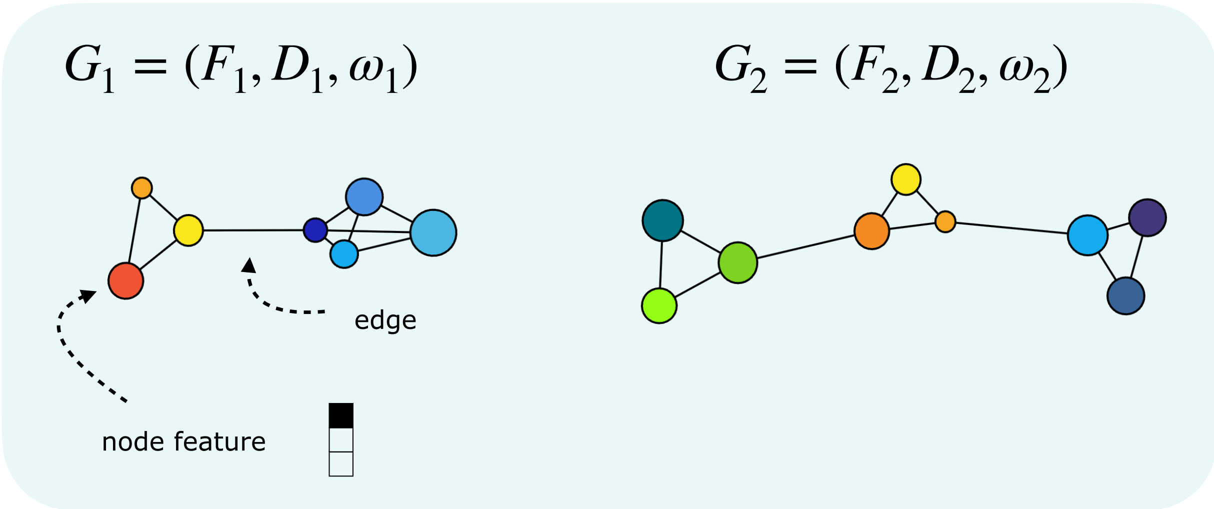
Unsupervised learning for Optimal Transport plan prediction between unbalanced graphs

Sonia Mazelet, Rémi Flamary, Bertrand Thirion

MIND meeting - 01/07/2025

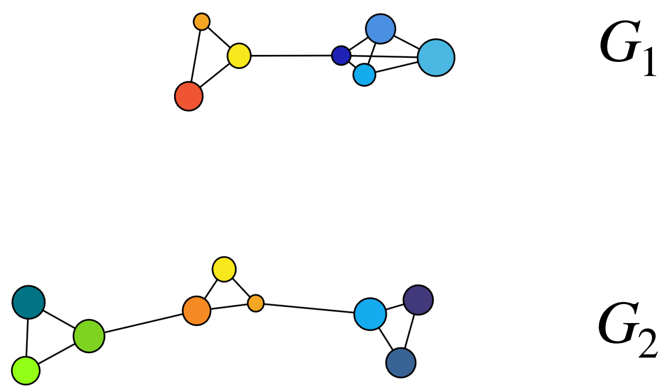
Graph matching

Graphs modeled as probability distribution, characterized by their geometry (adjacency matrix, shortest path distance matrix…), node features and node weights.



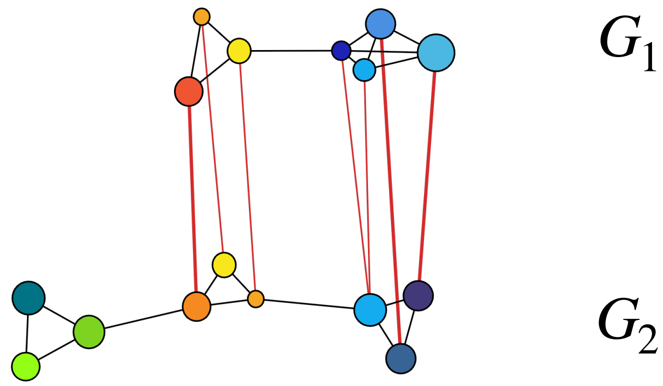
Graph matching

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.



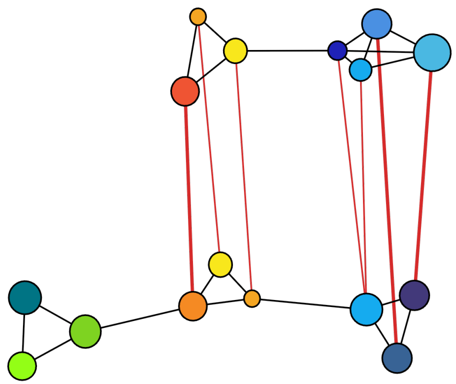
Graph matching

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.



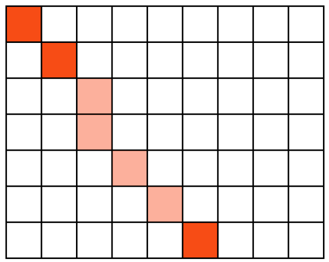
Graph matching

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

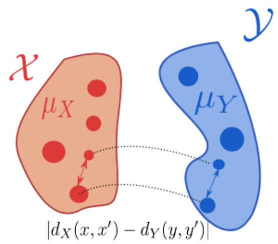
Fused Unbalanced Gromov Wasserstein (FUGW) optimal transport (OT) loss [Thual et al., 2022]

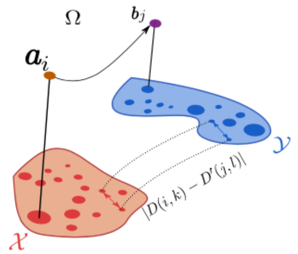
$$\mathsf{L}^{\alpha,\rho}(G_1, G_2, \mathbf{P}) = (1-\alpha) \sum_{i,j=1}^{n_1,n_2} \left\| (\mathbf{F}_1)_i - (\mathbf{F}_2)_j \right\|_2^2 \mathbf{P}_{i,j} + \alpha \sum_{i,j,k,l=1}^{n_1,n_2,n_1,n_2} \left| (\mathbf{D}_1)_{i,k} - (\mathbf{D}_2)_{j,l} \right|^2 \mathbf{P}_{i,j}\mathbf{P}_{k,l} + \rho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

match nodes with similar node features

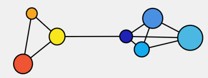
preserve local geometry

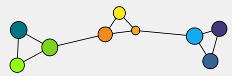
discard nodes that do not have a good match





$G_1 = (F_1, D_1, \omega_1)$ $\quad$ $G_2 = (F_2, D_2, \omega_2)$





Optimal transport distance between graphs

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

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

match nodes with similar node features

preserve local geometry

discard nodes that do not have a good match

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

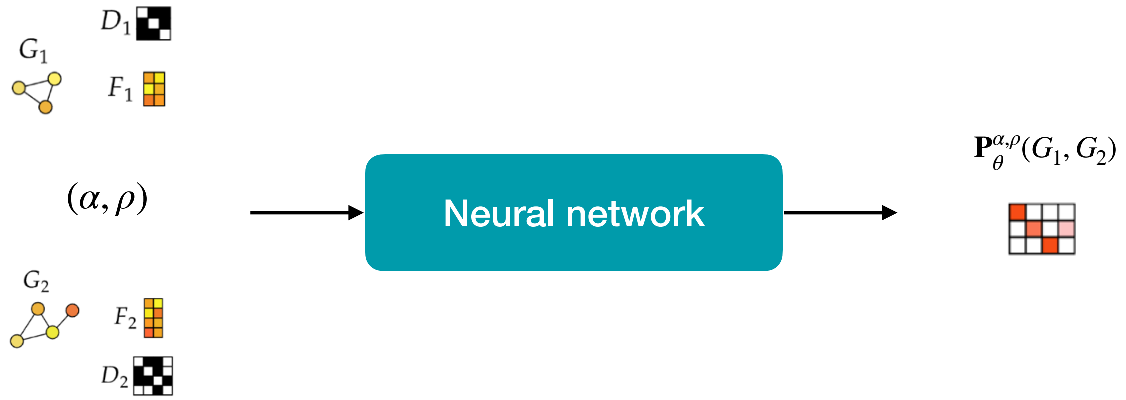
Solve the OT problem: batch coordinate descent with complexity $O(kn^3)$ for k the number of iterations and n the number of graph nodes.

→ unscalable for large graphs

Predicting FUGW plan

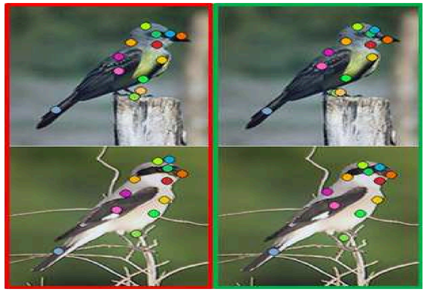
Goal: learn to predict FUGW plan $\mathbf{P}_\theta^{\alpha,\rho}(G_1, G_2)$ for all graph pairs $(G_1, G_2) \sim \mathscr{D}$ and parameters $(\alpha, \rho) \sim \mathscr{P}$.

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

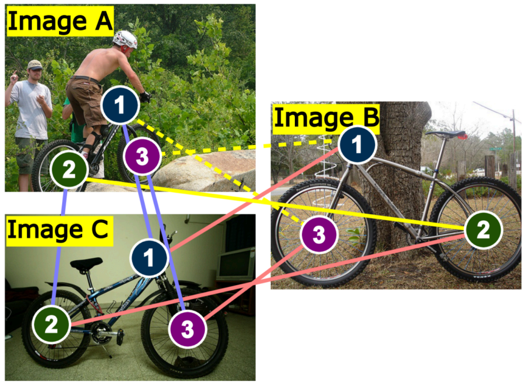


Training a graph matching neural network

Most graph matching neural network are trained in a supervised way *[Wang et al., 2019][Sarlin et al. 2020][Zanfir et al. 2020]* → ground truth correspondences are hard (if not impossible) to compute.

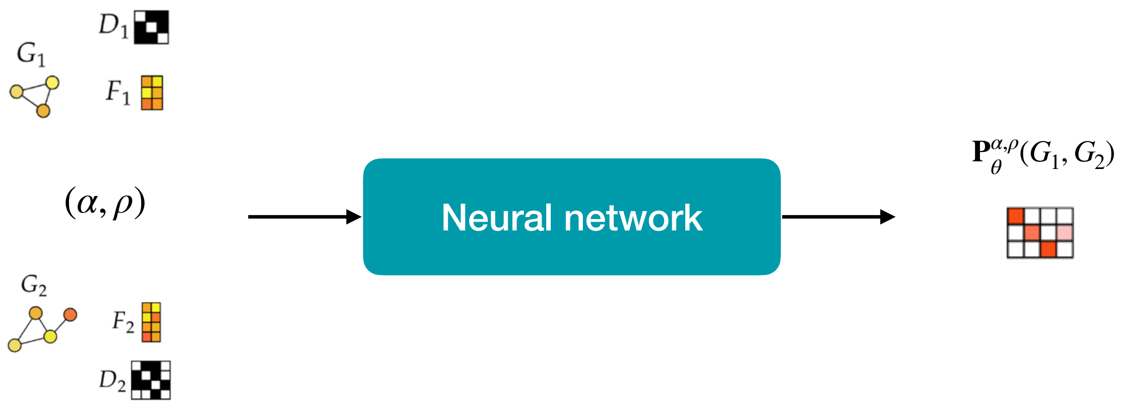


Methods trained in a unsupervised learn to match two copies of the same graph *[Liu et al. 2022],* or learn to minimize a criterion that is domain specific *[Tourani et al. 2024].*



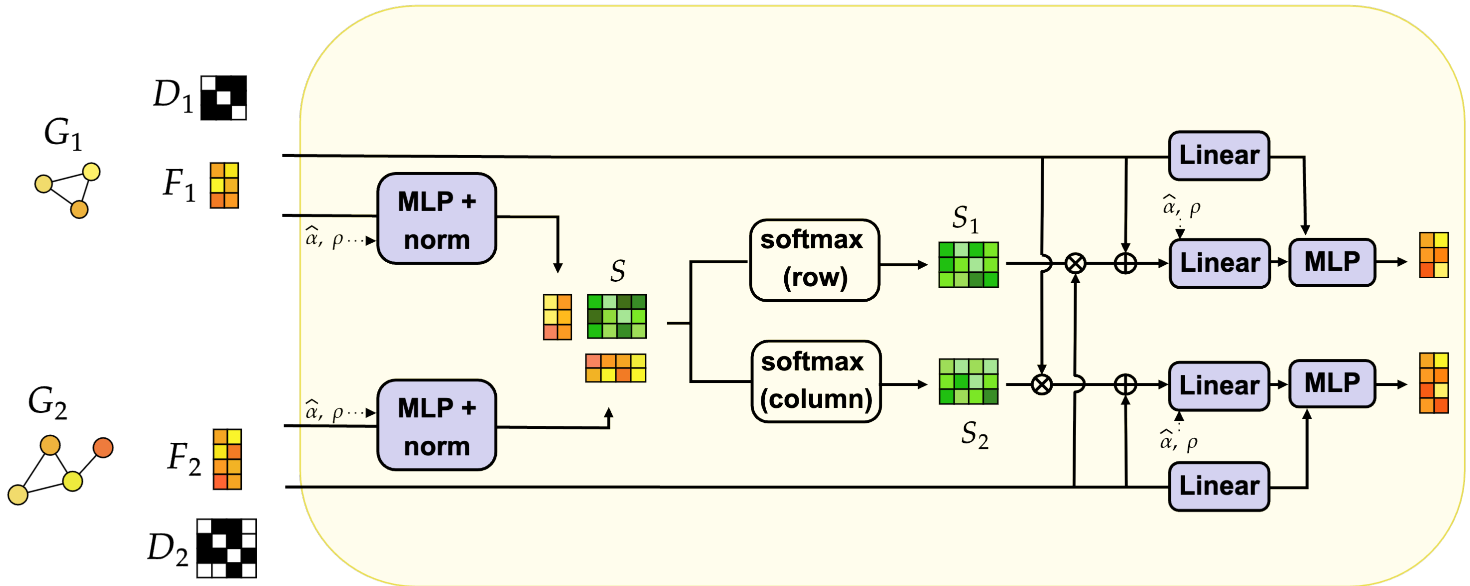
Predicting FUGW plan

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

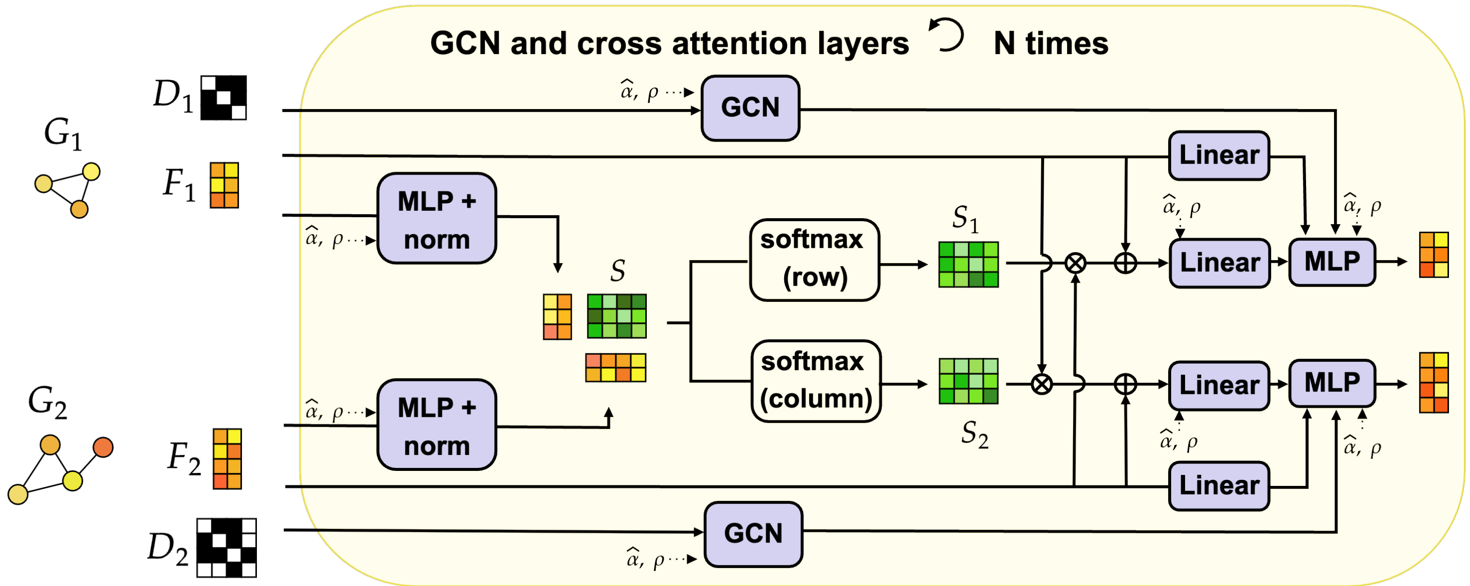


Optimisation problem: $\min_{\theta} \mathbb{E}_{G_1, G_2 \sim \mathcal{D}, \alpha, \rho \sim \mathcal{P}} \left[\mathsf{L}^{\alpha,\rho}(G_1, G_2, \mathbf{P}_\theta^{\alpha,\rho}(G_1, G_2)) \right]$ ⟶ unsupervised

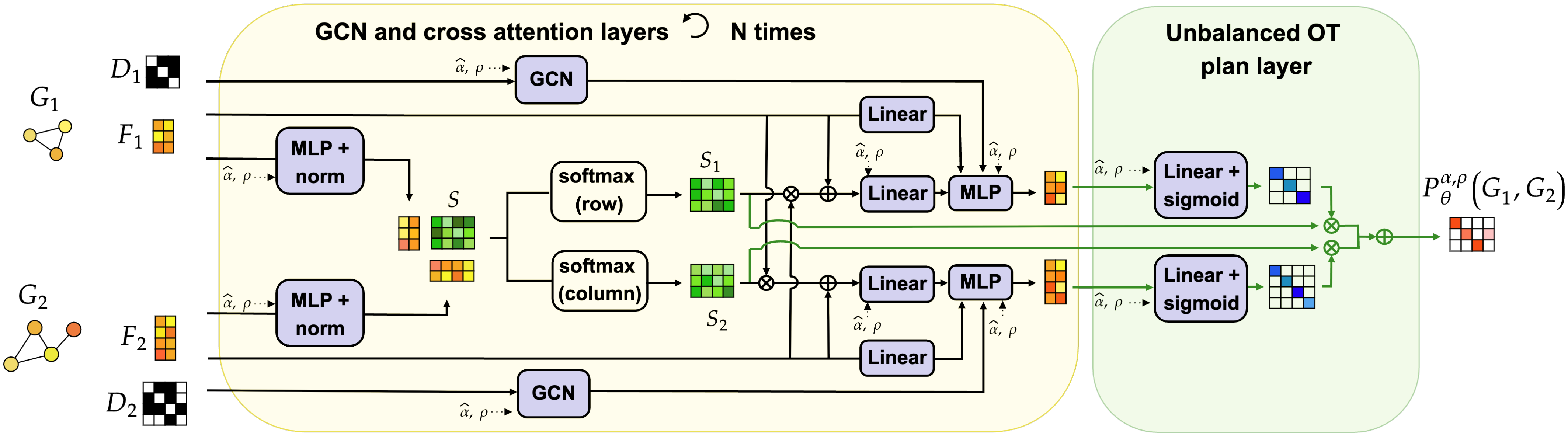
Unbalanced learning of Optimal Transport plans (ULOT)



Unbalanced learning of Optimal Transport plans (ULOT)

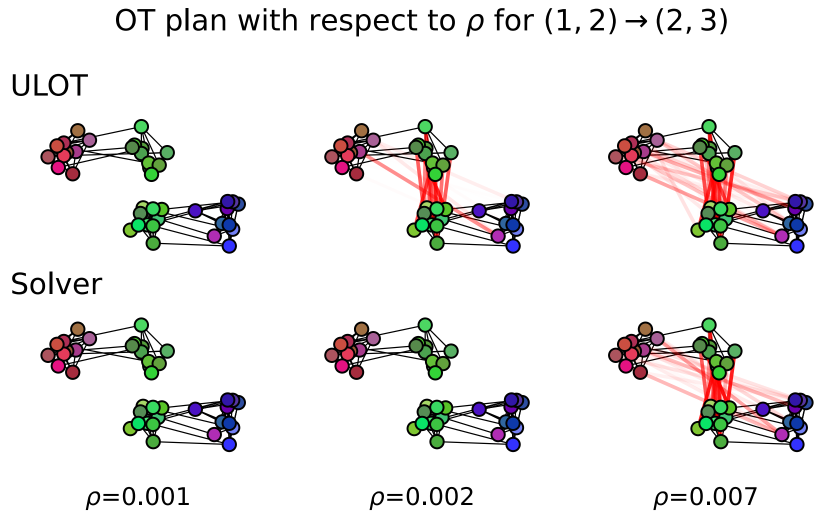


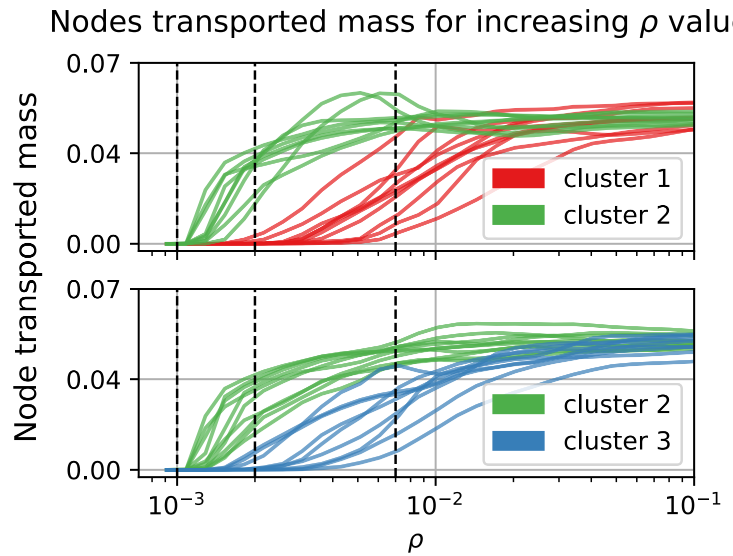
Unbalanced learning of Optimal Transport plans (ULOT)



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

Results on simulated graphs





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

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

ULOT trained on a dataset of Stochastic Block Model (SBM) graphs.

Predicted plans visualized on a pair of SBM graphs with one shared cluster: plans are similar to plans computed with classical solver and sometimes even better.

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

node features

structure

marginals regularization

Results on simulated graphs

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

node features

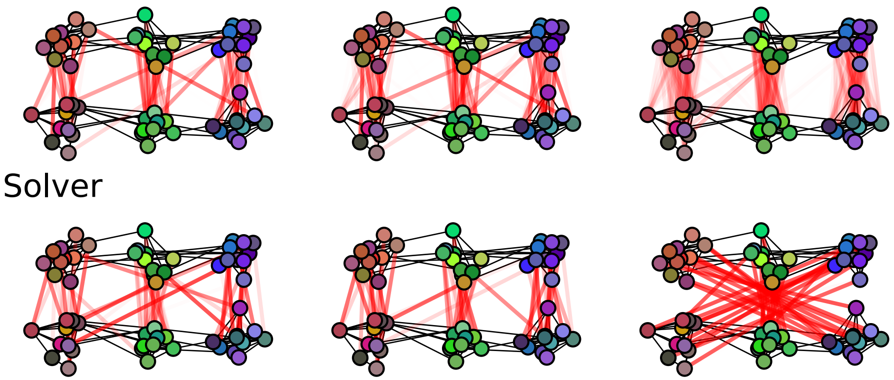
structure

marginals regularization

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

ULOT

Solver



α=0.0 α=0.5 α=1.0

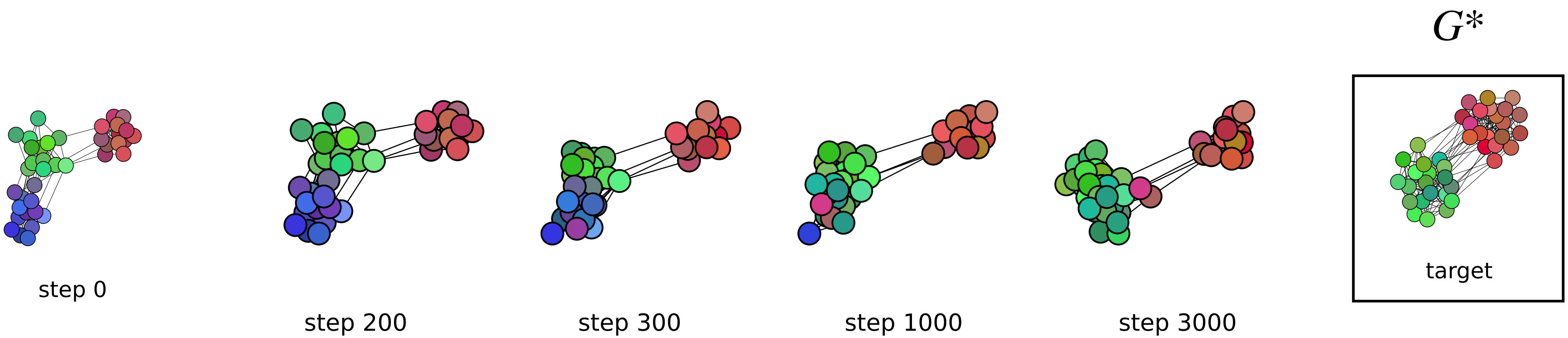
Predicted plans visualized on a pair of SBM graphs with three shared cluster: plans are similar to plans computed with classical solver.

For $\alpha = 1$ (graph structure only), ULOT pairs clusters correctly while the solver cannot differentiate the 1st and 3rd.

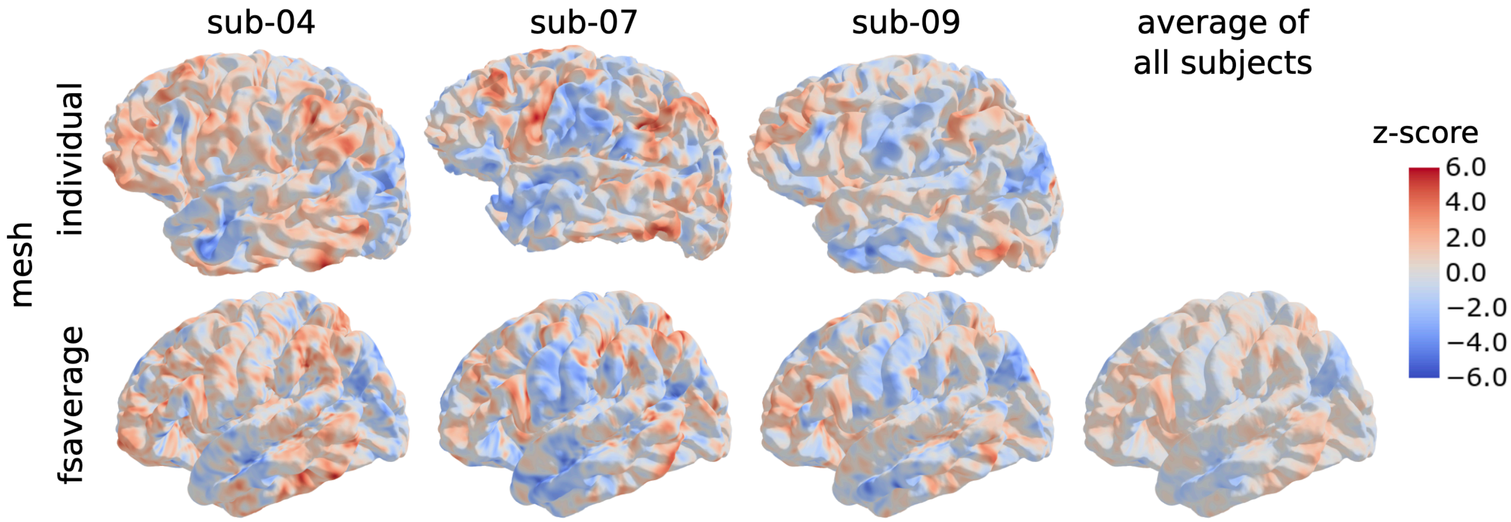
Application: minimizing functionals of the ULOT plan

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 on brain alignment



[Thual et al., 2022]

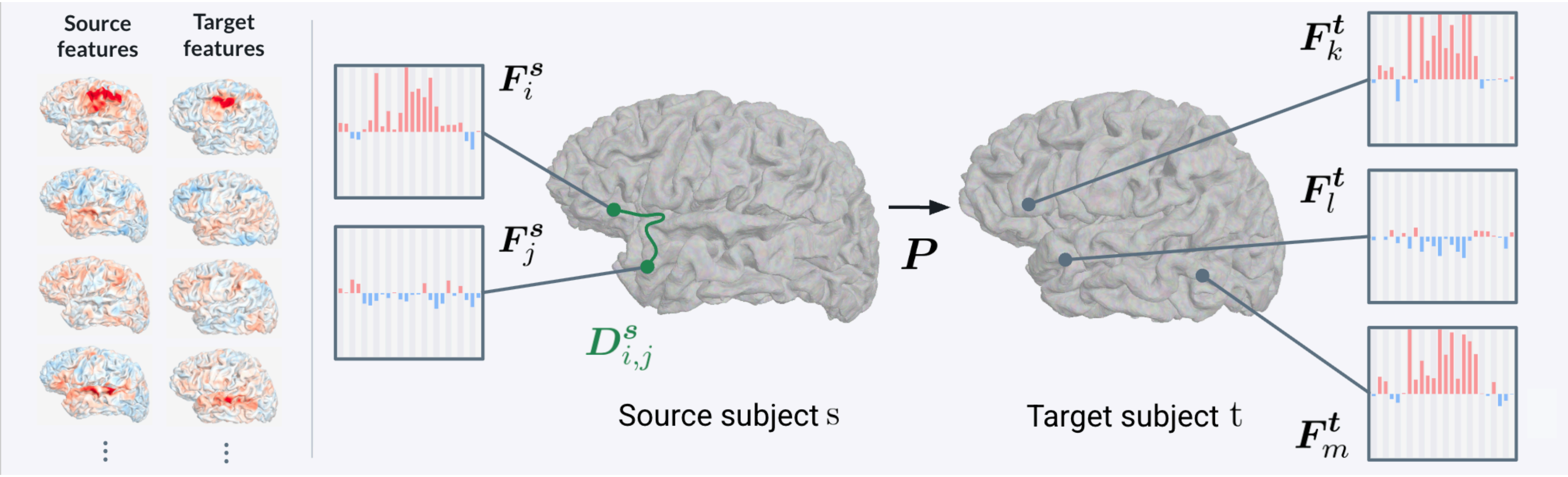
High inter subject variability (in terms of brain geometry or functional signatures) prevents generalization of observations made on a group of subjects.

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

FUGW for brain alignment

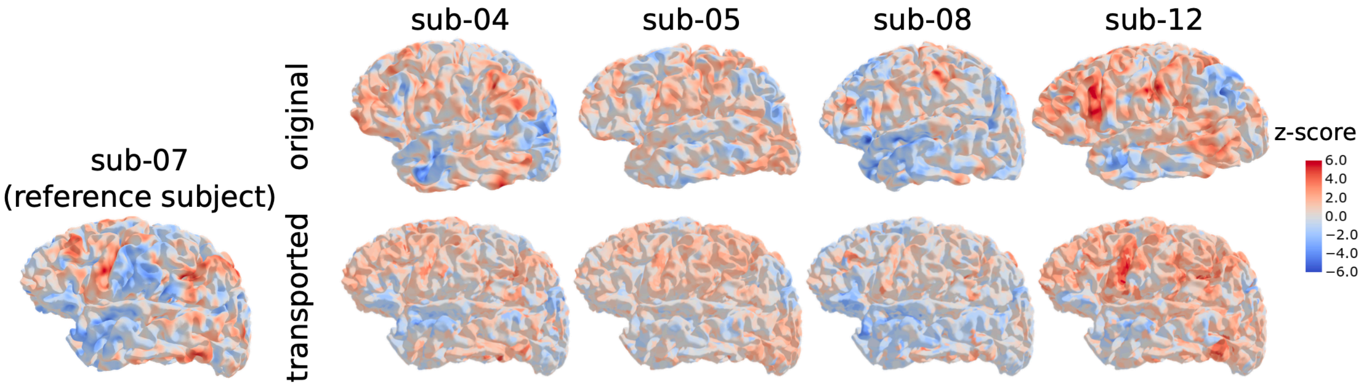
[Thual et al., 2022]: Brain alignment with FUGW transport plan.

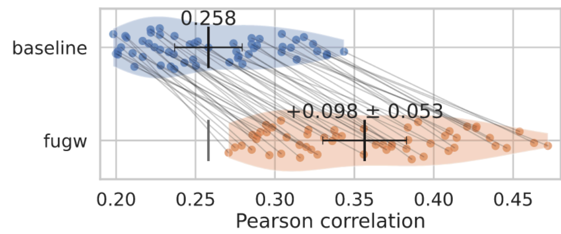
Graphs constructed from the brain surface geometries and fMRI activations for different tasks from the IBC dataset, 1000 nodes.



Results

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



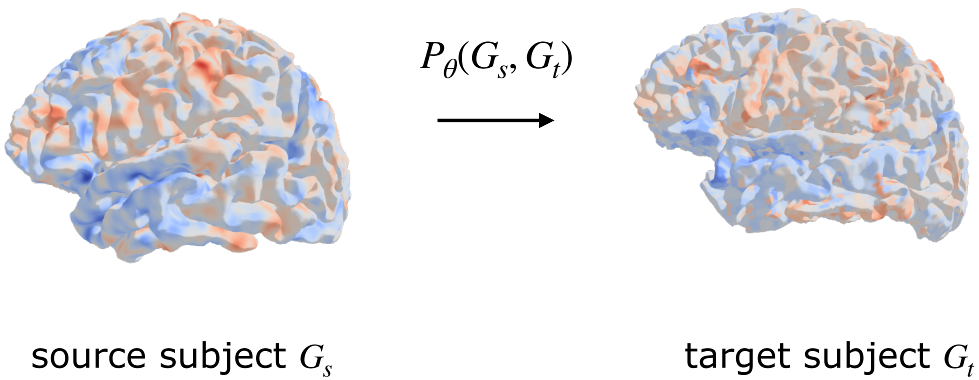


Limitations

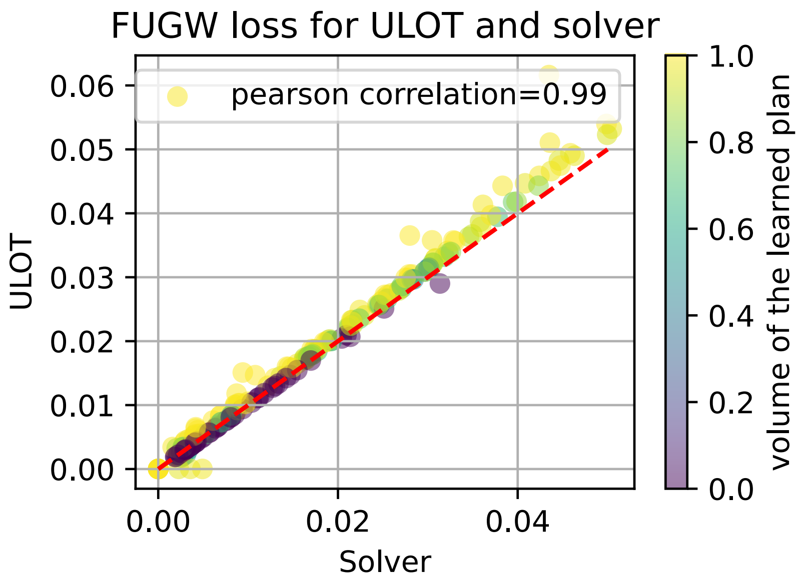
Computational time: 4 minutes for aligning one pair with 10k vertices on a single GPU. Limits applications such as computing barycenters, or computing alignments on large populations.

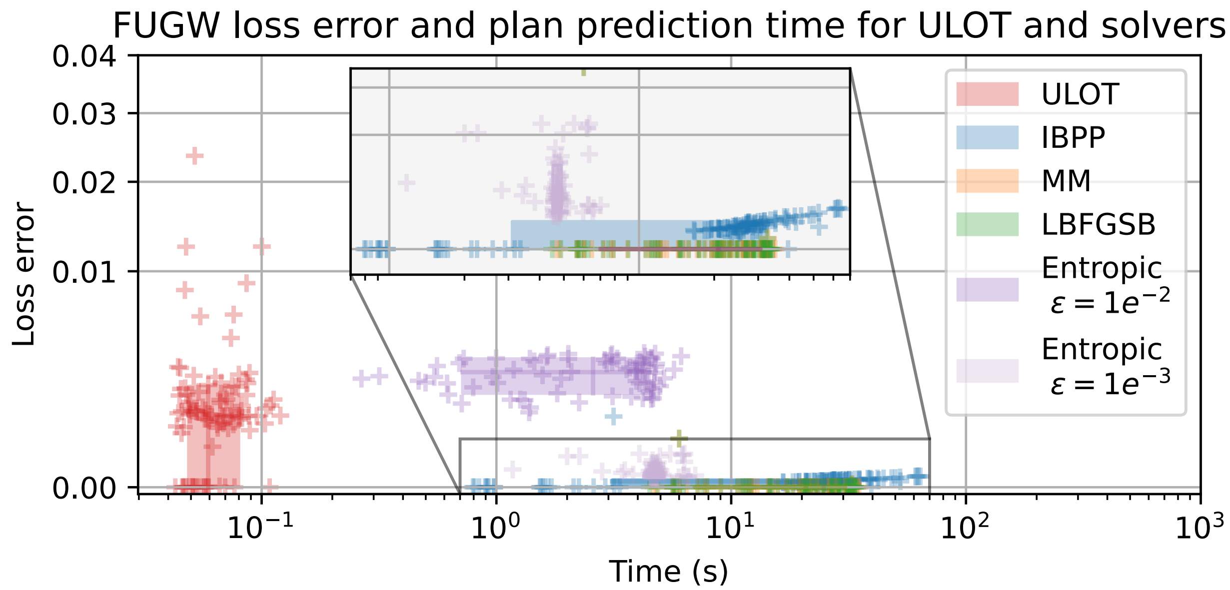
Choice of FUGW hyper parameters: hyper parameters highly influence the transport plan, and cannot be finely tuned because of high computational time.

Proposed method: used ULOT to align brains.



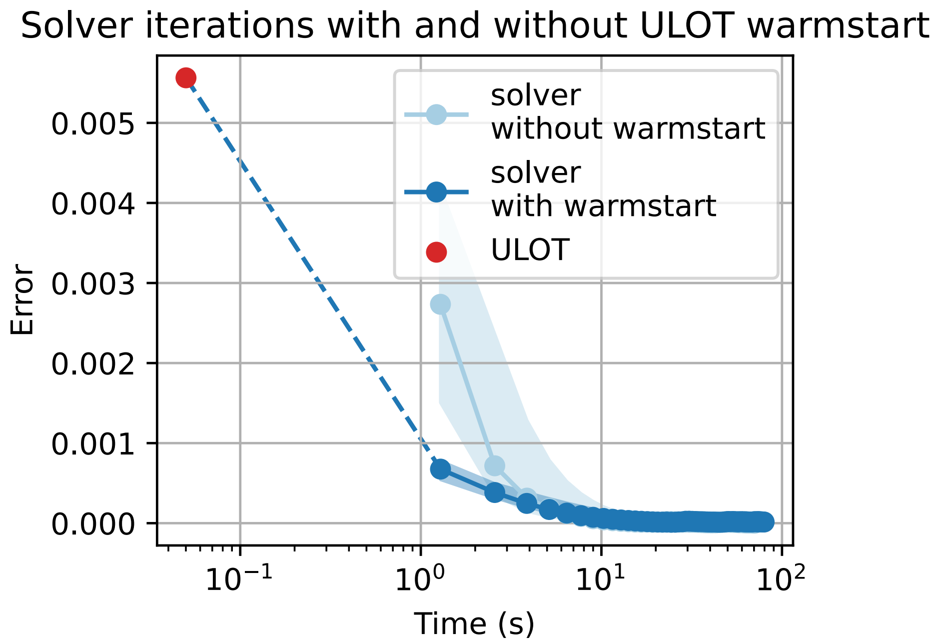
Applications on fMRI data





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

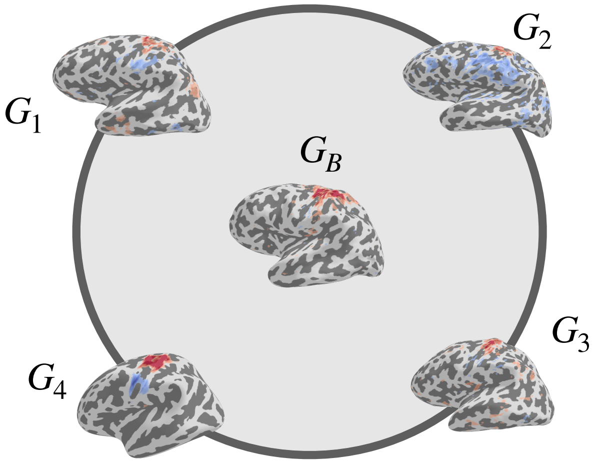
ULOT transport plan as warm start to solvers



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

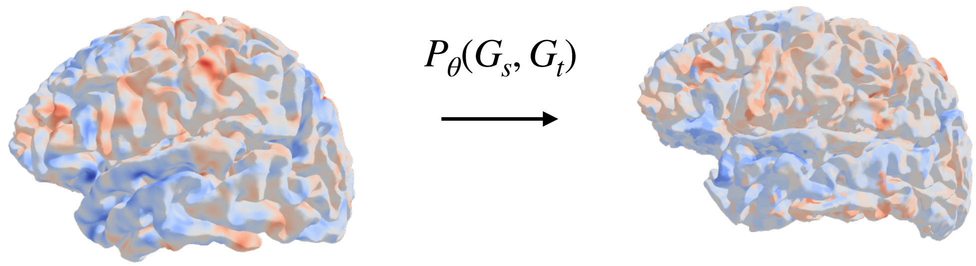
Future work on fMRI data

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: match results obtained by [Thual et al., 2022] and scale 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 plans (barycenters, minimization of functionals of the transport plan).
- Applications on fMRI dataset.
- Limitations and future work:
  - transport plan error can still be a problem in some applications where high precision is needed.
  - applications on neural dataset is limited because of the small size of datasets: need to investigate further data augmentation techniques.