# Unsupervised learning for Optimal Transport plan prediction between unbalanced graphs

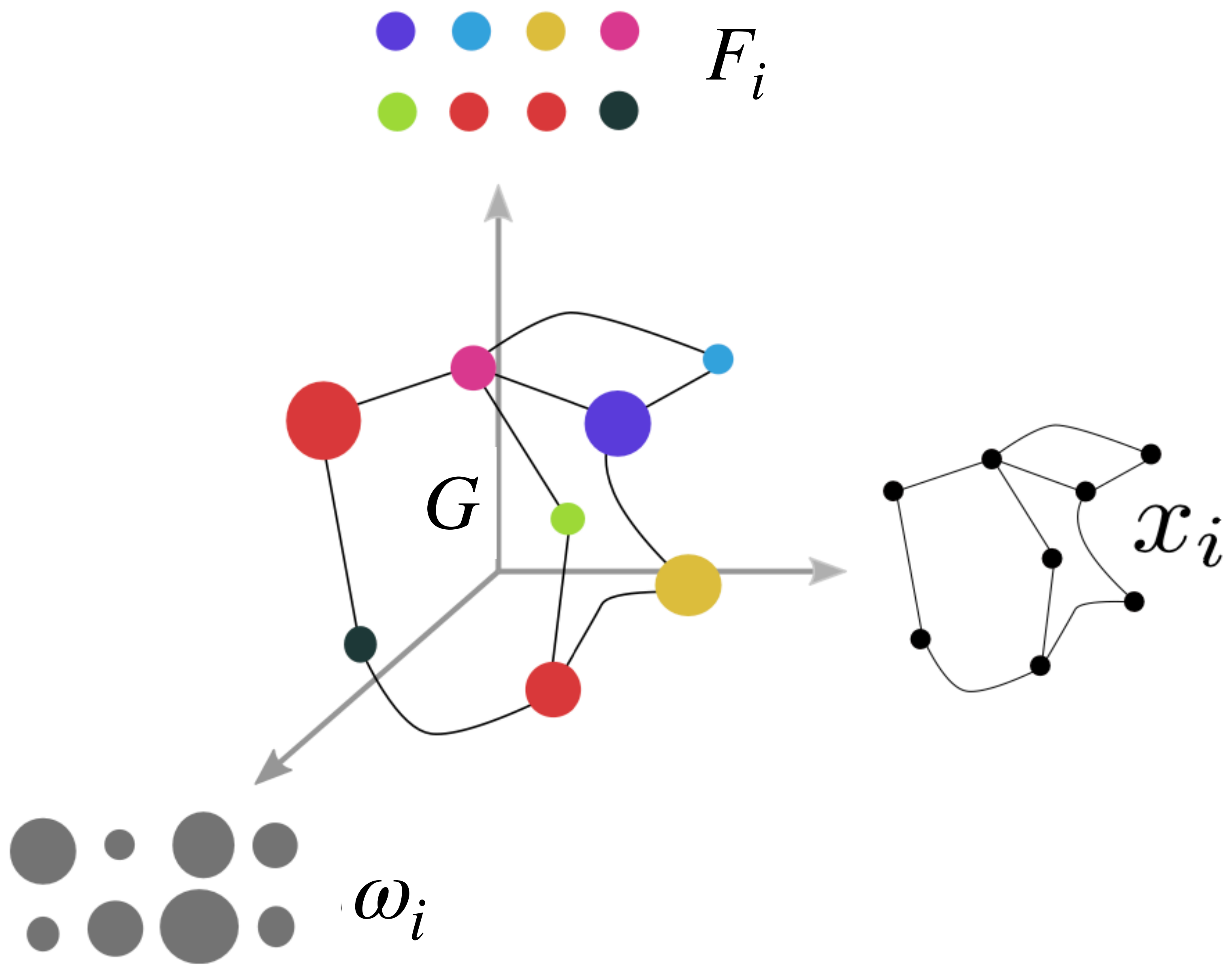
Sonia Mazelet, Rémi Flamary, Bertrand Thirion

GLOW reading group - 11 February 2026

# 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: $\omega$



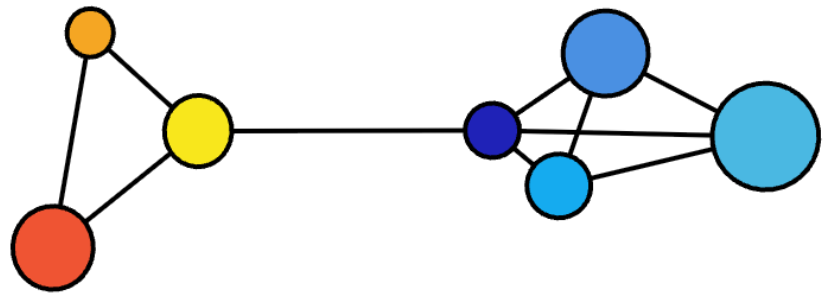

$$\mu = \sum_i \omega_i \delta_{(x_i, F_i)}$$

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

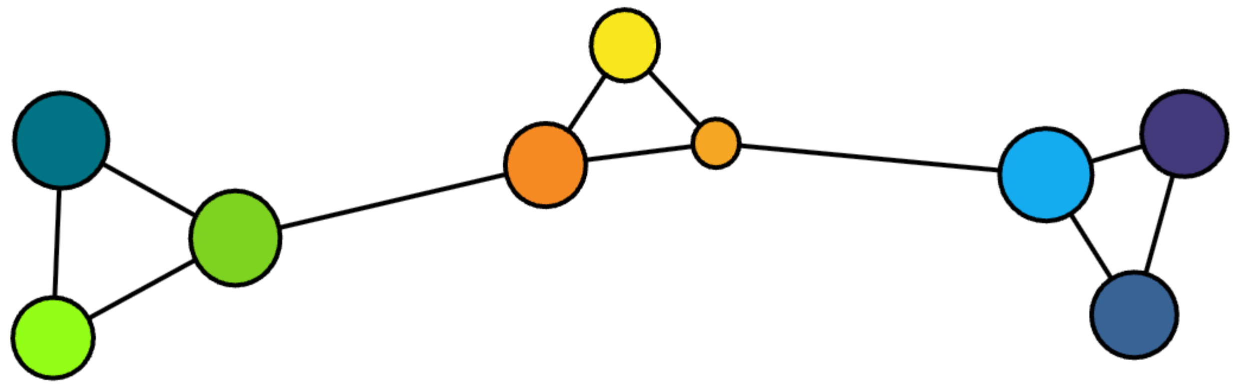
$$\mu_X = \sum_i \omega_i \delta_{x_i}$$

# 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.



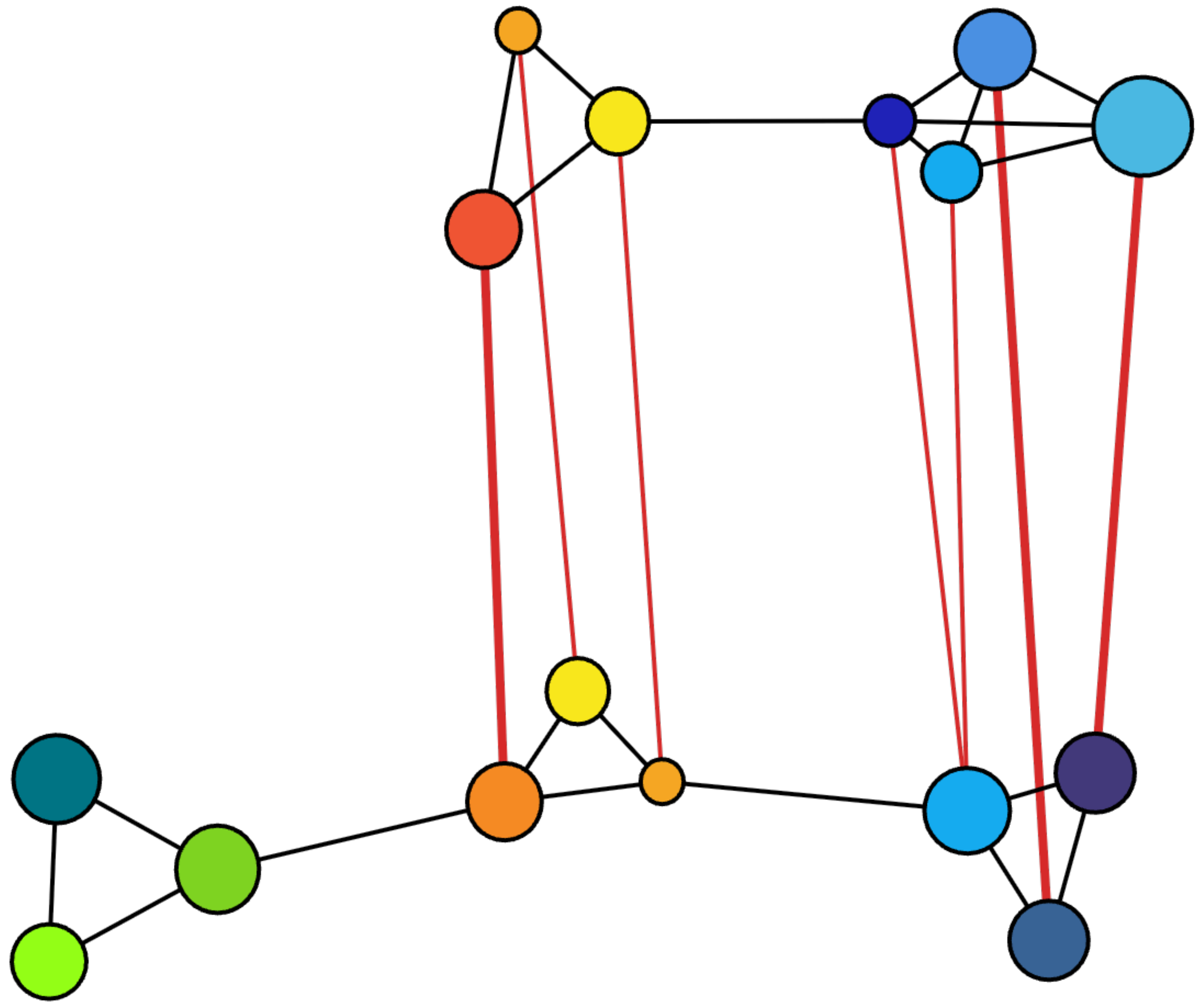
$$G_1 = (F_1, D_1, \omega_1)$$



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

# 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.

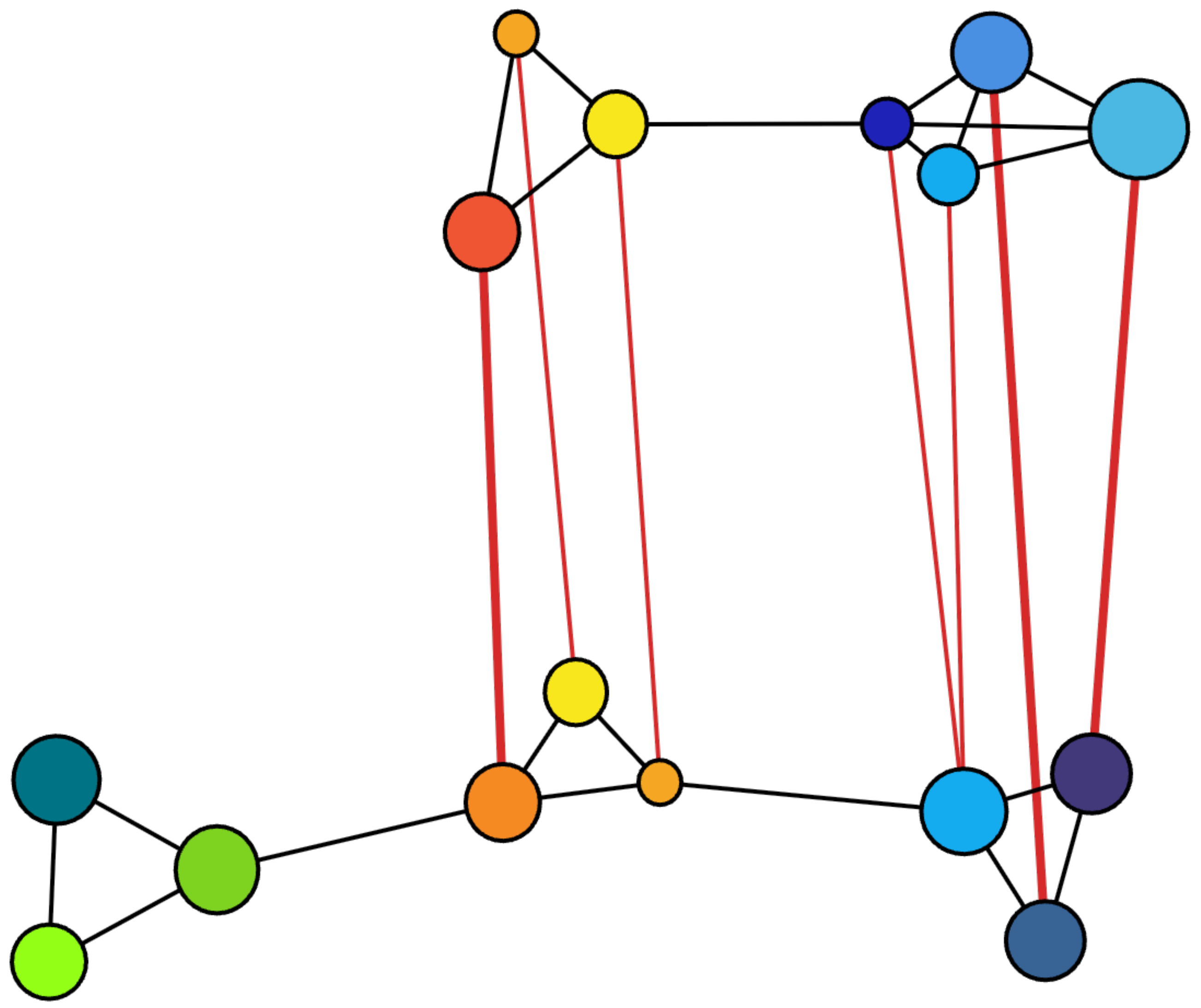


$$G_1 = (F_1, D_1, \omega_1)$$

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

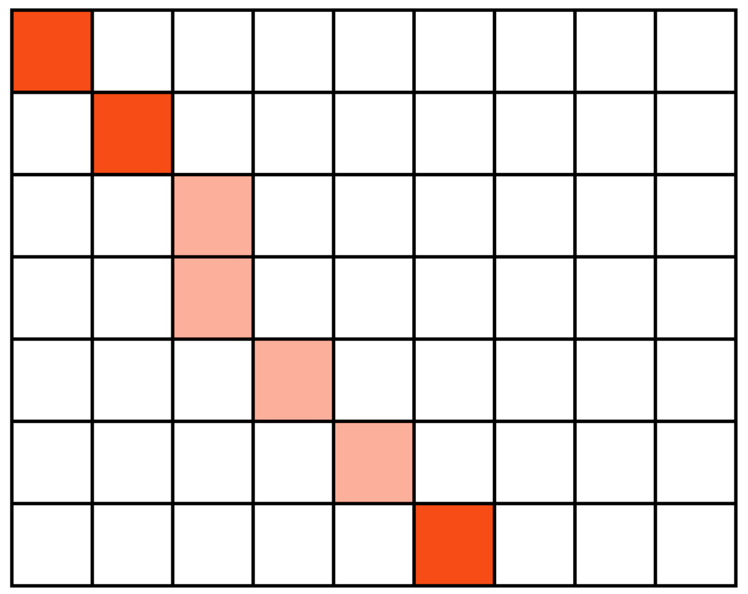
# Graph matching

**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

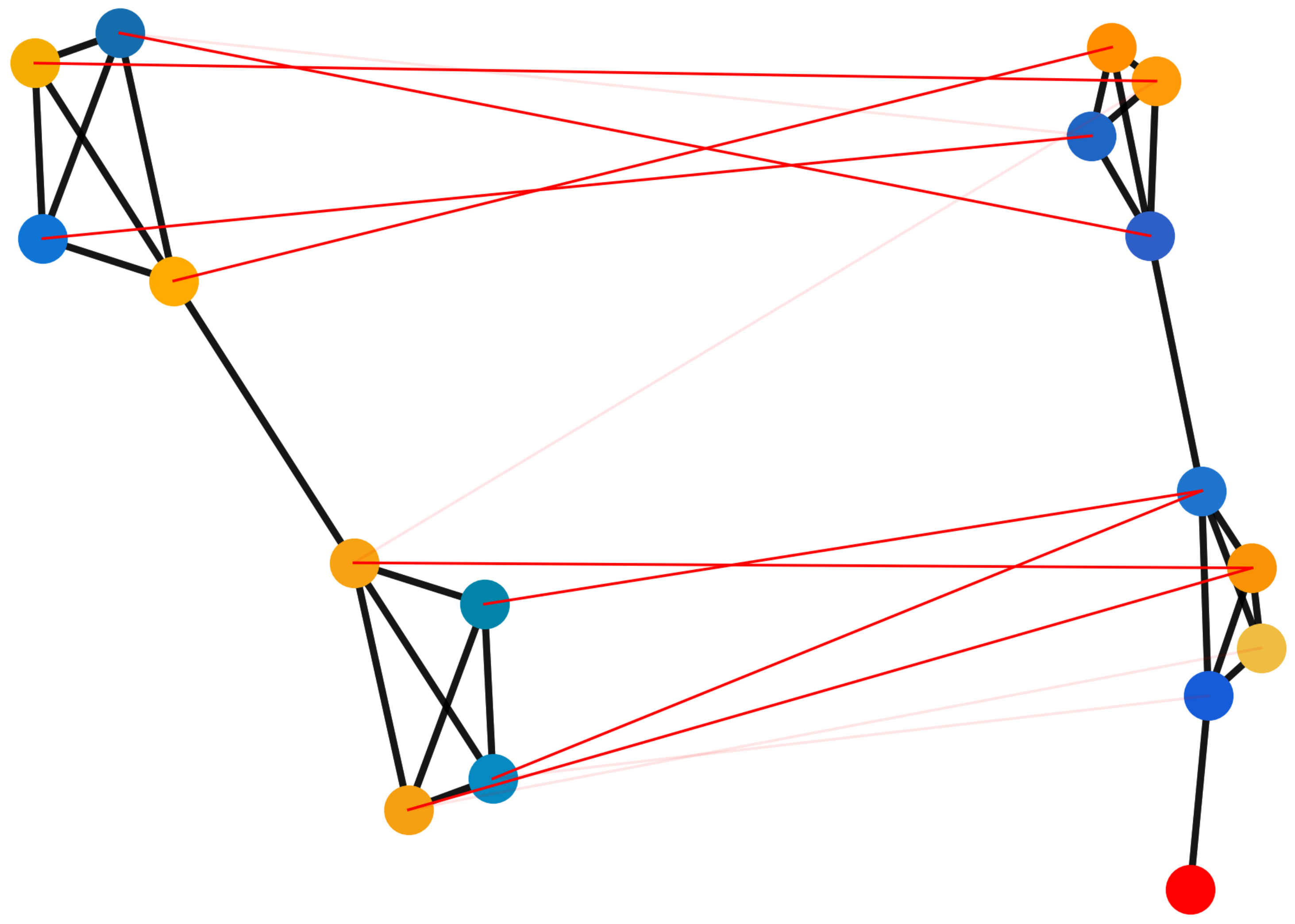
**Fused Unbalanced Gromov Wasserstein (FUGW) optimal transport loss [Thual et al., 2022]**

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

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]

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

match nodes with similar node features

preserve local geometry

discard nodes that do not have a good match

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

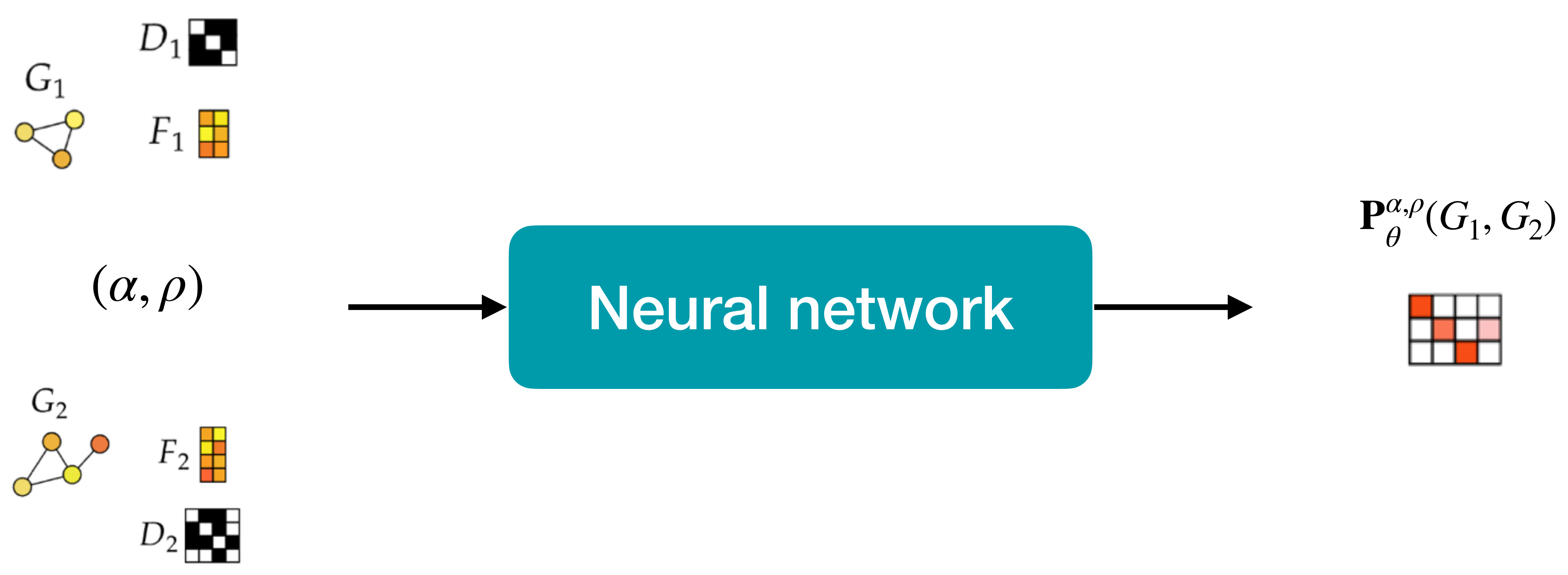
**Solve the OT problem**: batch coordinate descent with complexity $O(kn^3)$ for $k$ the number of iterations and $n$ the number of graph nodes.

→ unscalable for large graphs

# Predicting FUGW plans

**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 $(\alpha, \rho) \sim \mathscr{P}$.

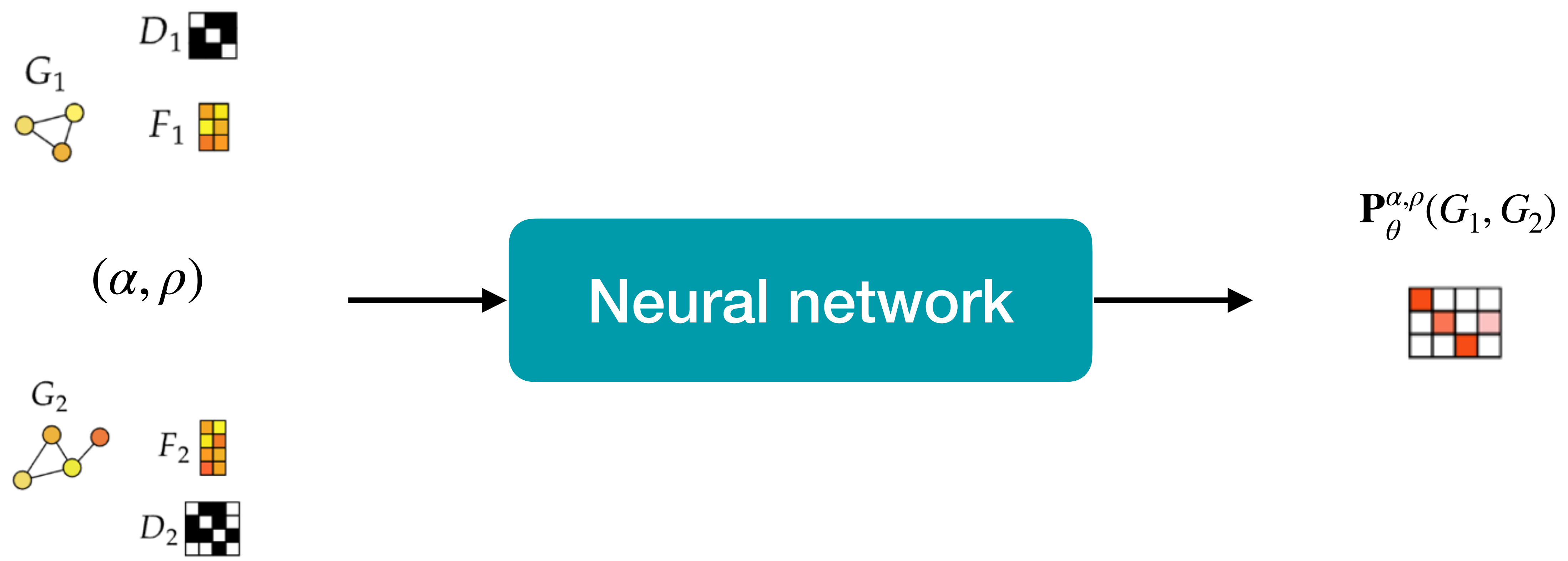
**Method:** Neural Network based on cross attention and Graph Convolutional Networks that predicts OT plans.



# Predicting FUGW plans

**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 on cross attention and Graph Convolutional Networks that predicts OT plans.

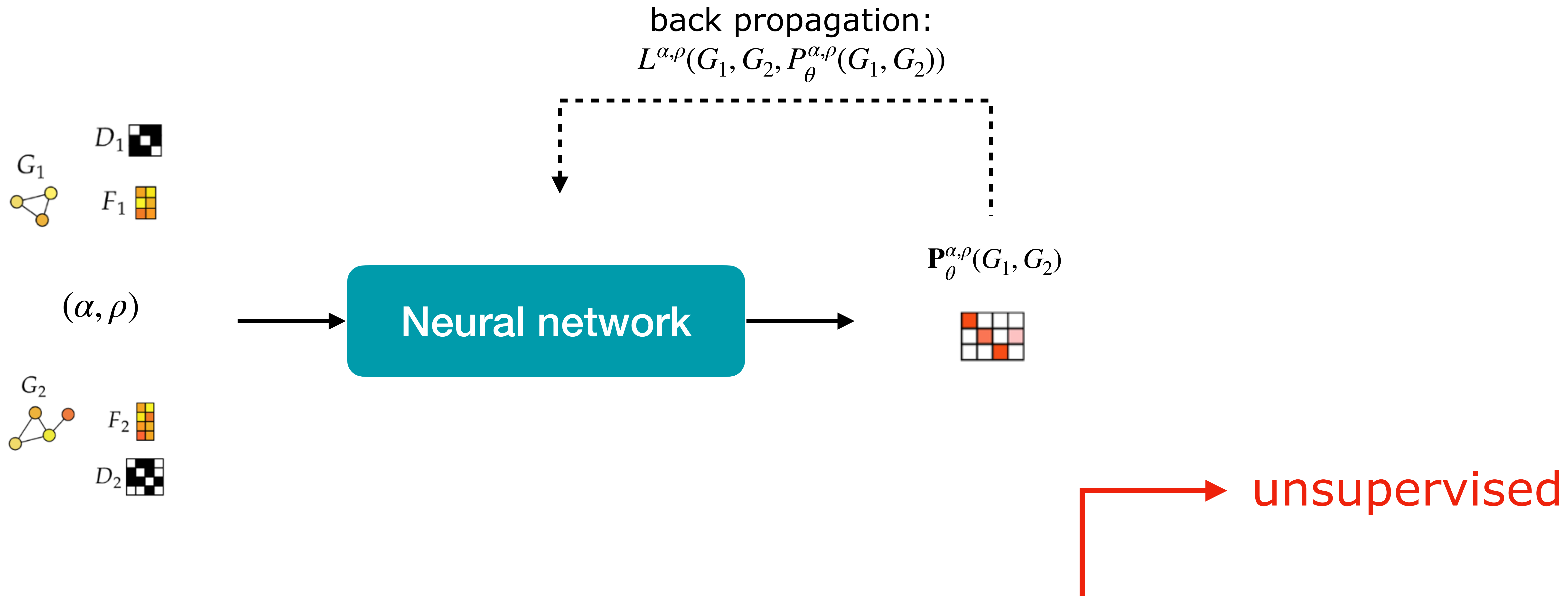



**Supervised training?** Unscalable

# Predicting FUGW plans

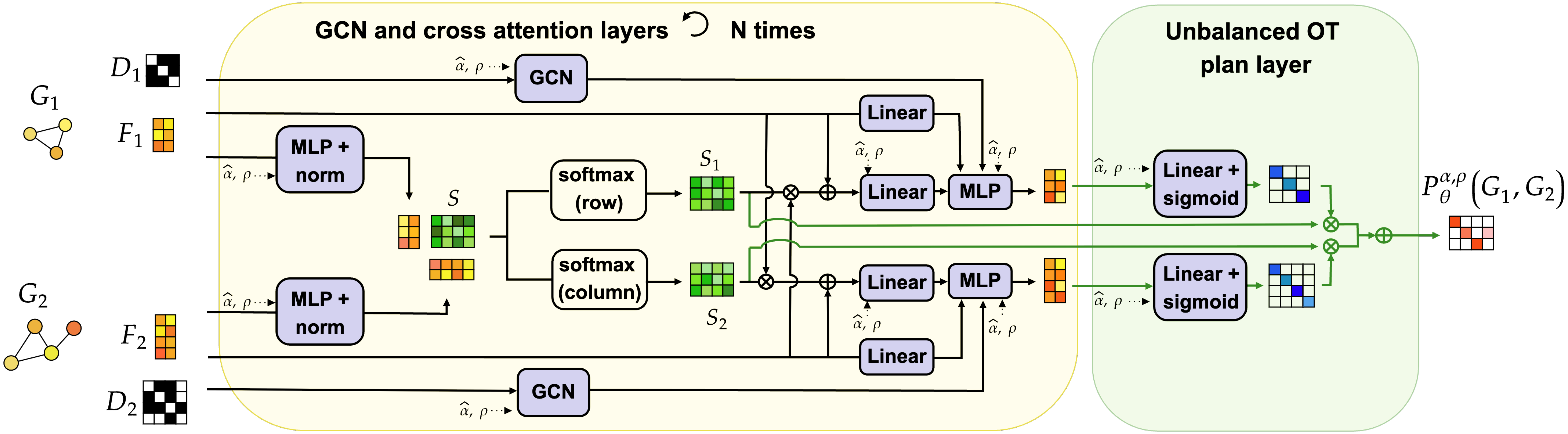
**Goal:** learn to predict FUGW plan $\mathbf{P}_\theta^{\alpha,\rho}(G_1, G_2)$ for all graph pairs $(G_1, G_2) \sim \mathscr{D}$ and parameters $(\alpha, \rho) \sim \mathscr{P}$.

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



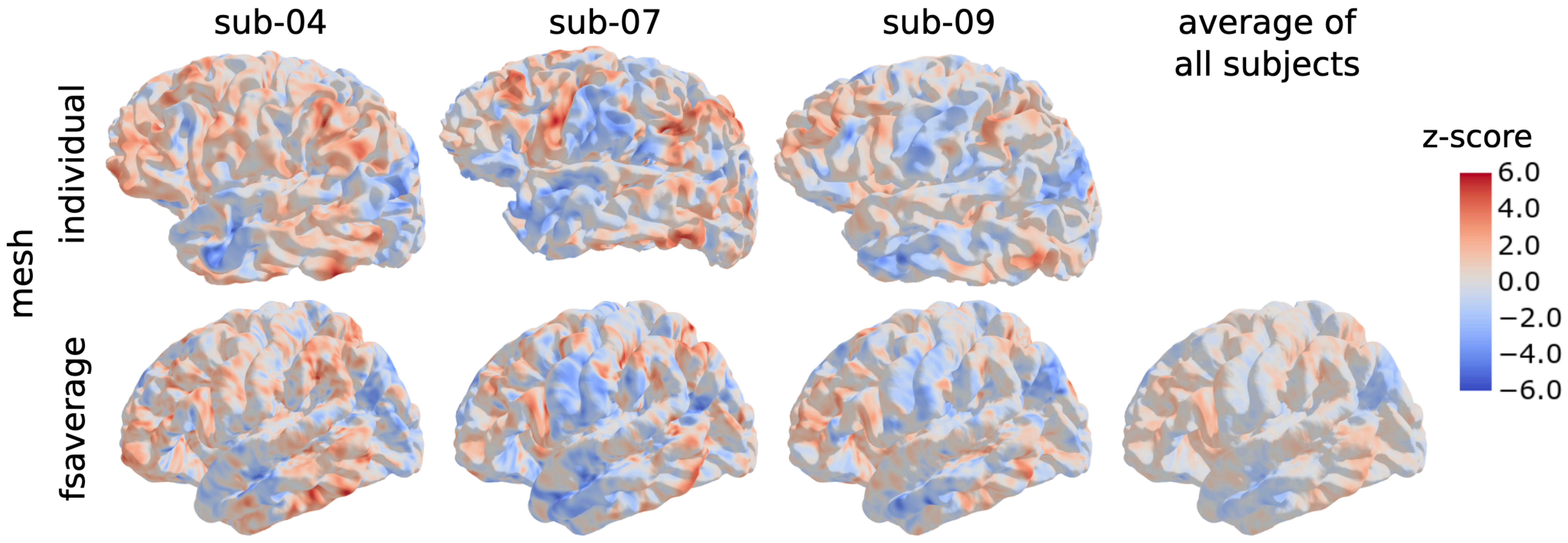

**Amortized optimisation (Amos et al., 2022):** $\min_\theta \mathbb{E}_{G_1, G_2 \sim \mathscr{D}, \alpha, \rho \sim \mathscr{P}} \left[ \mathsf{L}^{\alpha,\rho}(G_1, G_2, \mathbf{P}_\theta^{\alpha,\rho}(G_1, G_2)) \right]$

# Unbalanced learning of Optimal Transport plans (ULOT)




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

# 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.

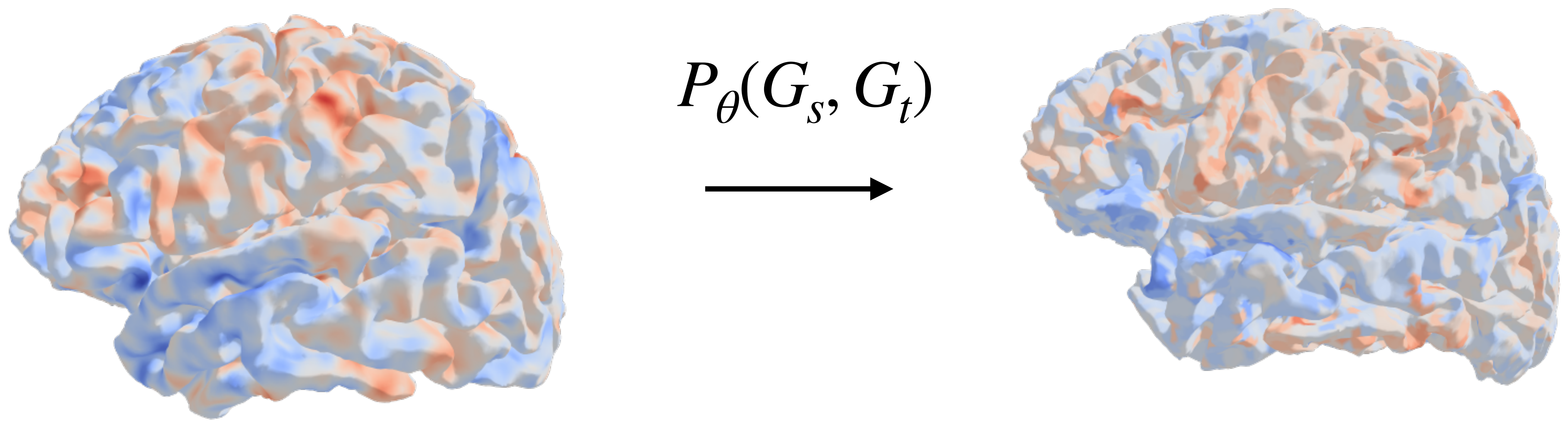
Graphs constructed from the brain surface geometries and functional MRI activations for different tasks from the Individual Brain Charting dataset, 1000 nodes.

# FUGW for brain alignment

**Computational time:**

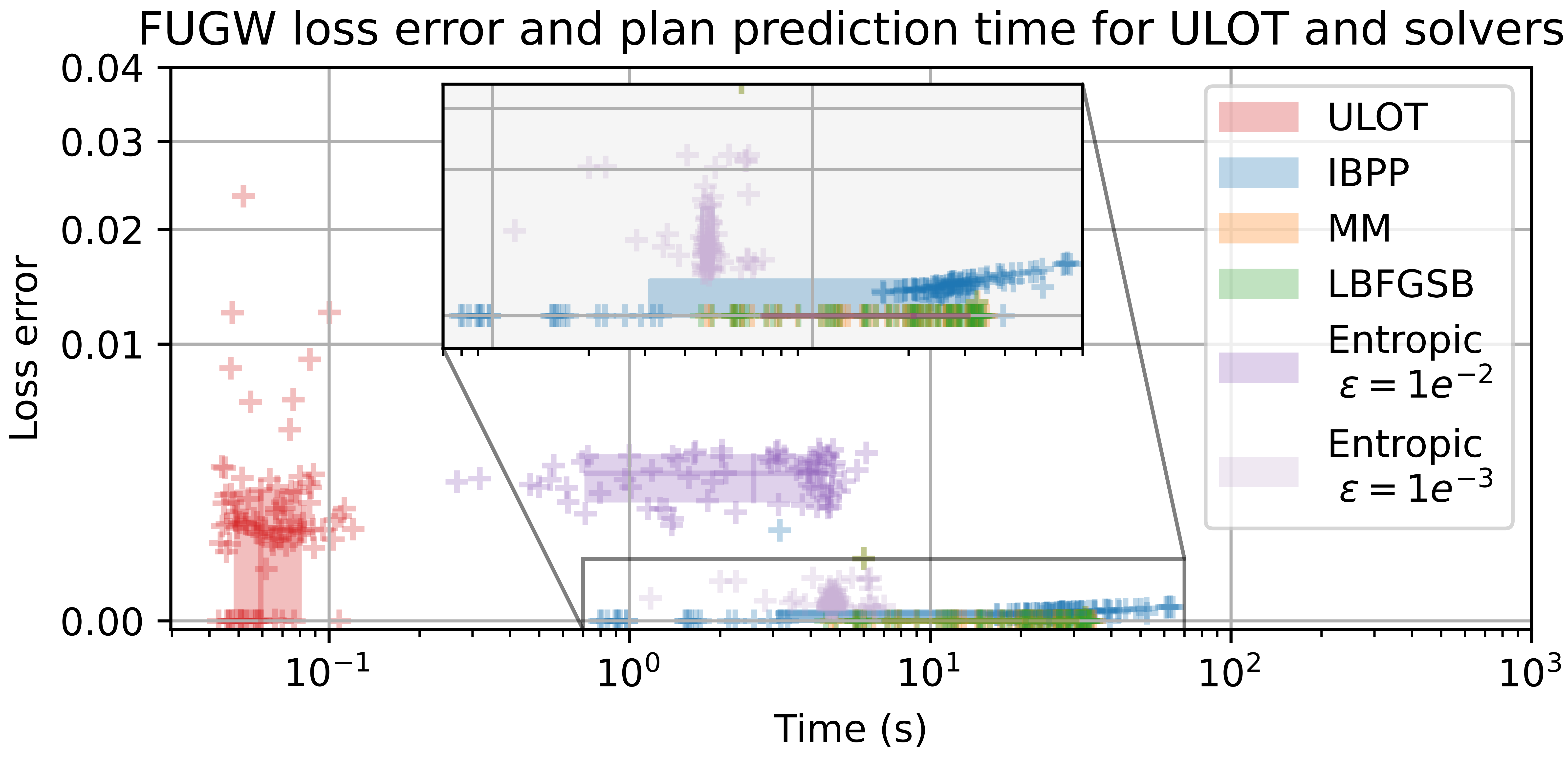
- 4 minutes for aligning one pair with 10k vertices on a single GPU.
- Limits applications for computing alignments on large populations.

**Proposed solution:** Align brains with ULOT transport plans




source subject $G_s$ target subject $G_t$

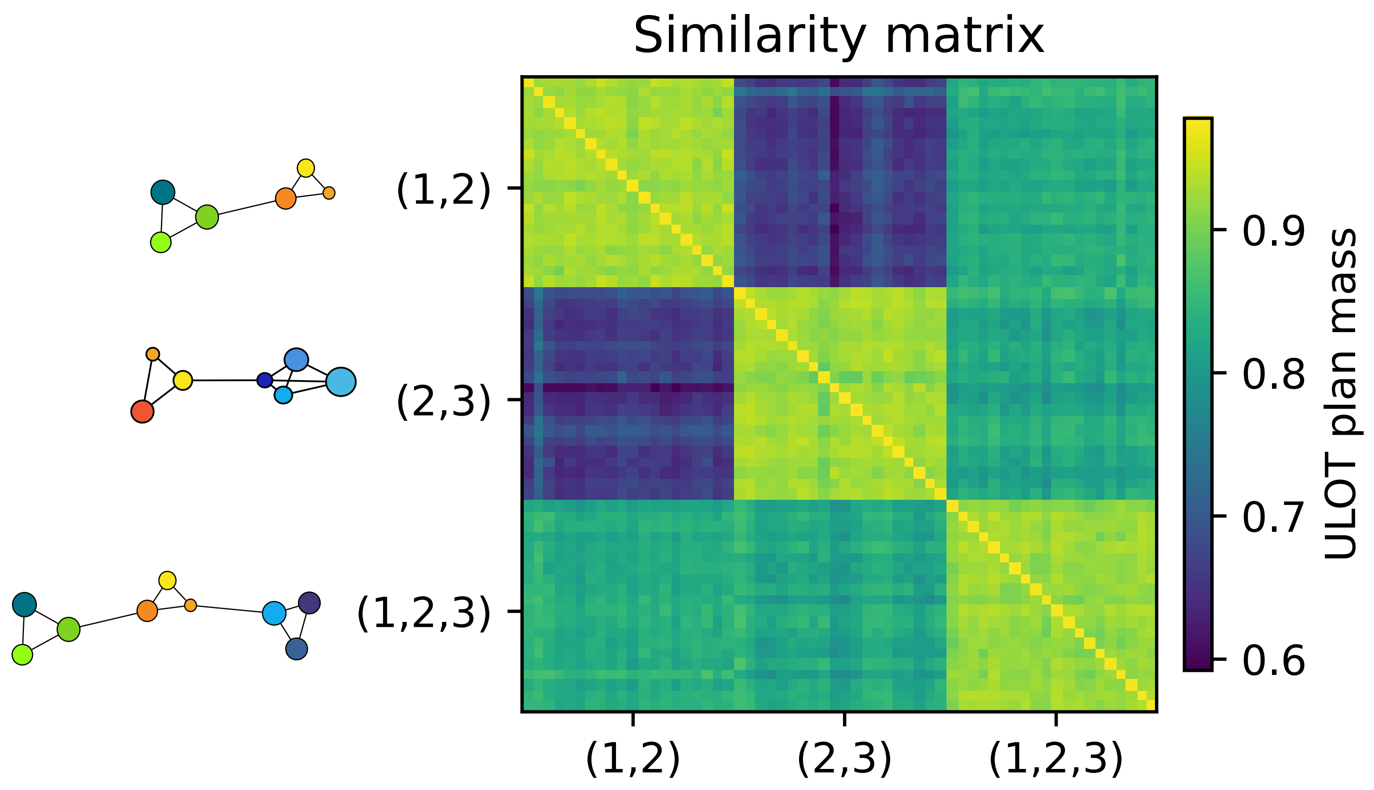
# Results on fMRI data



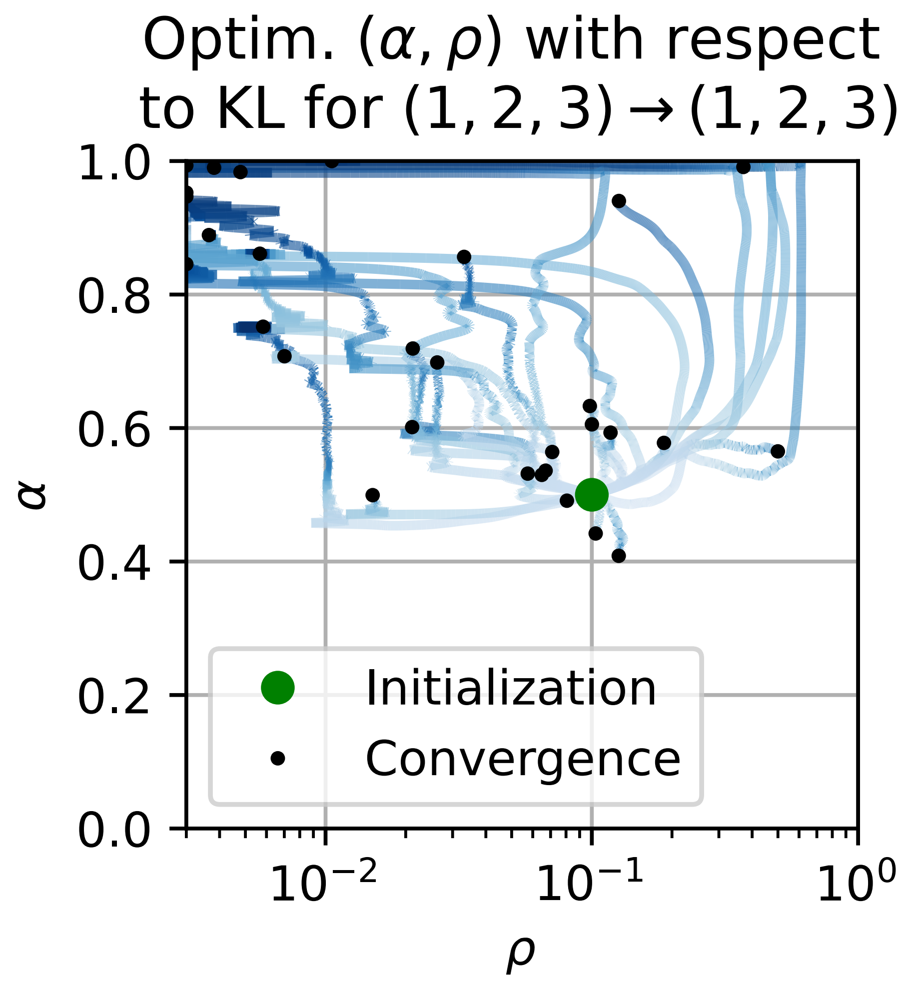

ULOT predicted plans with low error compared to solvers, and up to 100 times faster: allows extensive parameter selection and scalability to large graphs.

# Applications

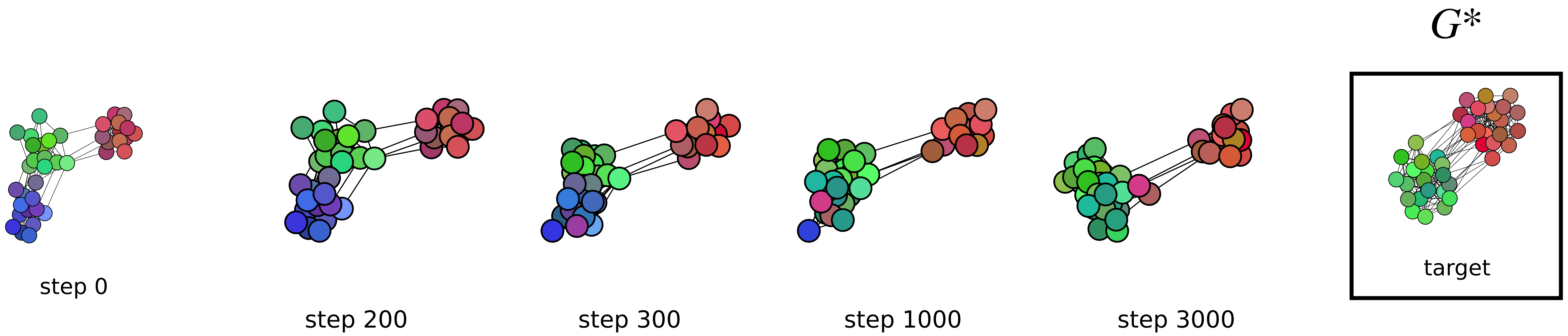
- A **quadratic graph similarity measure**: mass of the ULOT transport plans




- **FUGW hyperparameter selection** with bi level optimization



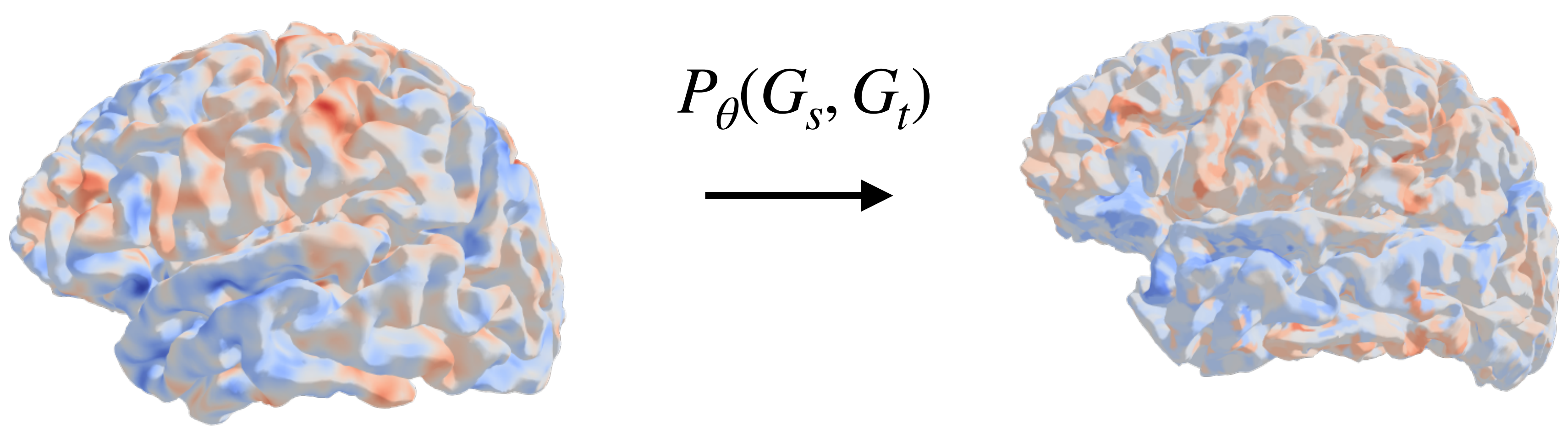

- Optimization of **functionals of the ULOT plan**



# 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).
- Limitation:
  - transport plan error can still be a problem in some applications where high precision is needed: ULOT can be used to initialize an OT solver.

**paper**




source subject $G_s$ target subject $G_t$