Unsupervised learning for Optimal Transport plan prediction between unbalanced graphs

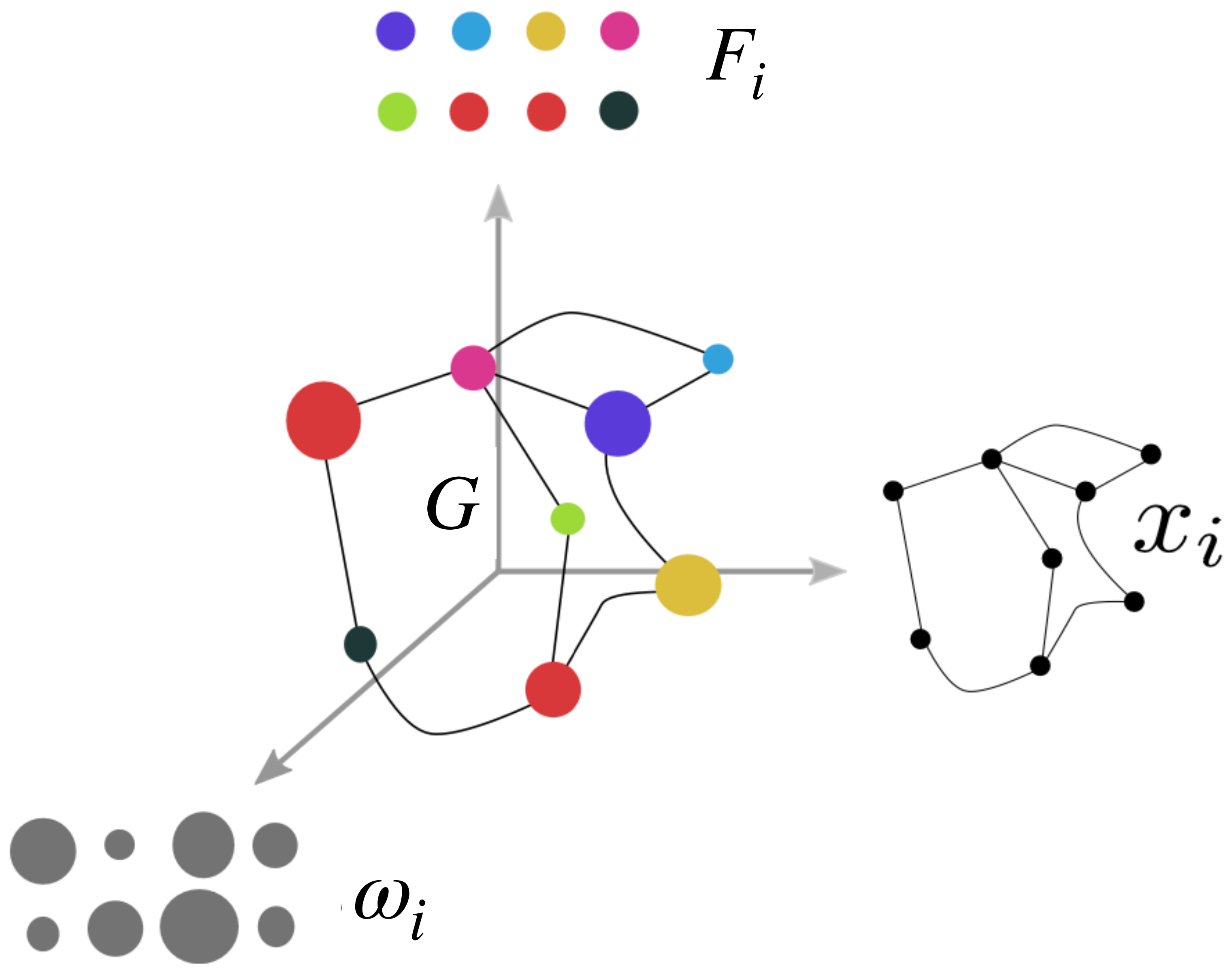
Sonia Mazelet, Rémi Flamary, Bertrand Thirion

Workshop Fondements mathématiques de l'IA - 10/12/2025

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



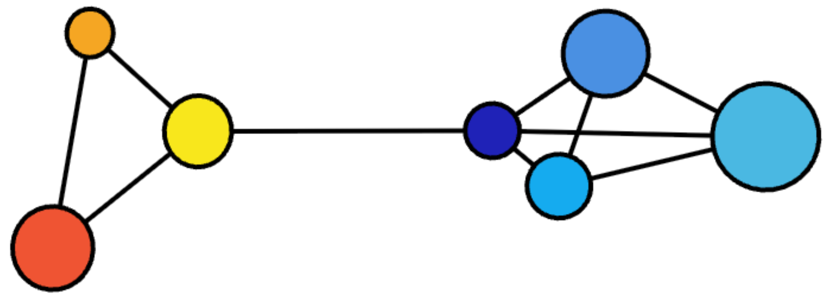
$$\mu = \sum_i \omega_i \delta_{(x_i, F_i)}$$

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

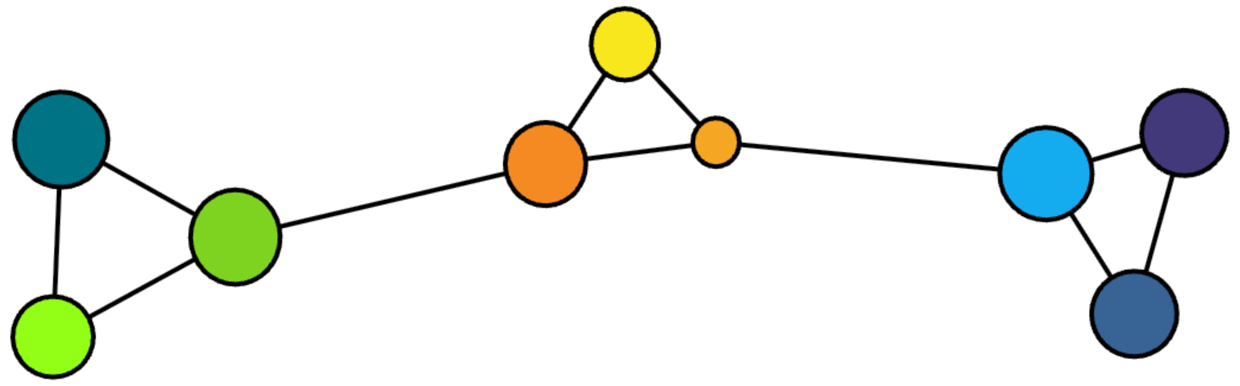
$$\mu_X = \sum_i \omega_i \delta_{x_i}$$

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.



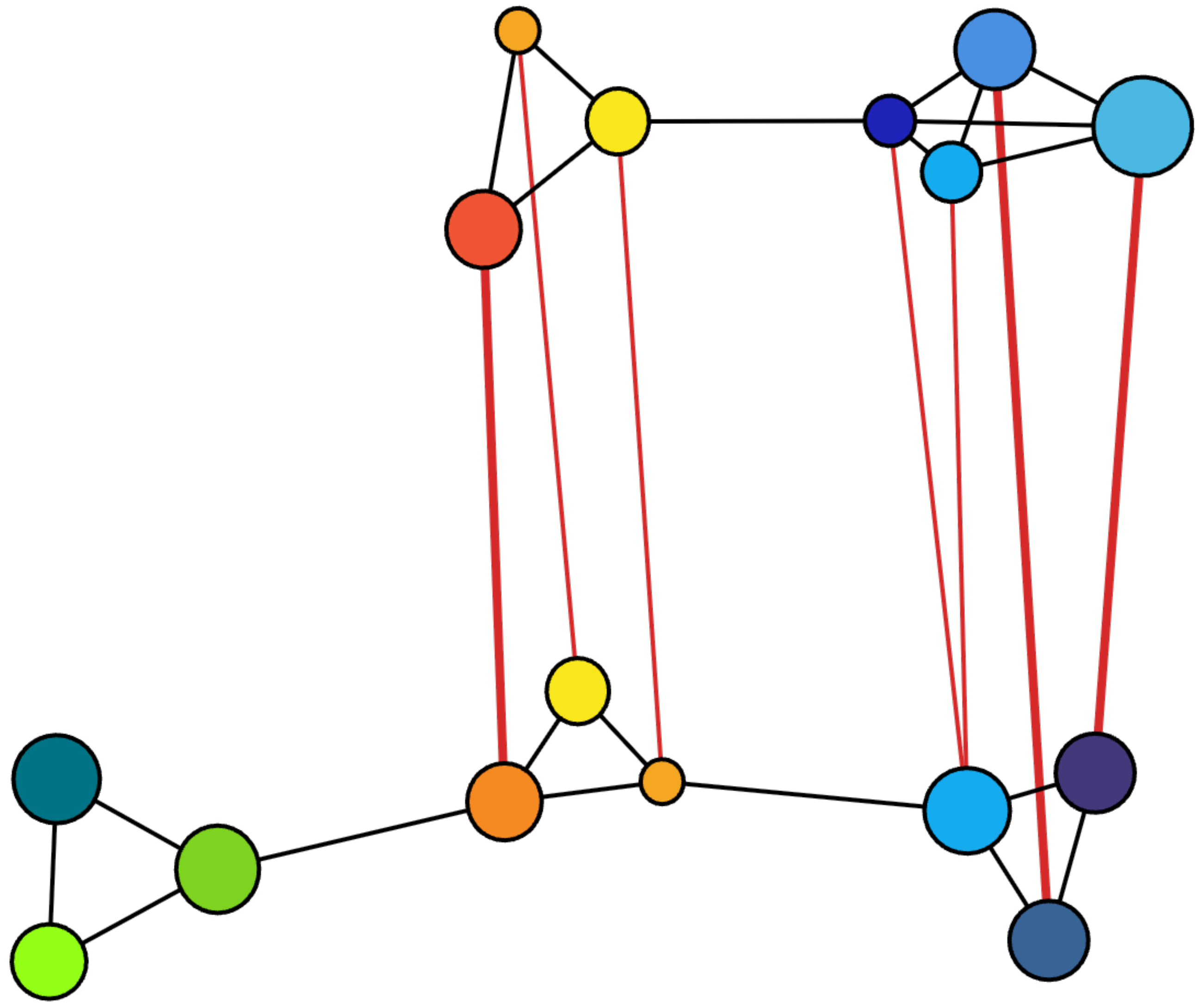
$$G_1 = (F_1, D_1, \omega_1)$$



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

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.

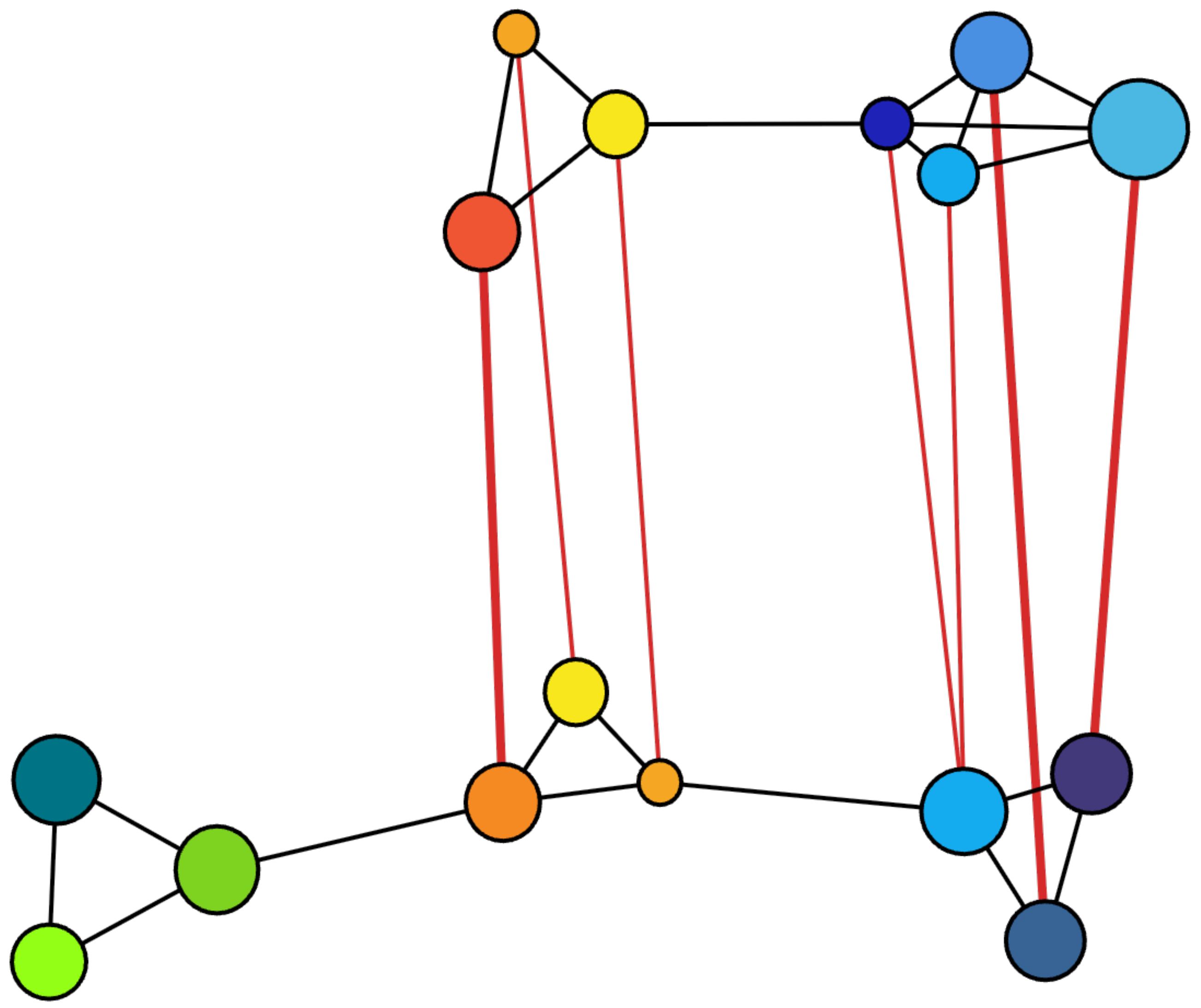


$$G_1 = (F_1, D_1, \omega_1)$$

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

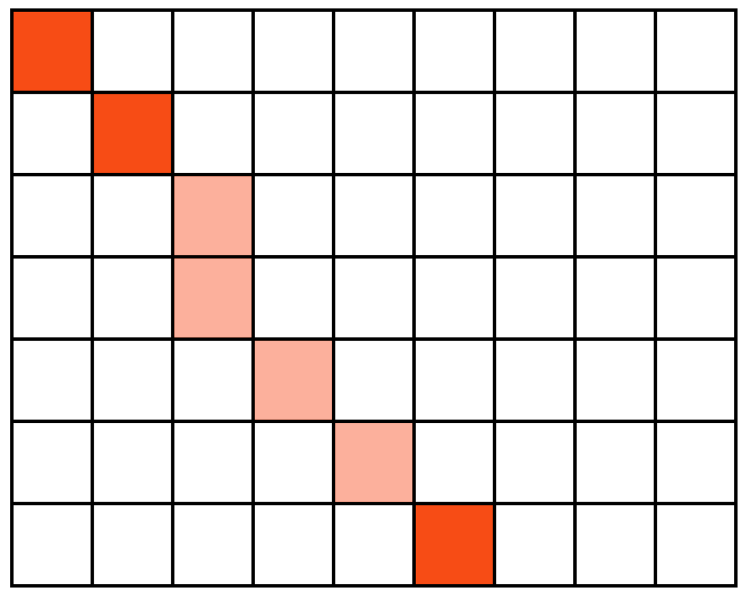
Graph matching

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

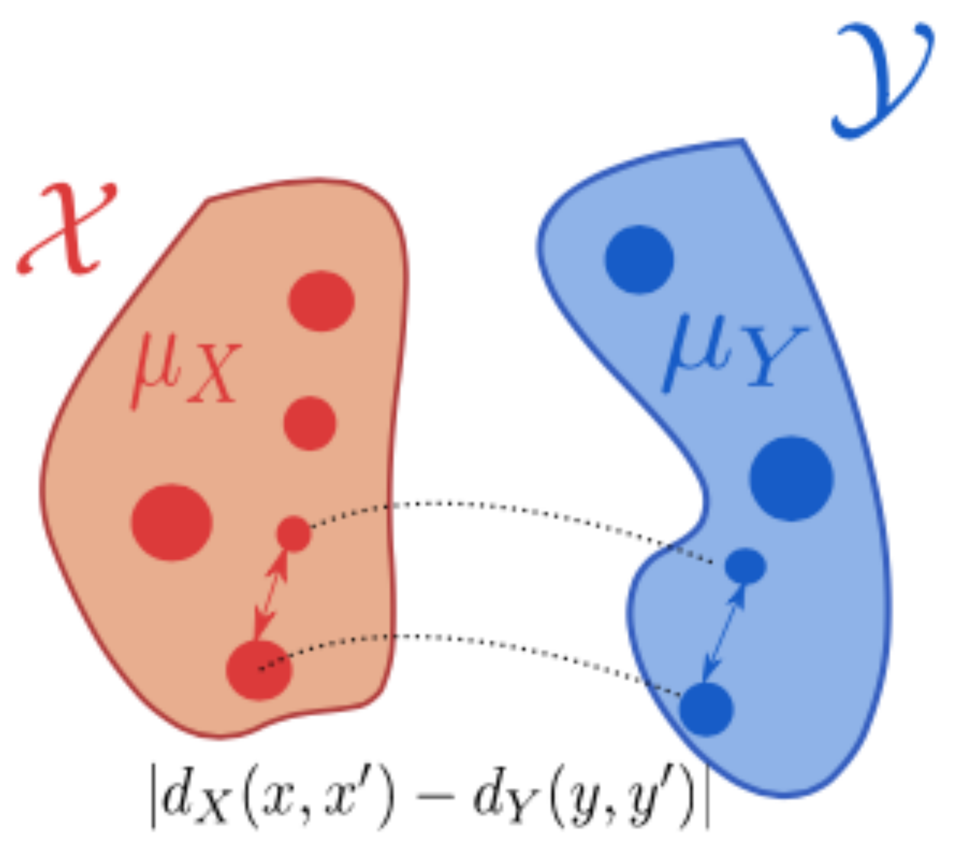
Fused Unbalanced Gromov Wasserstein (FUGW) optimal transport loss [Thual et al., 2022]

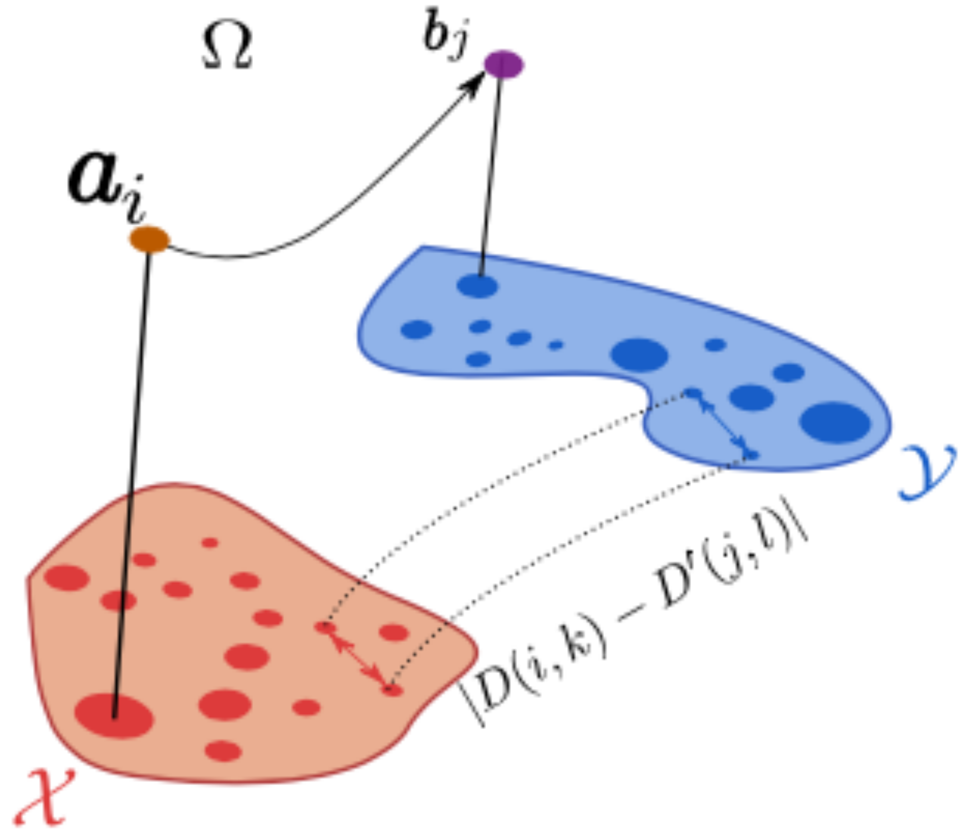
$$\mathcal{L}^{\alpha,\rho}(G_1, G_2, \mathbf{P}) = (\mathbf{1}-\alpha)\sum_{i,j=1}^{\mathbf{n_1},\mathbf{n_2}} \left\|(\mathbf{F_1})_\mathbf{i} - (\mathbf{F_2})_\mathbf{j}\right\|_\mathbf{2}^\mathbf{2} \mathbf{P_{i,j}} + \alpha \sum_{i,j,k,l=1}^{\mathbf{n_1},\mathbf{n_2},\mathbf{n_1},\mathbf{n_2}} \left|(\mathbf{D_1})_{\mathbf{i,k}}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

match nodes with similar node features

preserve local geometry

discard nodes that do not have a good match





Optimal transport distance between graphs

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

$$\mathcal{L}^{\alpha,\rho}(G_1, G_2, \mathbf{P}) = (\mathbf{1}-\alpha) \sum_{i,j=1}^{\mathbf{n_1,n_2}} \left\|(\mathbf{F_1})_\mathbf{i} - (\mathbf{F_2})_\mathbf{j}\right\|_2^2 \mathbf{P_{i,j}} + \alpha \sum_{i,j,k,l=1}^{\mathbf{n_1,n_2,n_1,n_2}} \left|(\mathbf{D_1})_\mathbf{i,k} - (\mathbf{D_2})_\mathbf{j,l}\right|^2 \mathbf{P_{i,j}} \mathbf{P_{k,l}} + \rho \left(\mathbf{KL}(\mathbf{P}_{\#1} \otimes \mathbf{P}_{\#1} | \omega_1 \otimes \omega_1) + \mathbf{KL}(\mathbf{P}_{\#2} \otimes \mathbf{P}_{\#2} | \omega_2 \otimes \omega_2)\right)$$

match nodes with similar node features

preserve local geometry

discard nodes that do not have a good match

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

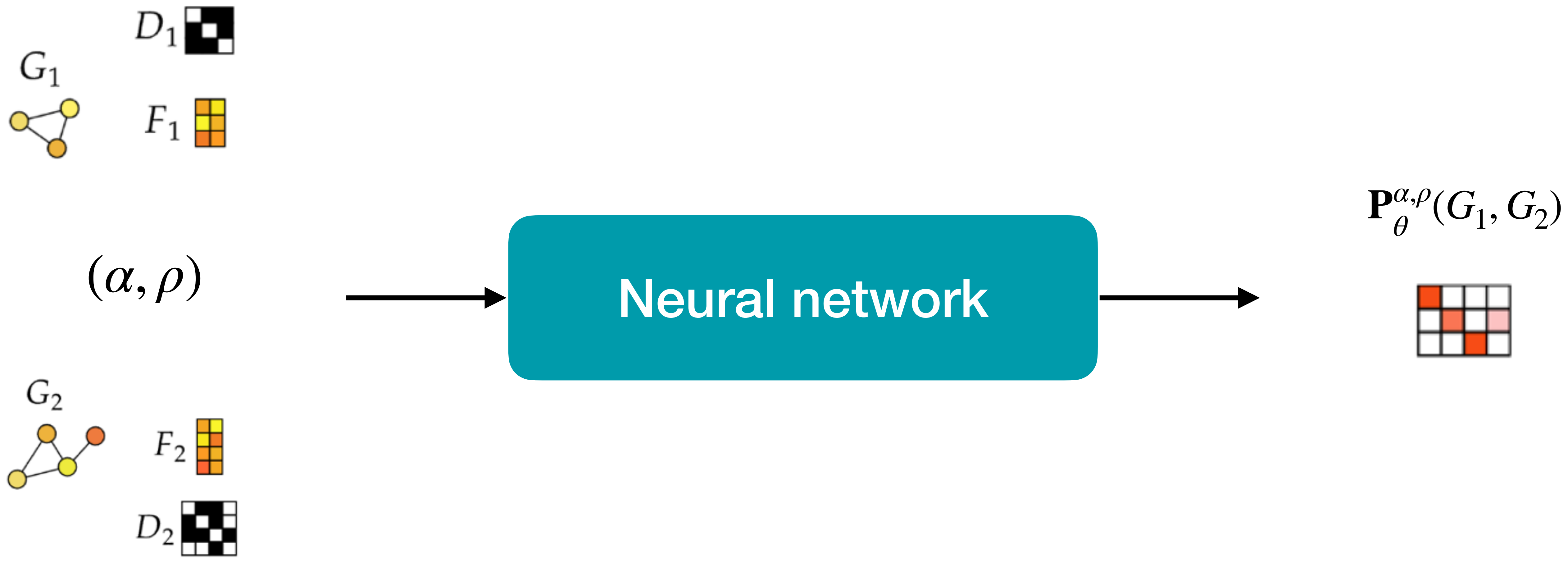
Solve the OT problem: batch coordinate descent with complexity $O(kn^3)$ for k the number of iterations and n the number of graph nodes.

→ 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 $(\alpha, \rho) \sim \mathscr{P}$.

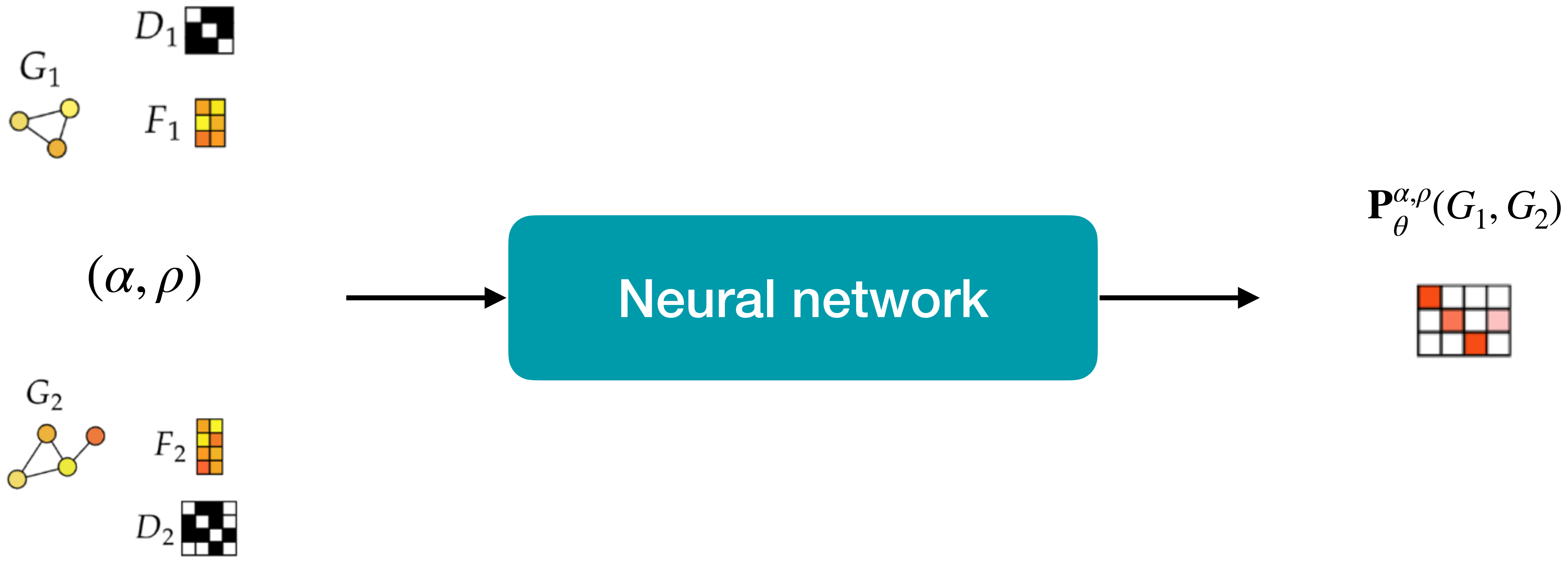
Method: Neural Network based cross attention and Graph Convolutional Networks that predicts OT plans.



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.

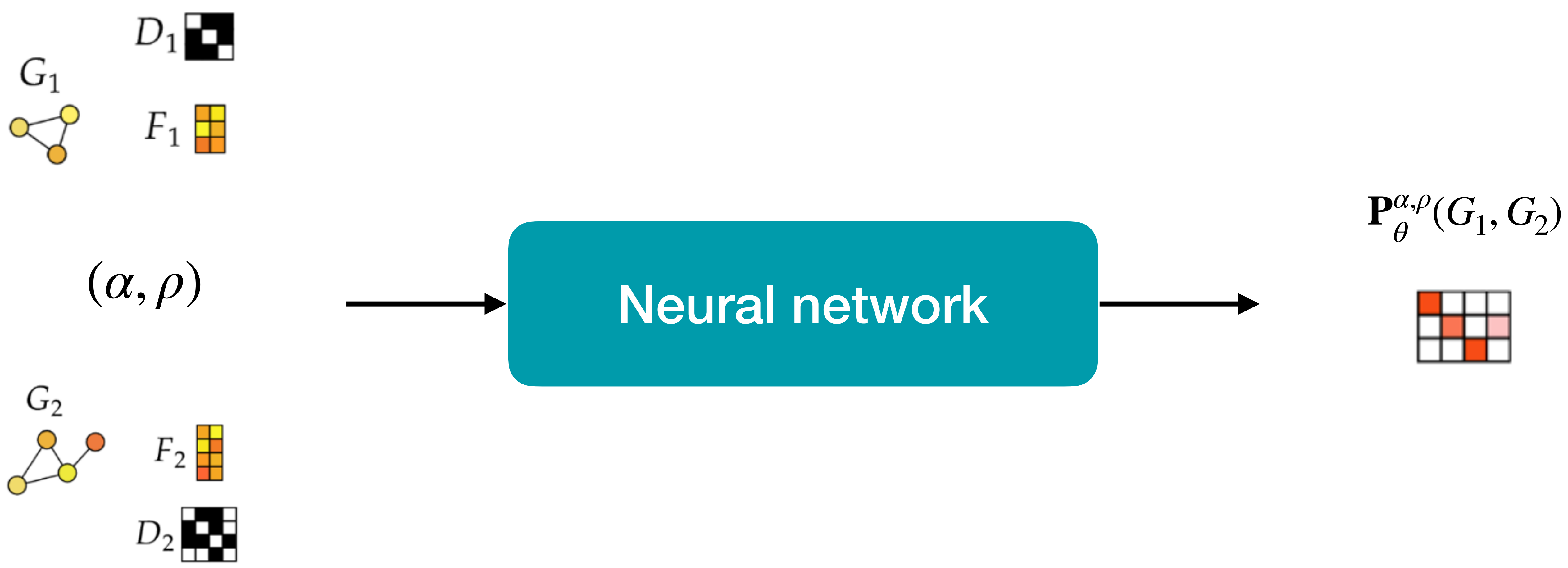


Supervised training? Unscalable

Predicting FUGW plan

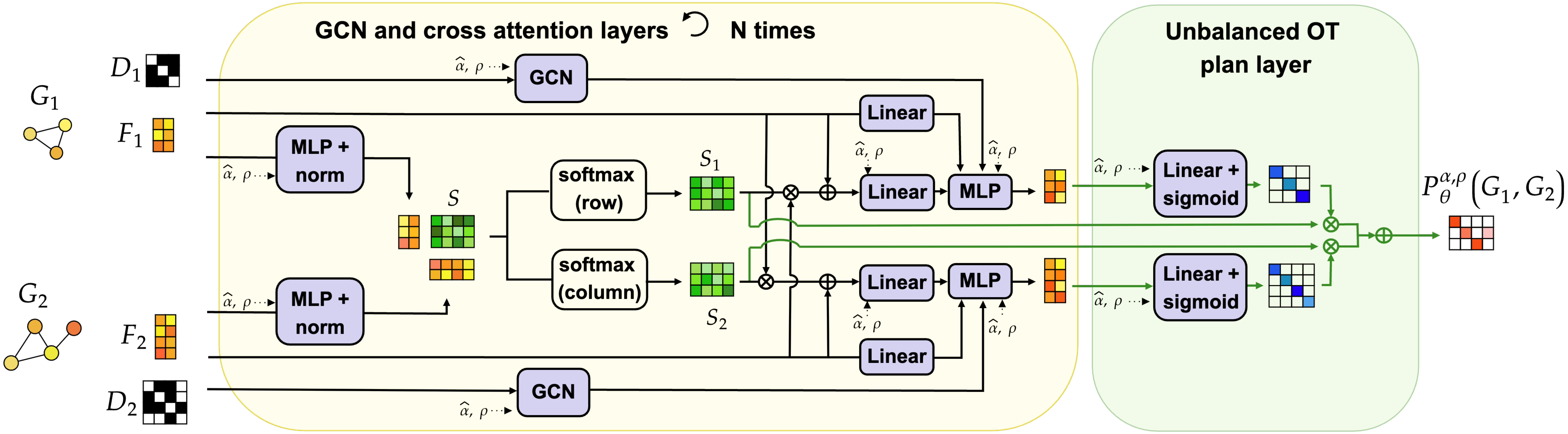
Goal: learn to predict FUGW plan $\mathbf{P}_\theta^{\alpha,\rho}(G_1, G_2)$ for all graph pairs $(G_1, G_2) \sim \mathscr{D}$ and parameters $(\alpha, \rho) \sim \mathscr{P}$.

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



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

Unbalanced learning of Optimal Transport plans (ULOT)

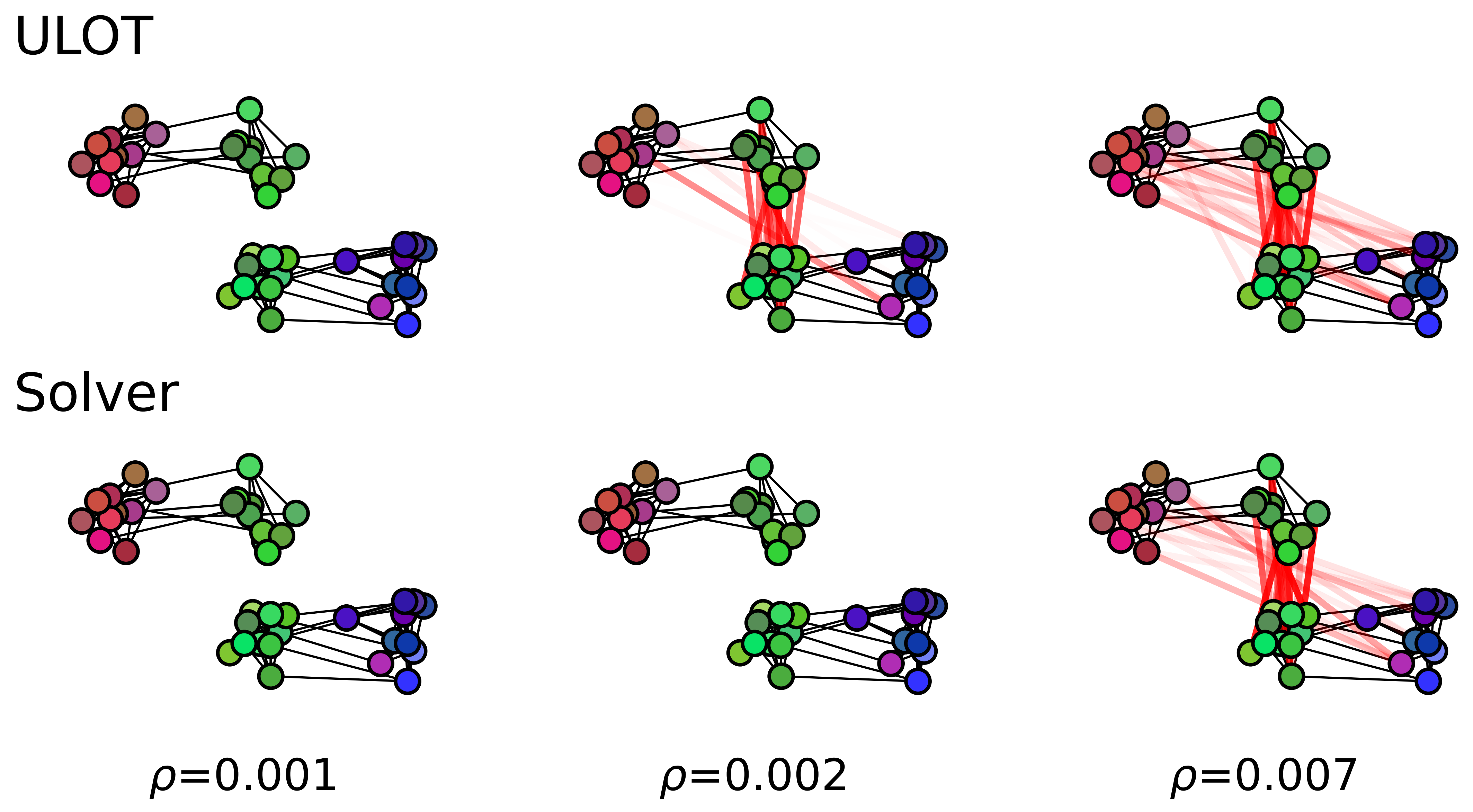


Complexity: $O(n^2)$ for n the number of graph nodes

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

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



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

node features

structure

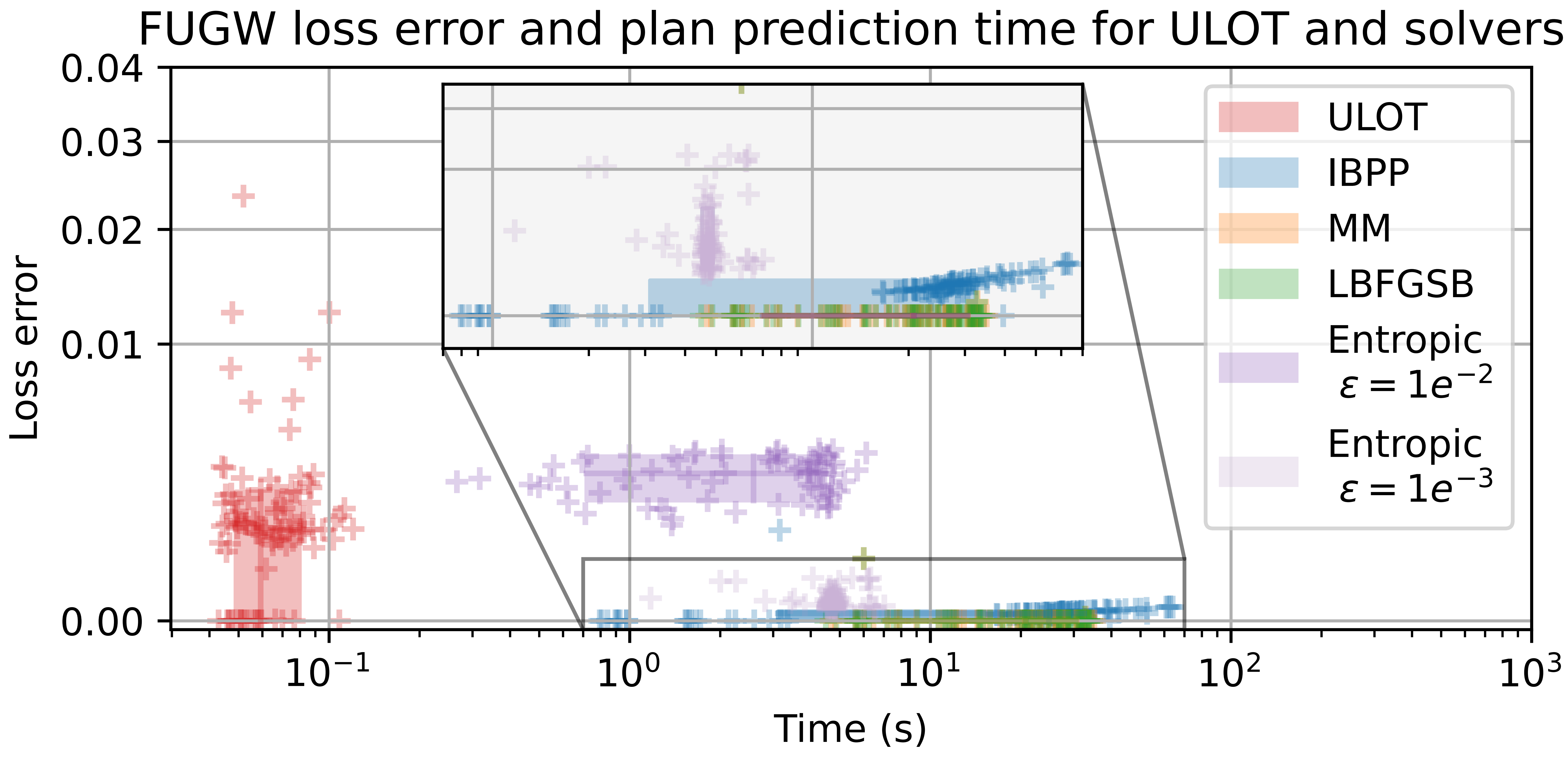
marginals regularization

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.

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.

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