Stacking multiple optimal transport policies to map functional connectomes

Javid Dadashkarimi

Matthew Rosenblatt, Amin Karbasi, Dustin Scheinost

Computer Science Department
Yale University

57th Annual Conference on Information Sciences and Systems

Baltimore, MD

March, 2023

Contact information:
javid.dadashkarimi@yale.edu

- Functional magnetic resonance imaging (fMRI) revolutionized the field of neuroscience.
- We have access to a vastly large amount of insightful data from our brains.
- Researchers use these data to understand how the human brain works, to associate the brain with our behaviors, to investigate individual differences, or to study brain alterations in neuropsychiatric disorders.

How does the human brain function?

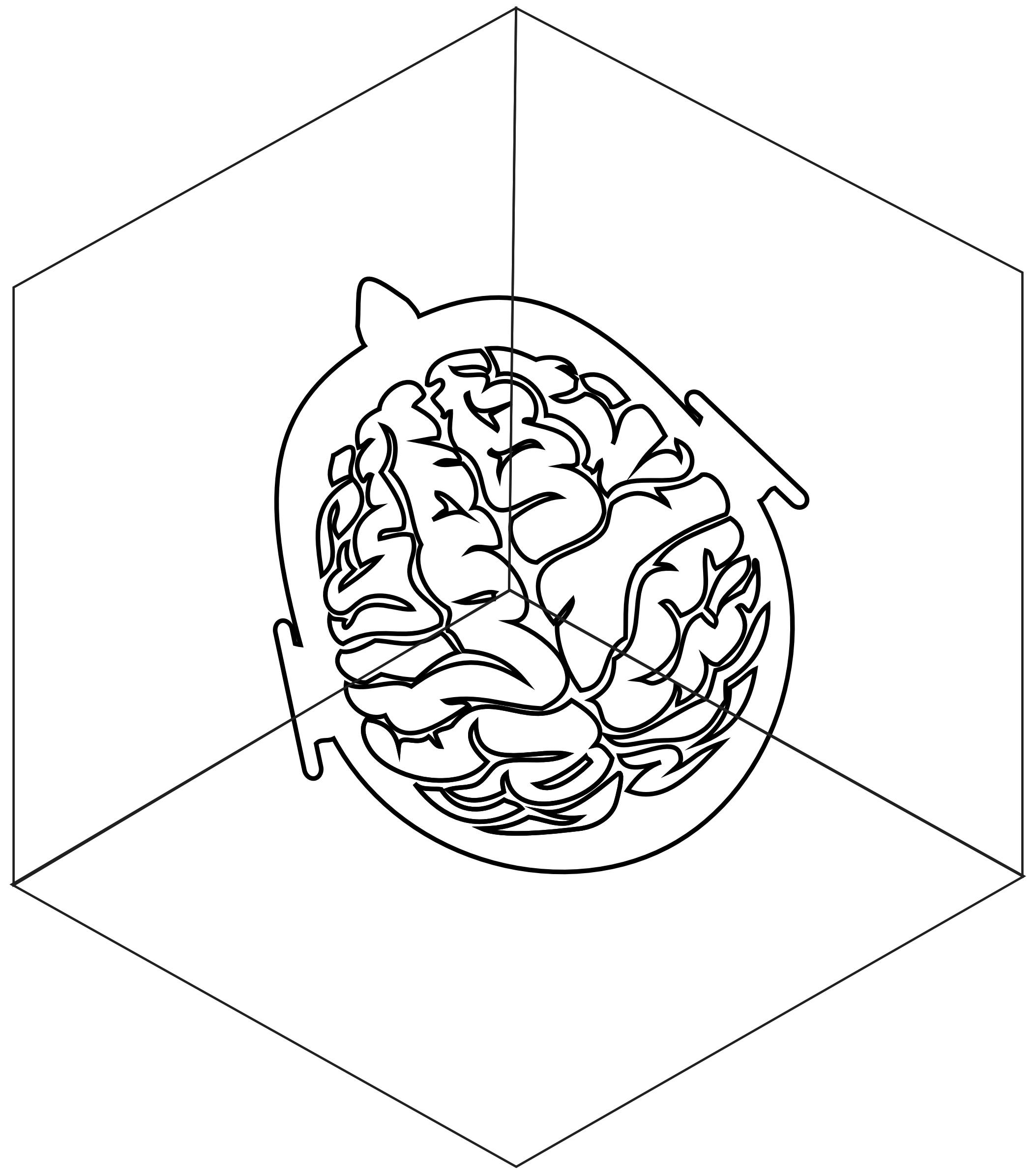
Functional magnetic resonance imaging (fMRI)

1. Associating brain and behavior
2. Studying group differences

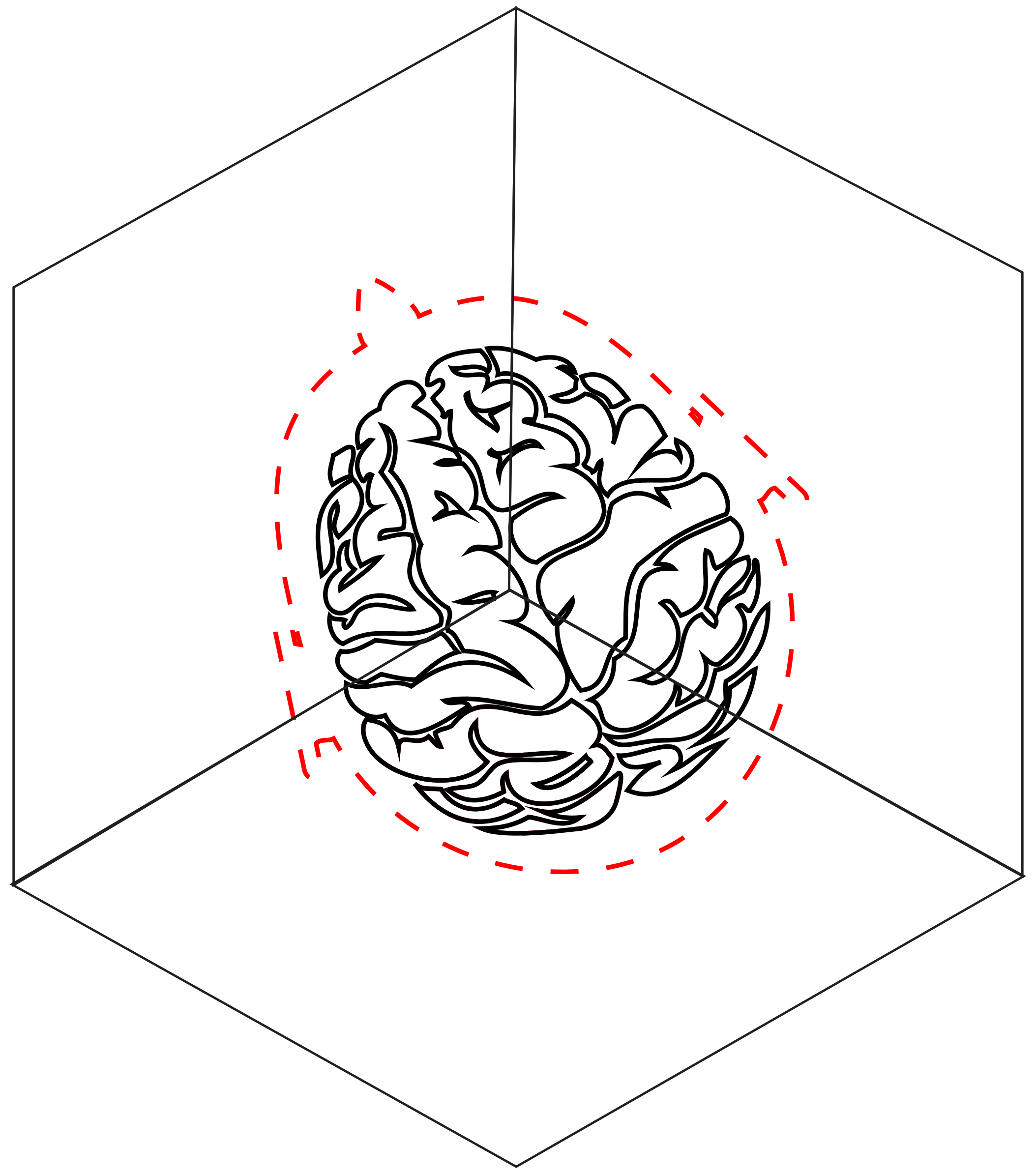
Functional Connectivity

Widely used in neuroscience to understand the functional organization of the brain.

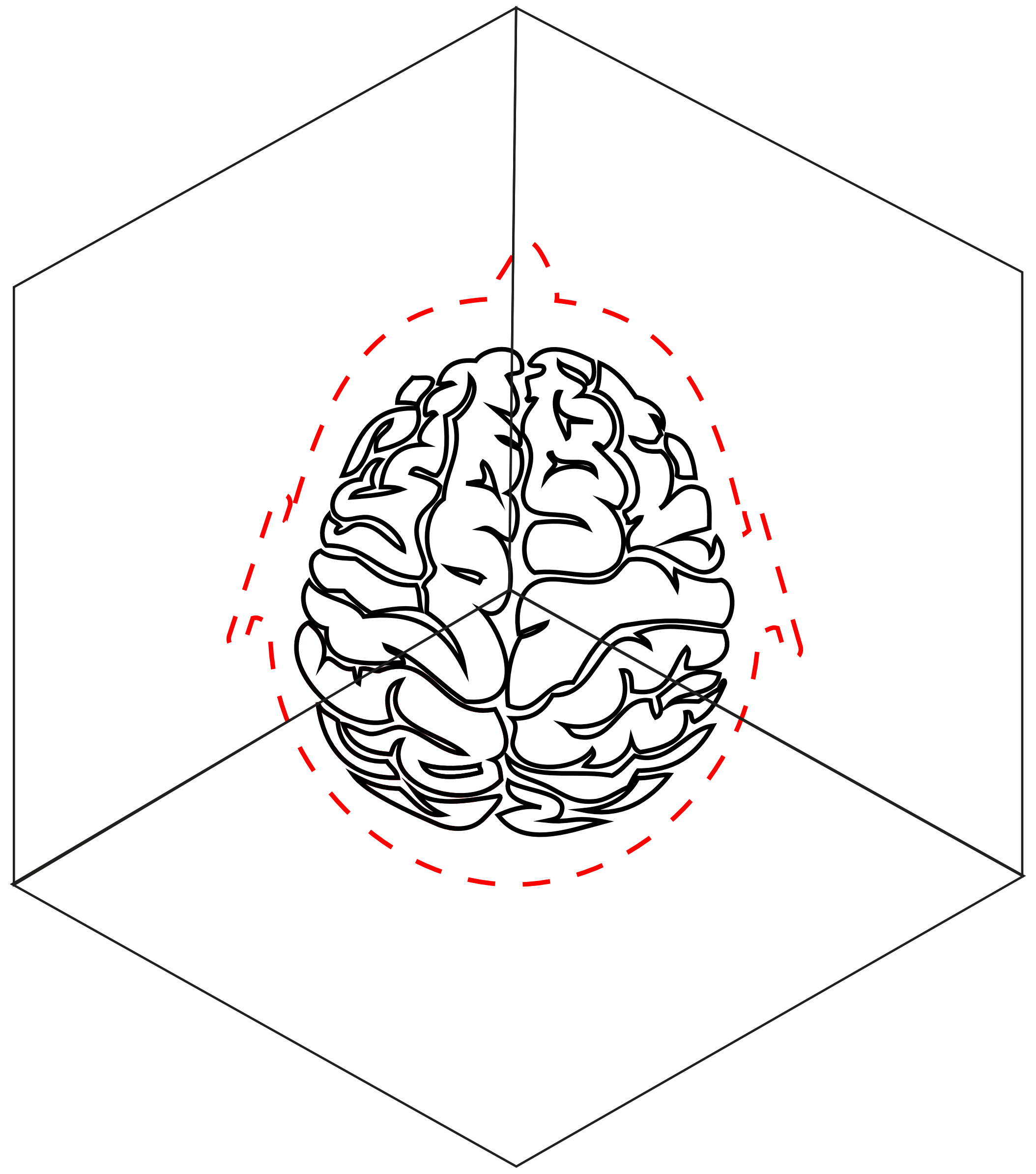
1. What are connectomes
2. How to make functional connectivity
3. Applications in neuroscience



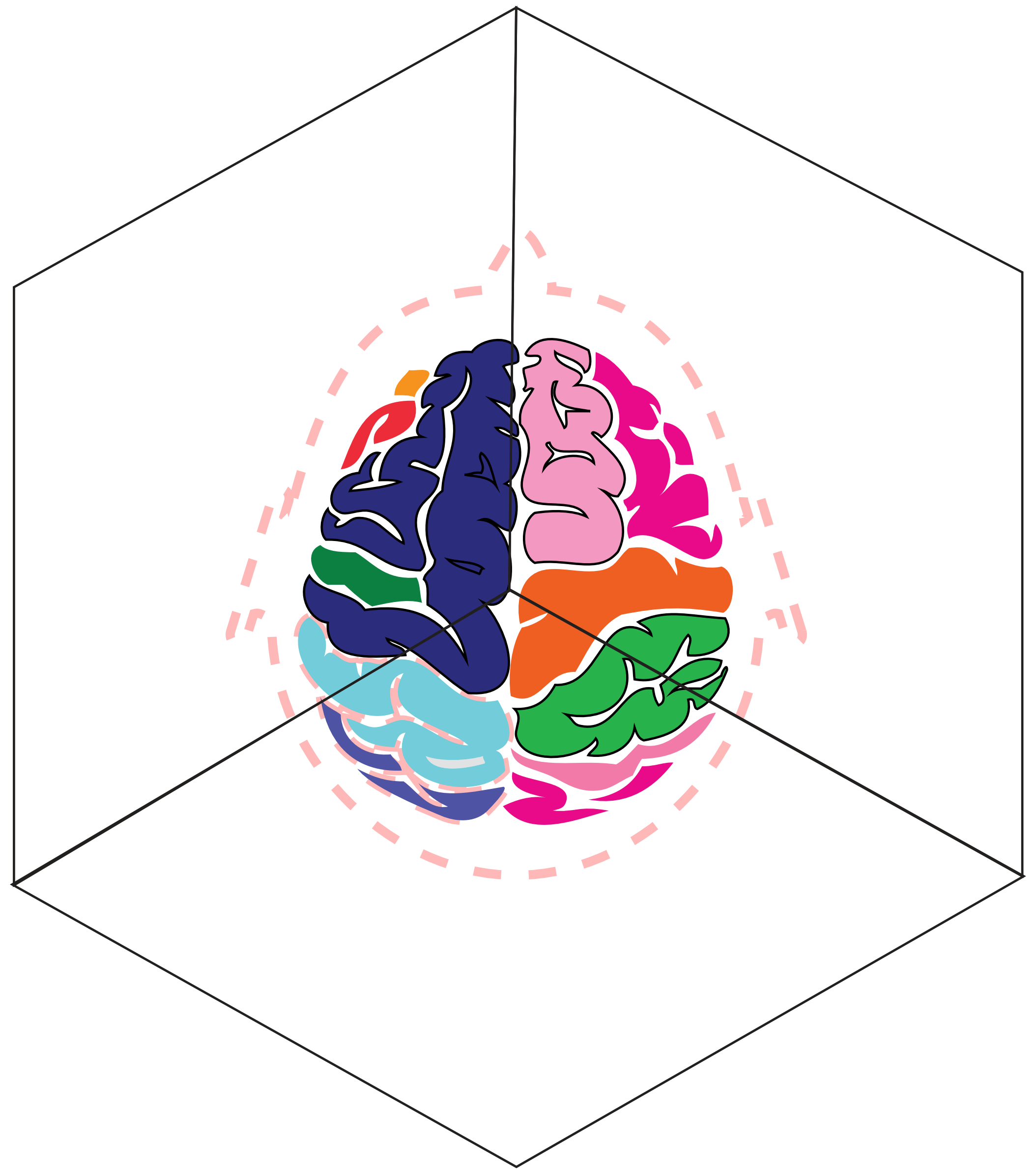
Motion correction



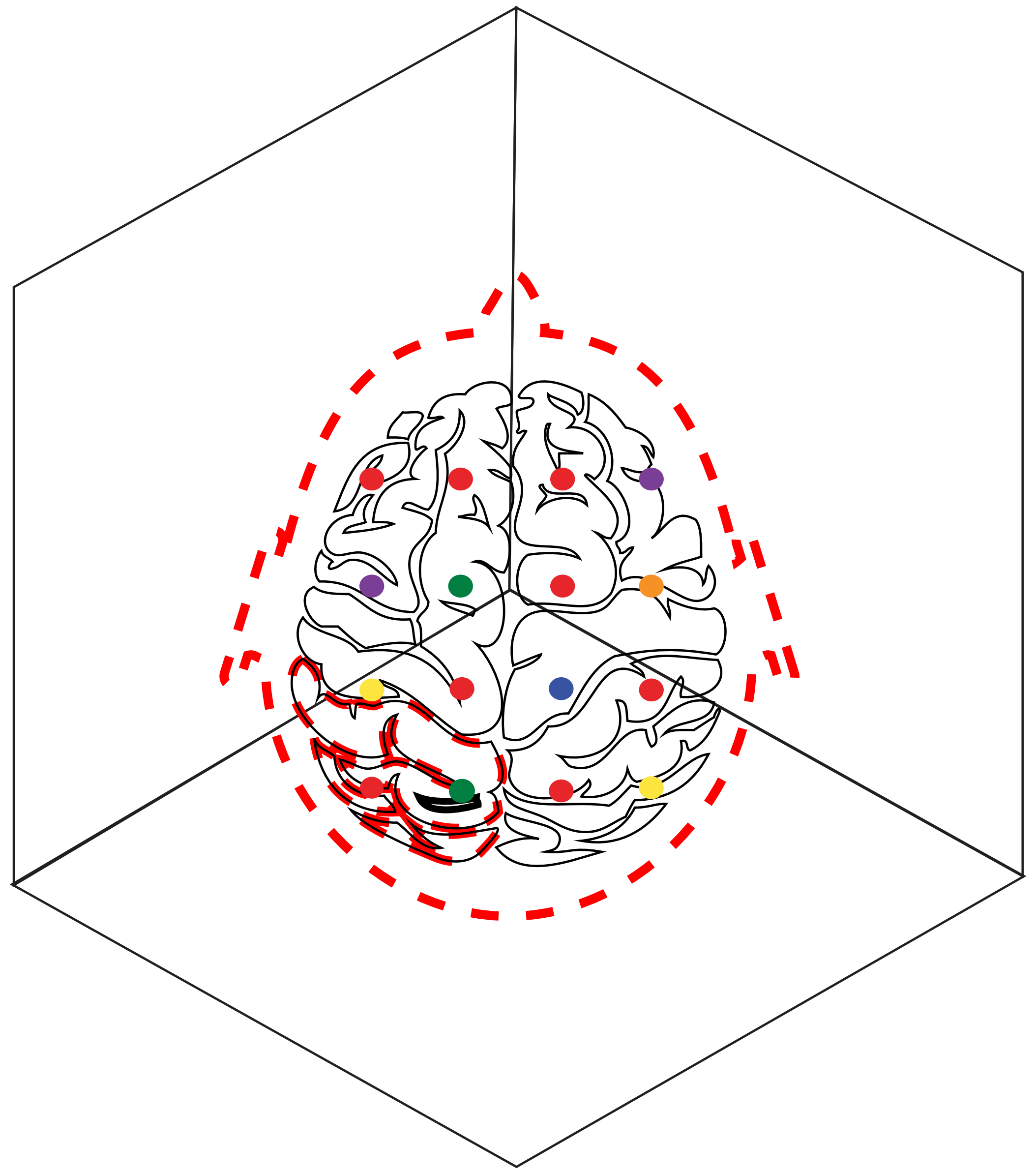
Skull stripping



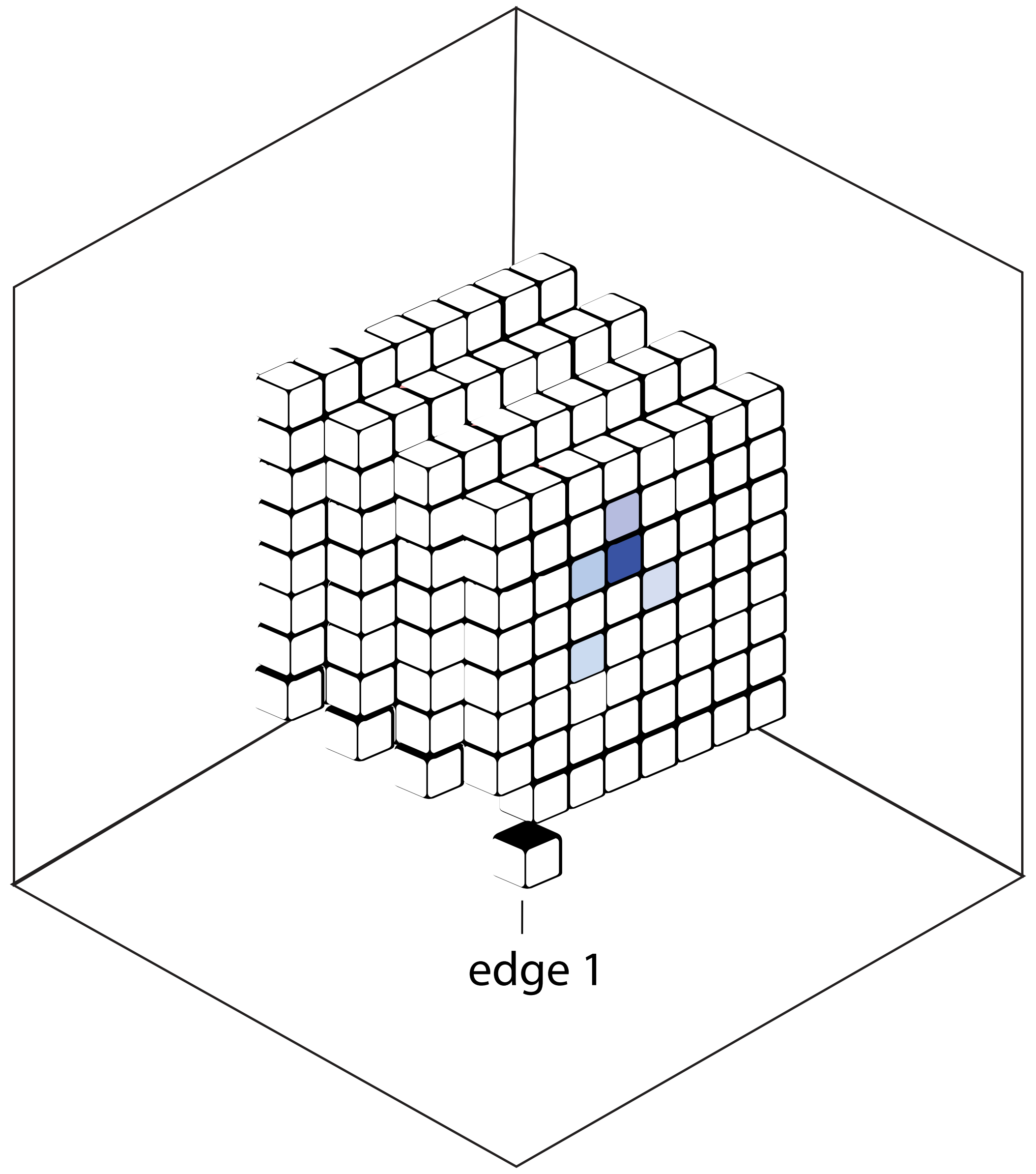
Registering to a template



Voxel wise parcellation

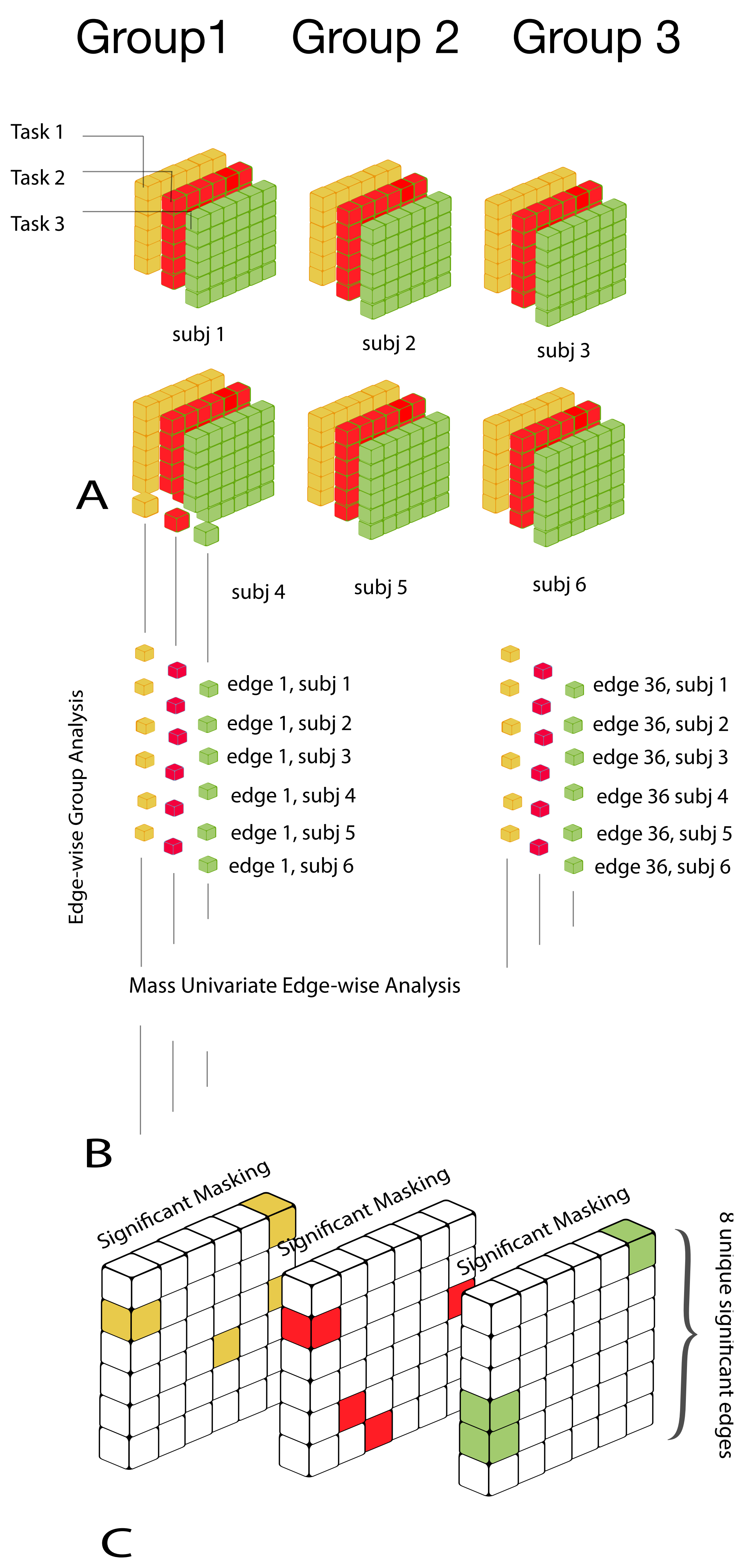


ROI-based parcellation

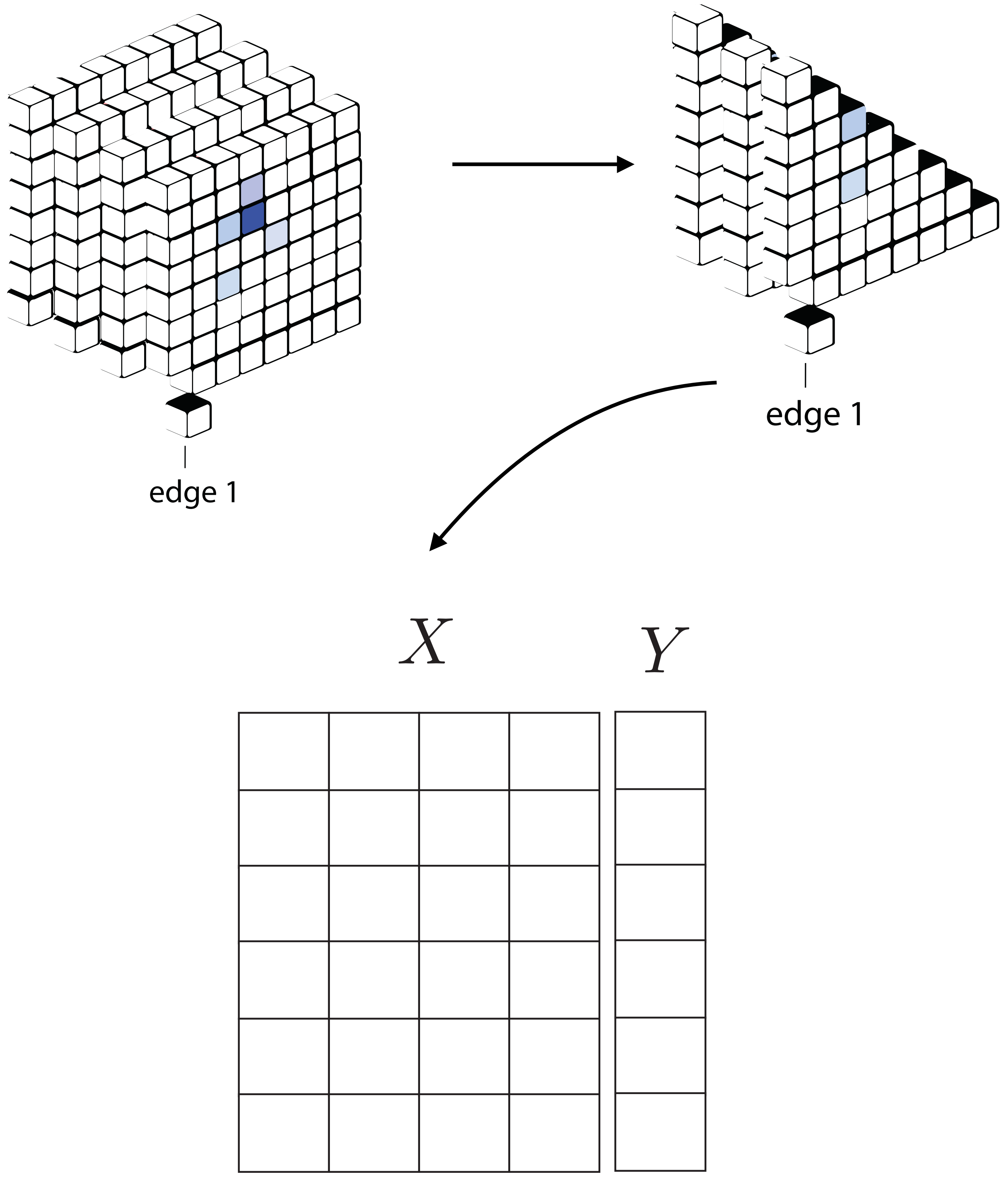


Functional connectomes

Explanatory Analysis



Predictive Modeling

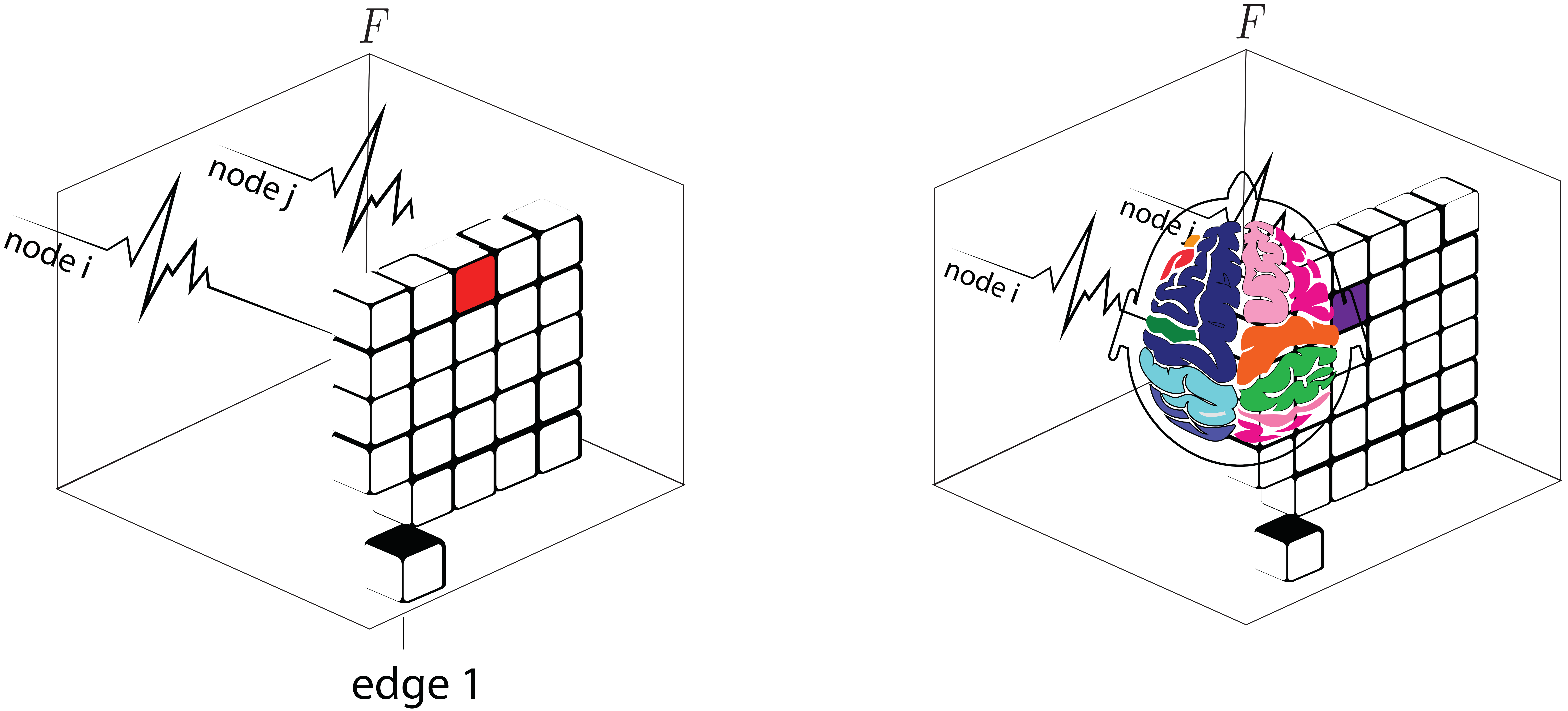


$$Y = X_1\beta_1 + X_2\beta_2 + .. + X_N\beta_N + \beta_0$$

Functional Connectivity

Widely used in neuroscience to understand the functional organization of the brain.

1. What are connectomes
2. How to make functional connectivity
3. Applications in neuroscience

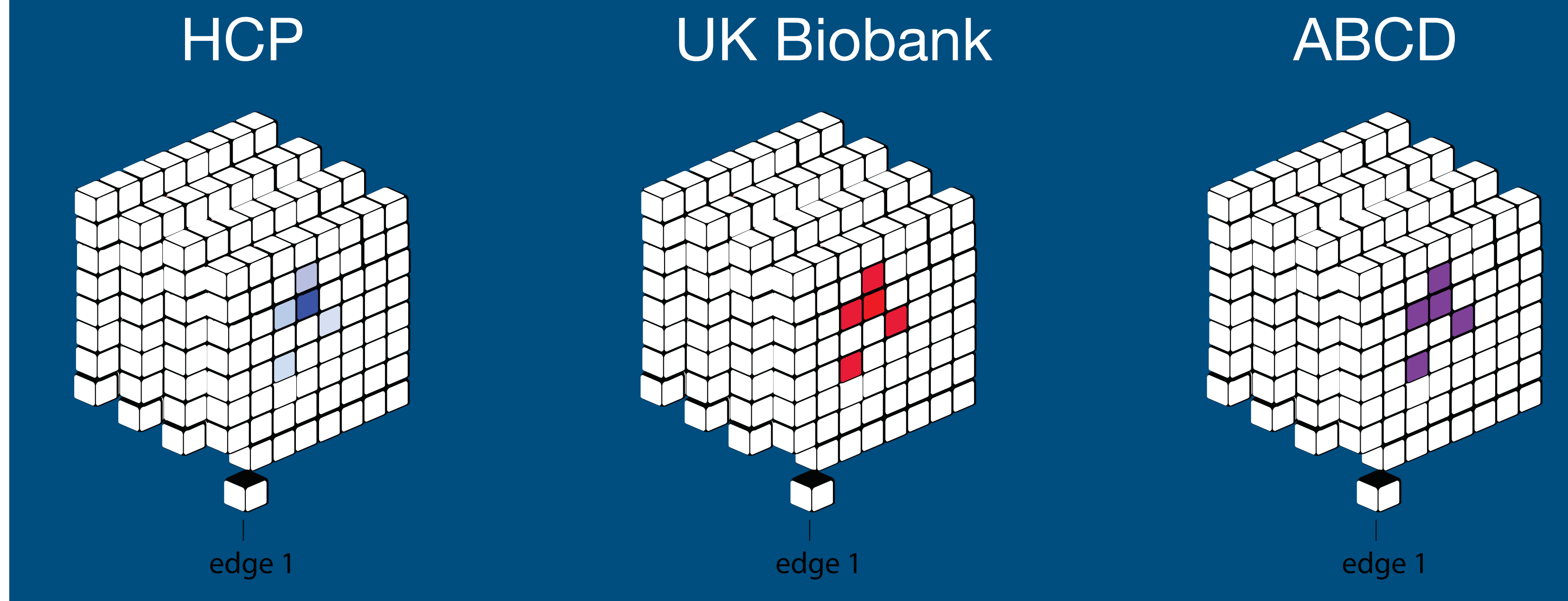


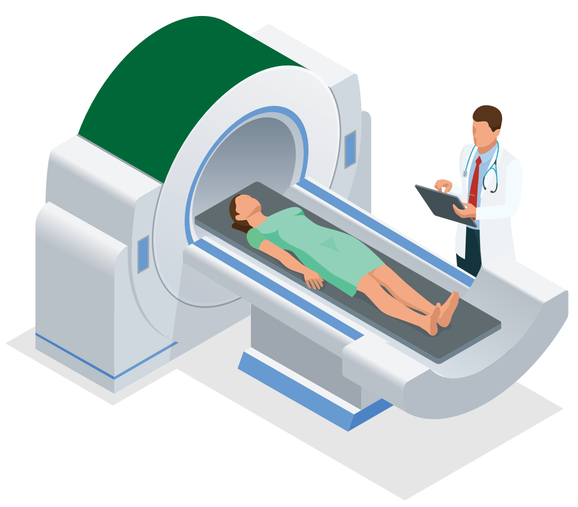
- The need for an atlas to create a connectome hinders comparisons across studies.
- Different atlases divide the brain into different regions of varying size and topology.
- Thus connectomes created from different atlases are not directly comparable.

Real-world challenges

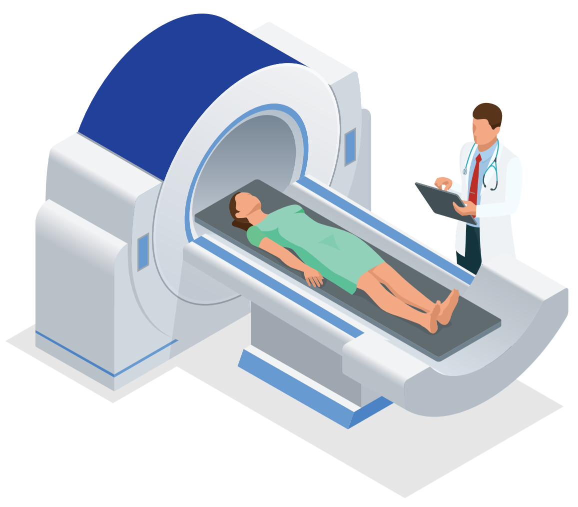
Different studies have different standards and limitations

1. Generalizability
2. Storage concerns
3. Privacy concerns

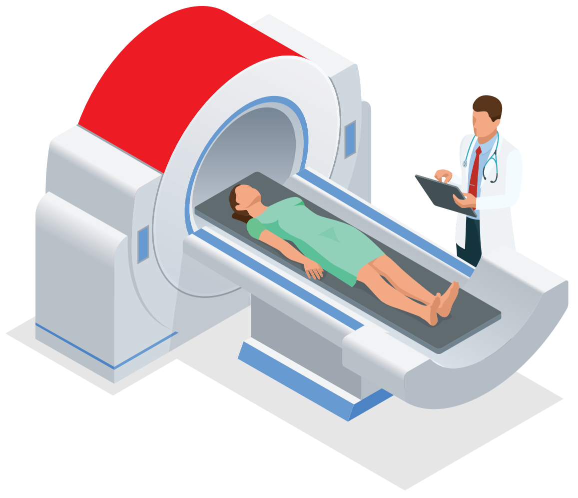




- **Generalizability**:
  - Currently, no solutions exist to extend previous results to a connectome generated from a different atlas.
  - This prevents these datasets from being combined without reprocessing data.

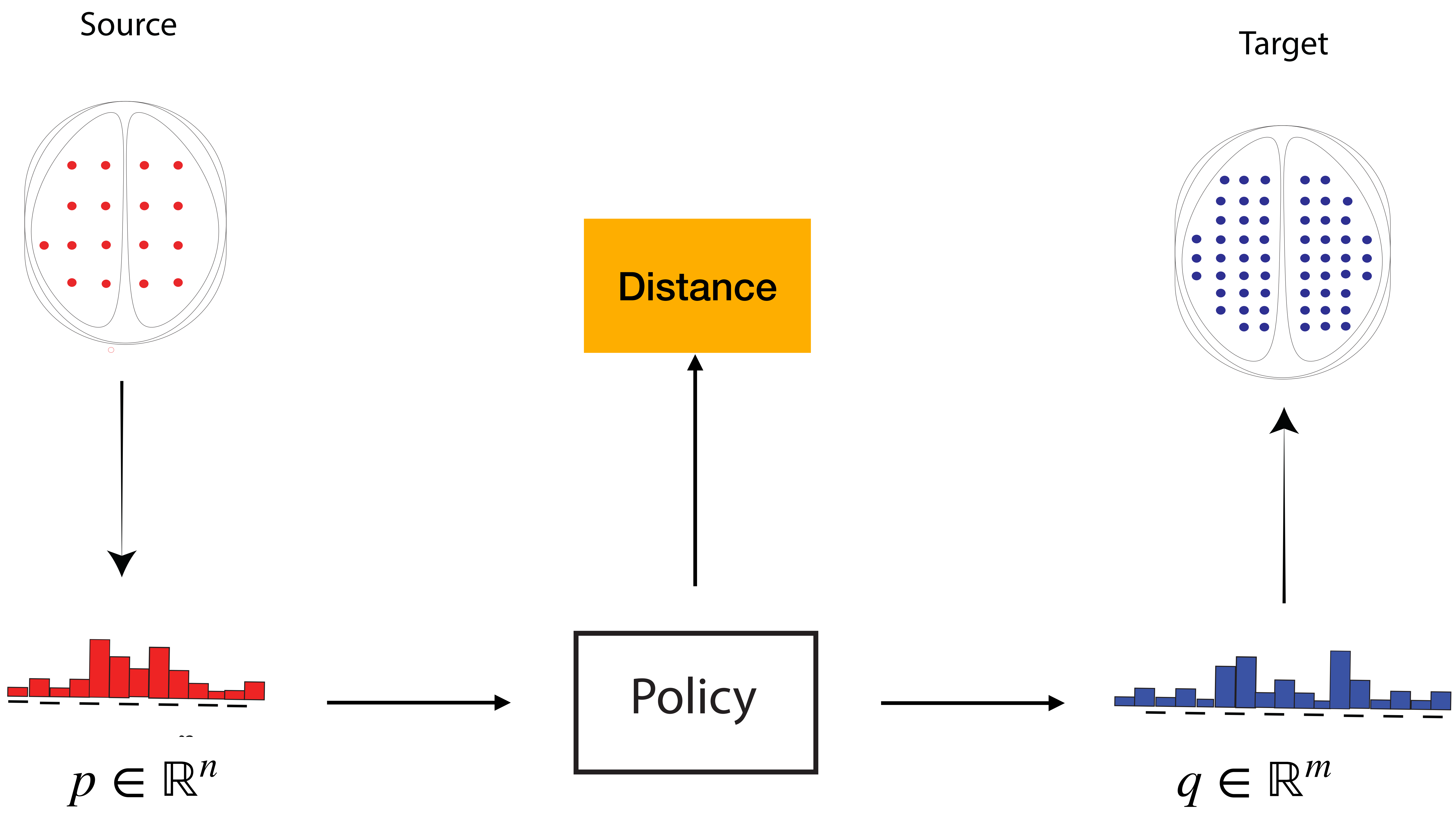


- **Storage and time complexity**:
  - Smaller labs might not have the resources to store and reprocess these data from scratch.



- **Privacy concerns**:
  - Due to privacy some datasets are only released as fully processed connectomes.
  - Critically, in this case, it is not possible to go to the data to create connectomes from another atlas.

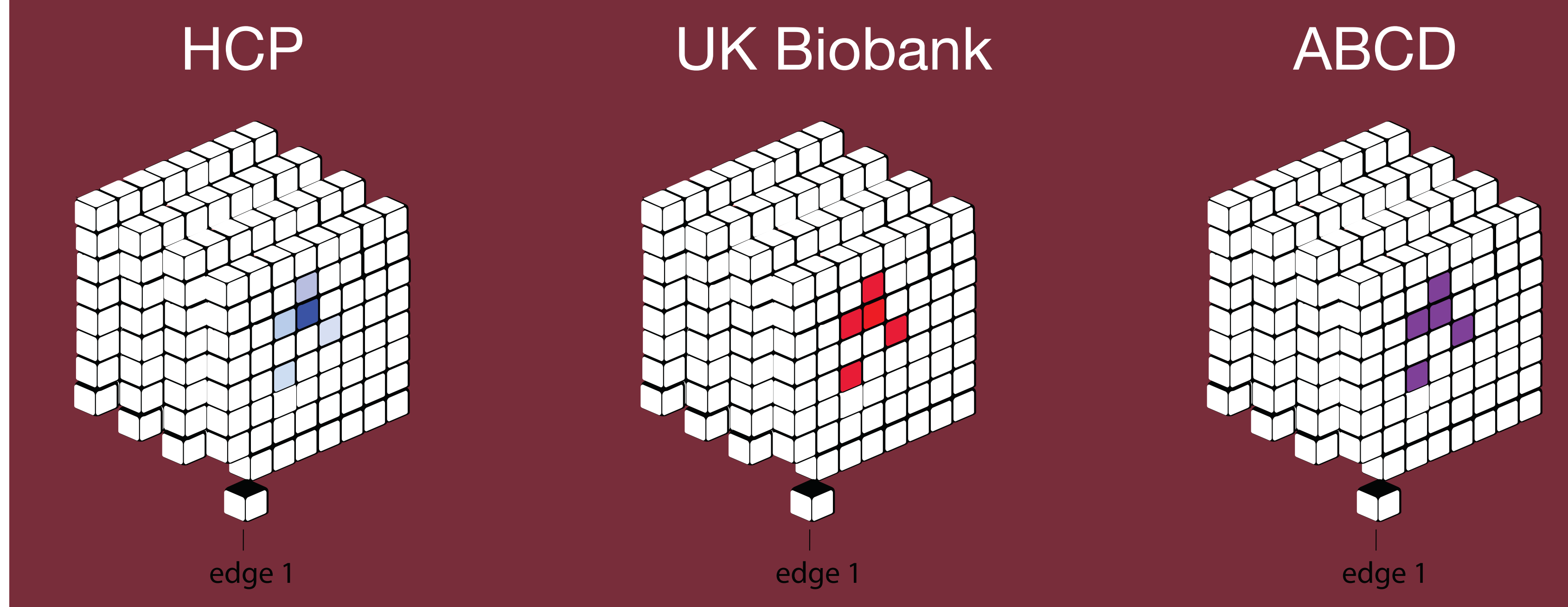
A moment-to-moment transportation method



Our solution: dataset harmonization

Estimating connectomes in a missing form

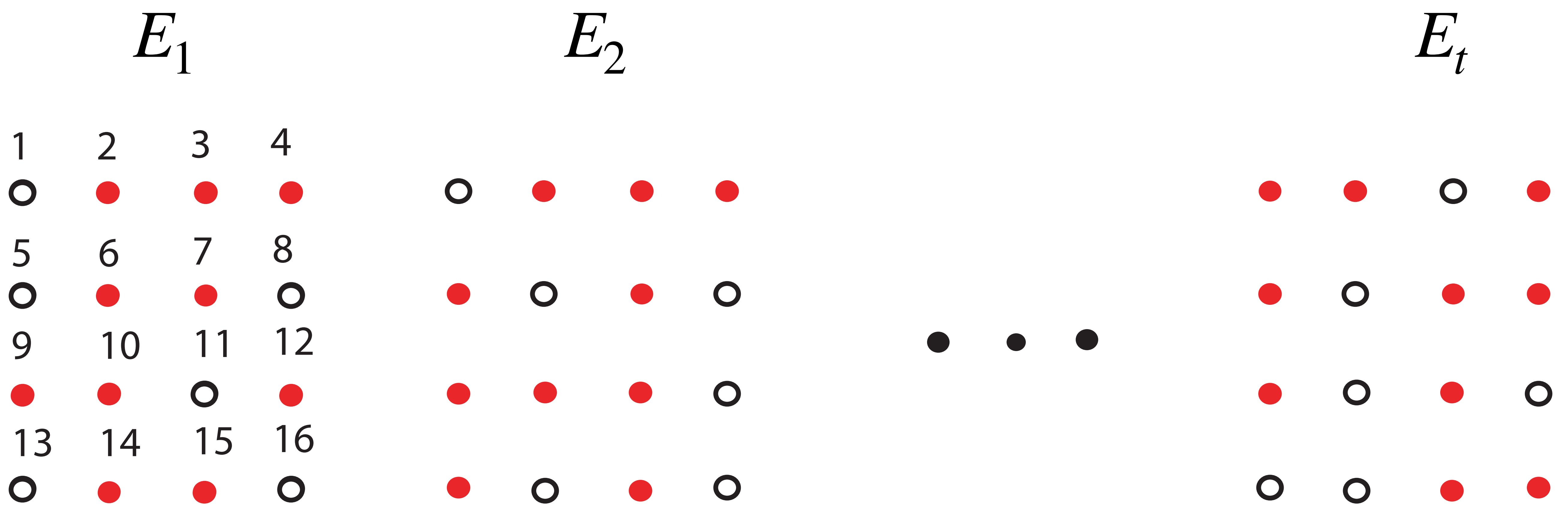
1. Time series-based approach
2. Transforming distribution of ROIs across atlases



Our solution: dataset harmonization

Estimating connectomes in a missing form

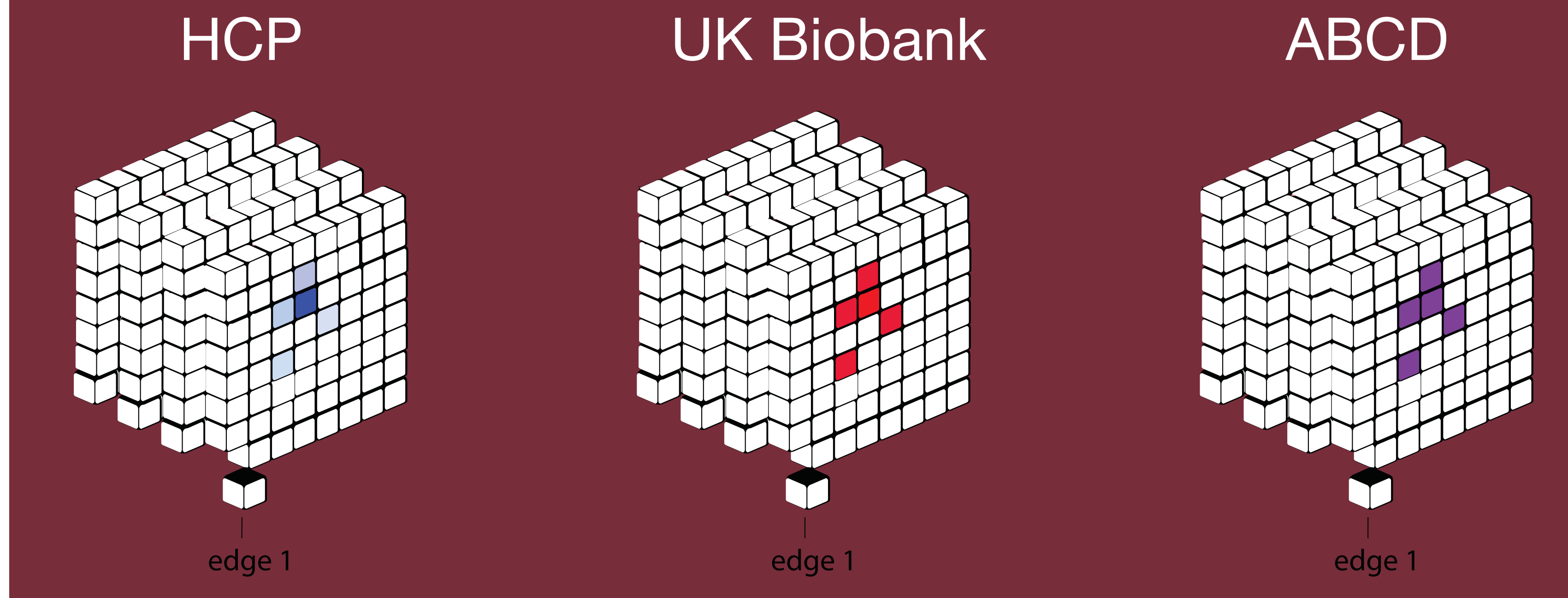
1. Time series-based approach
2. Transforming distribution of ROIs across atlases



$$p_1^{N_A}(1-p_1)^{N_I}\; p_2^{N_A}(1-p_2)^{N_I} \;\cdots\; p_{16}^{N_A}(1-p_{16})^{N_I}$$

$$q_1^{N_A}(1-q_1)^{N_I}\; q_2^{N_A}(1-q_2)^{N_I} \;\cdots\; q_{16}^{N_A}(1-q_{16})^{N_I}$$

$$\frac{P(\text{observation}\,|\,p)}{P(\text{observation}\,|\,q)} = \frac{p_1^{N_A^1}(1-p_1)^{N_I^1}\; p_2^{N_A^1}(1-p_2)^{N_I^1} \cdots p_{16}^{N_A^1}(1-p_{16})^{N_I^1}}{q_1^{N_A^1}(1-q_1)^{N_I^1}\; q_2^{N_A^1}(1-q_2)^{N_I^1} \cdots q_{16}^{N_A^1}(1-q_{16})^{N_I^1}}$$



normalized relative likelihood $= \left(\frac{p_1^{N_A^1} p_2^{N_A^2} \ldots (1-p_{16})^{N_I}}{q_1^{N_A^1} q_2^{N_A^2} \ldots (1-q_{16})^{N_I}}\right)^{\frac{1}{N}}$

$$= \frac{1}{N} \log\left(\frac{p_1^{N_A} p_2^{N_A} \ldots (1-p_{16})^{N_I}}{q_1^{N_A} q_2^{N_A} \ldots (1-q_{16})^{N_I}}\right)$$

$$= \frac{1}{N} \log p_1^{N_A^1} + \quad \frac{1}{N} \log(1-p_1)^{N_I^1} \quad \ldots \quad -\frac{1}{N} \log q_{16}^{N_A} \quad -\frac{1}{N} \log(1-q_{16})^{N_I}$$

$$= \frac{N_A^1}{N} \log p_1 + \quad \frac{N_I^1}{N} \log(1-p_1) \quad \ldots \quad -\frac{N_A}{N} \log q_{16} \quad -\frac{N_I}{N} \log(1-q_{16})$$

$$= p_1 \log p_1 \; + (1-p_1)\log(1-p_1) \cdots -q_{16} \log q_{16}$$

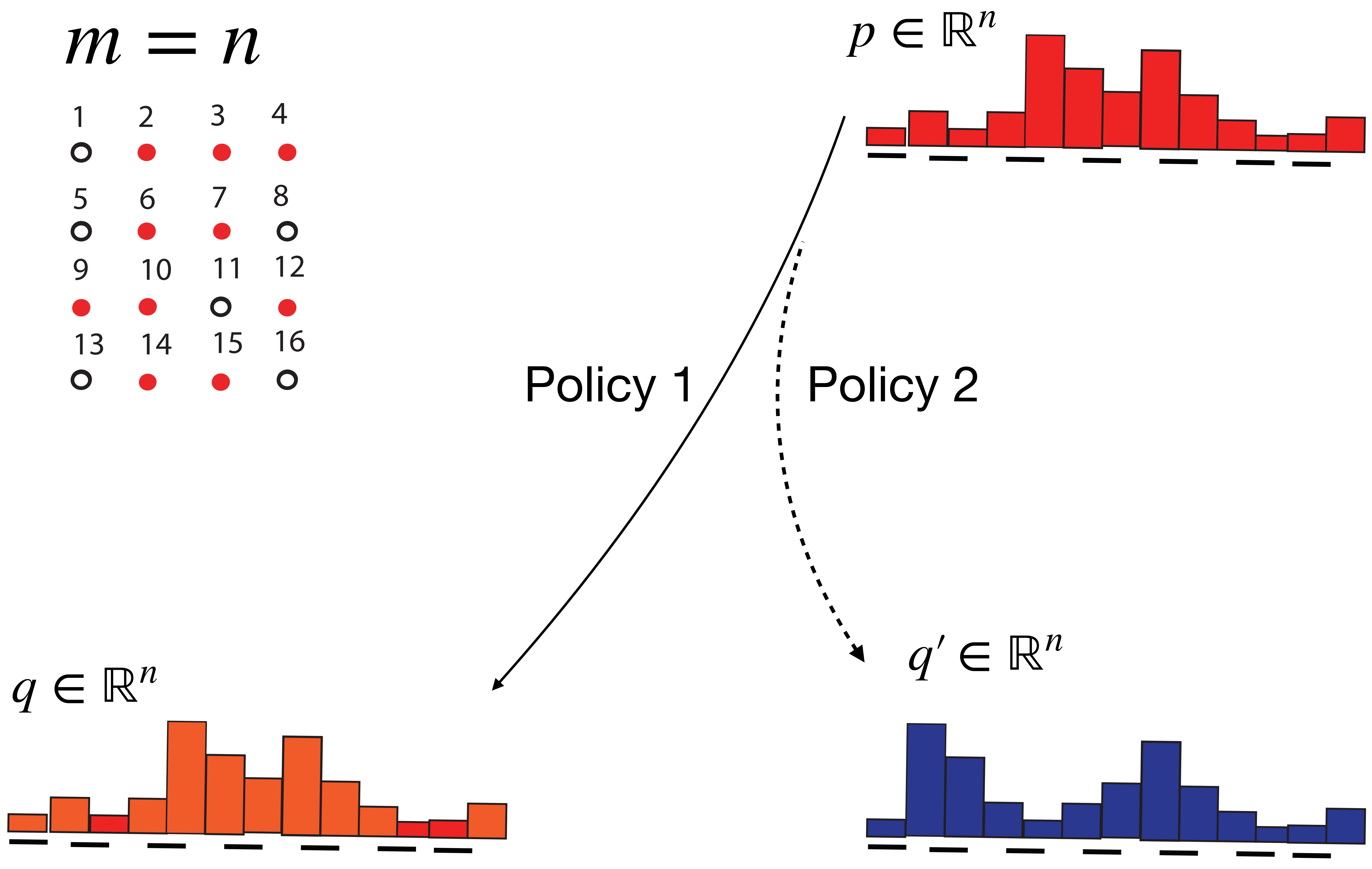
$$-(1-q_{16})\log(1-q_{16})$$

$$= D_{KL}(p\,||\,q) = \sum_{x \in \mathcal{X}} p(x) \log\left(\frac{p(x)}{q(x)}\right)$$

Kullback–Leibler divergence

Measures exactly the same thing

1. Log properties, product to addition, division to subtraction
2. How likely $q(x)$ would generate samples from $p(x)$

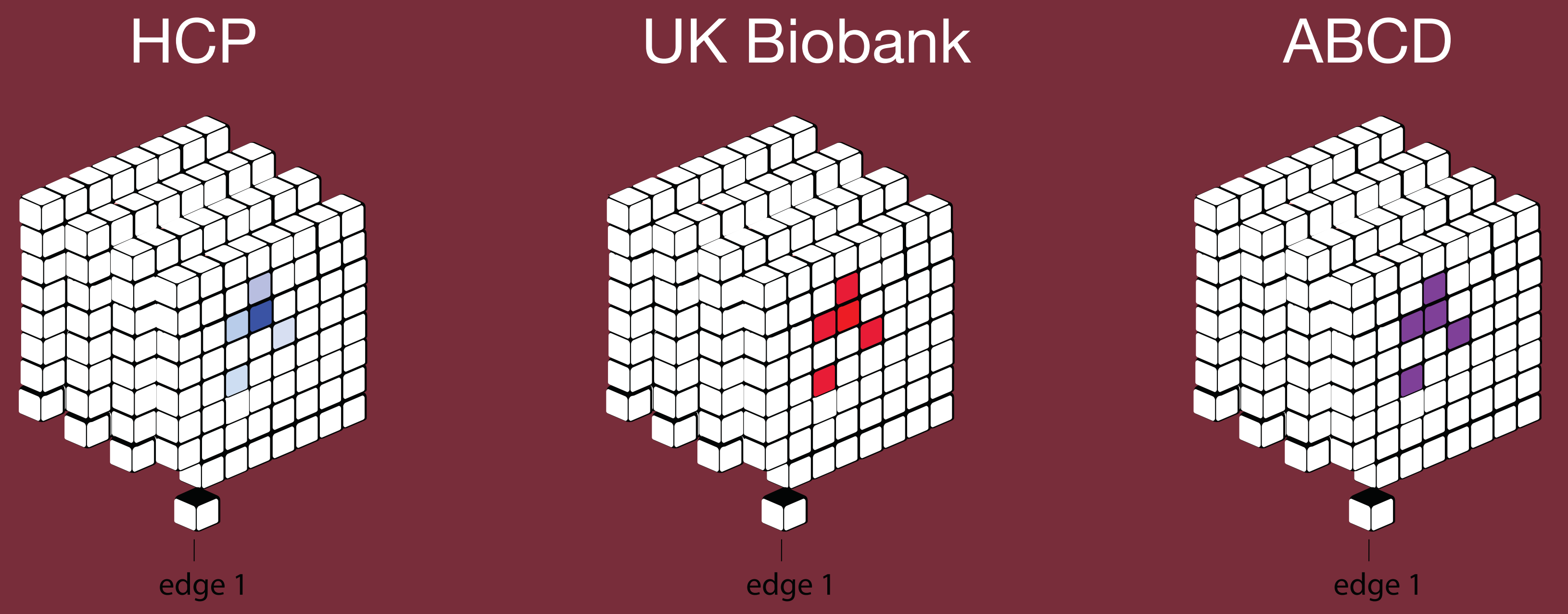


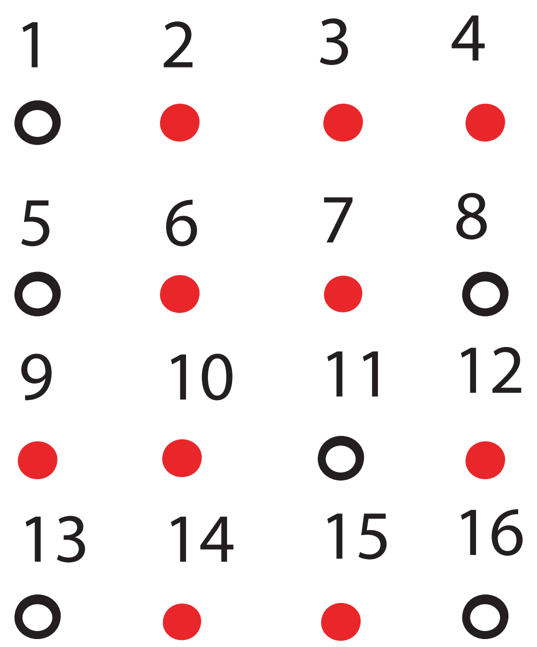
$$KL(p\,||\,q) < KL(p\,||\,q')$$

Our solution: dataset harmonization

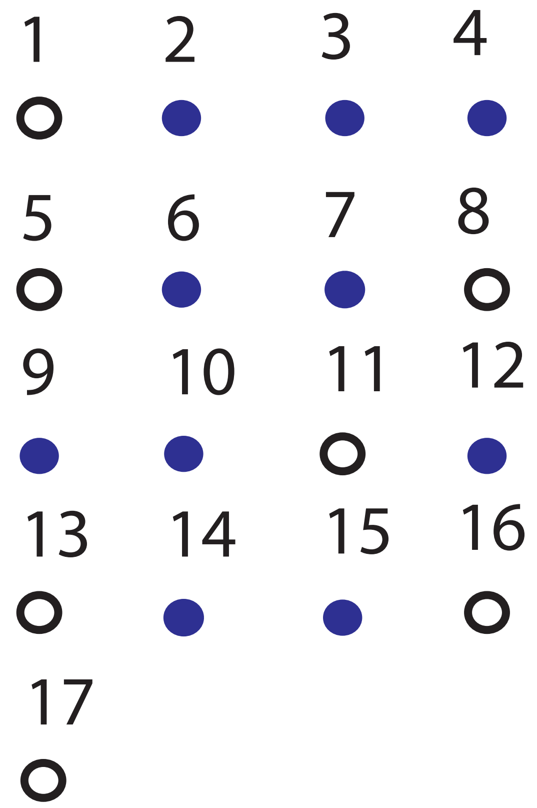
Estimating connectomes in a missing form

1. Time series-based approach
2. Transforming distribution of ROIs across atlases





How about when $m \neq n$



$$\frac{P(\text{observation}\mid p)}{P(\text{observation}\mid q)} = \frac{p_1^{N_A}(1-p_1)^{N_I}\, p_2^{N_A}(1-p_2)^{N_I} \cdots p_{16}^{N_A}(1-p_{16})^{N_I}\times 0}{q_1^{N_A}(1-q_1)^{N_I}\, q_2^{N_A}(1-q_2)^{N_I}\cdots q_{17}^{N_A}(1-q_{17})^{N_I}}$$

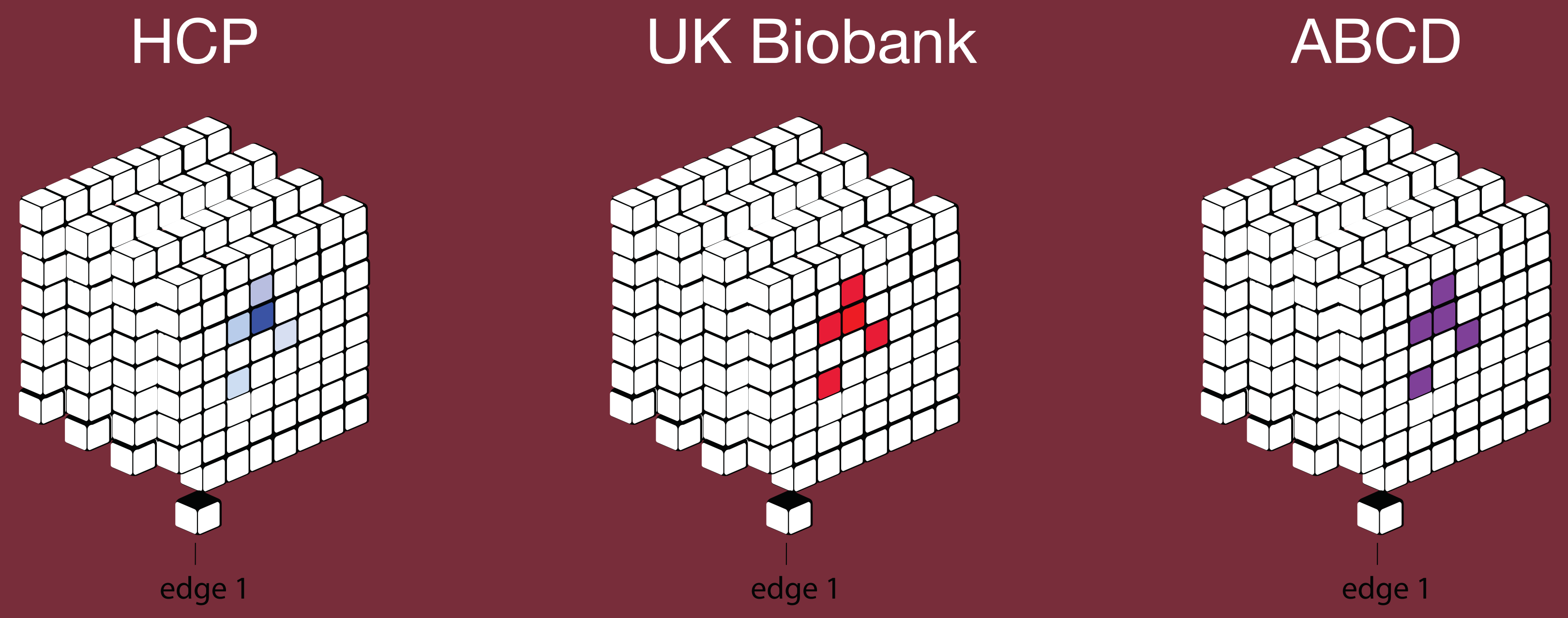
$$= 0$$

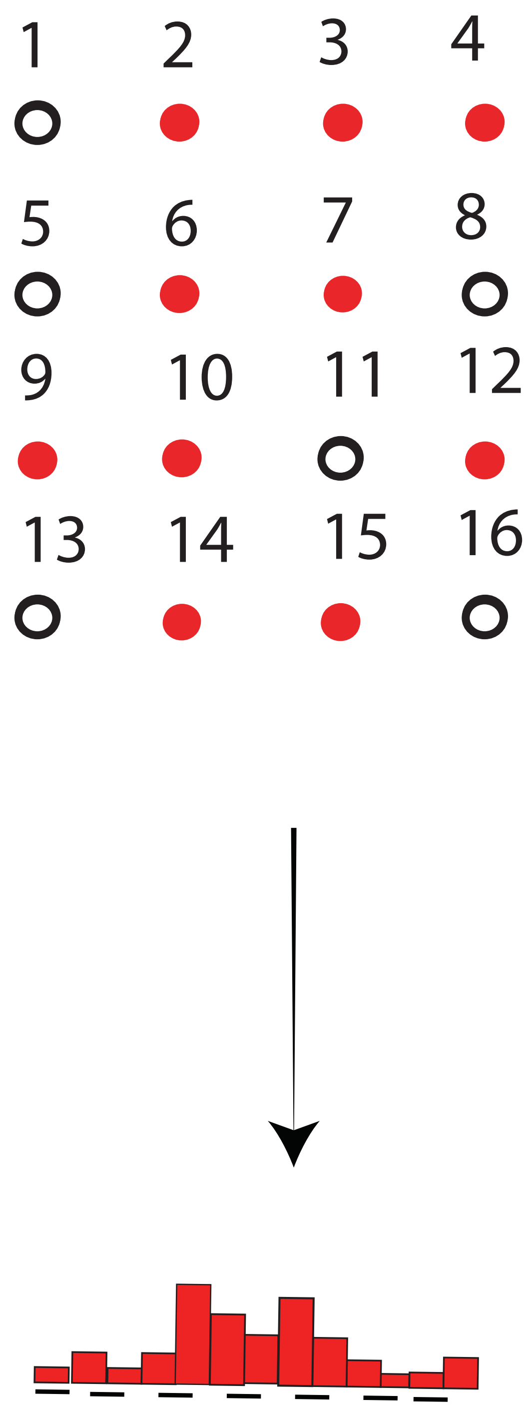
KL divergence fails in this scenario.

Our solution: dataset harmonization

Estimating connectomes in a missing form

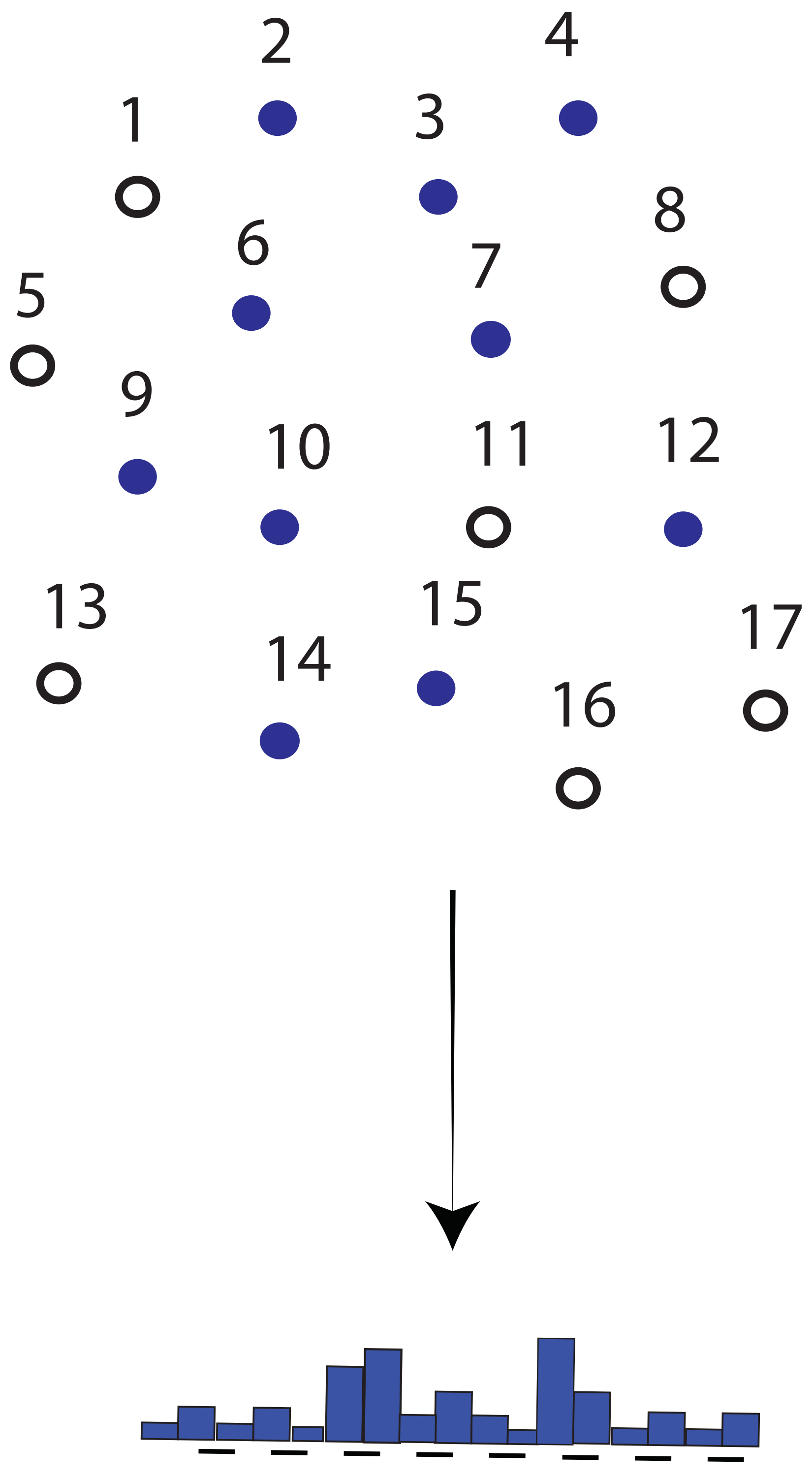
1. Time series-based approach
2. Transforming distribution of ROIs across atlases





$\mu_N \in \mathbb{R}^{n_N}$

$$\mathscr{M}_1 \neq \mathscr{M}_2$$



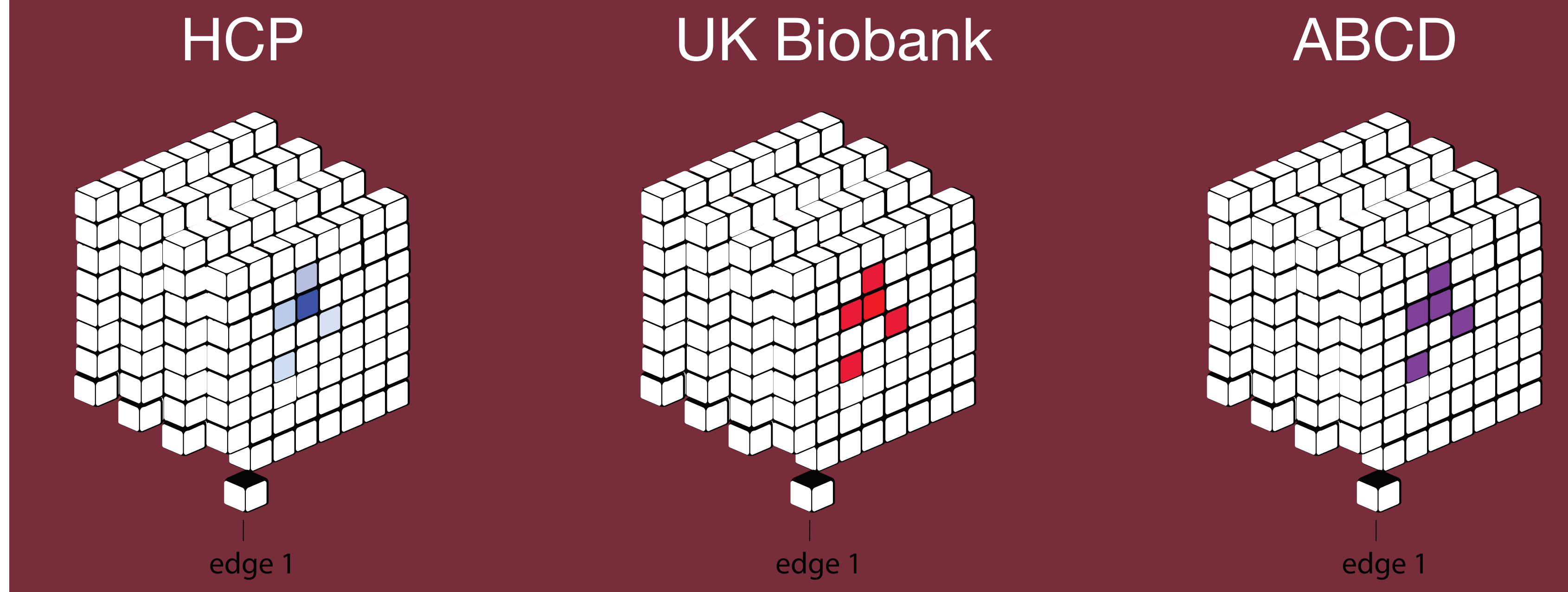
$\nu \in \mathbb{R}^m$

1. How about when the two distributions are defined in completely different spaces?
2. Optimal Transport captures both geometry and inconsistency of dimensions between p and q.

Our solution: dataset harmonization

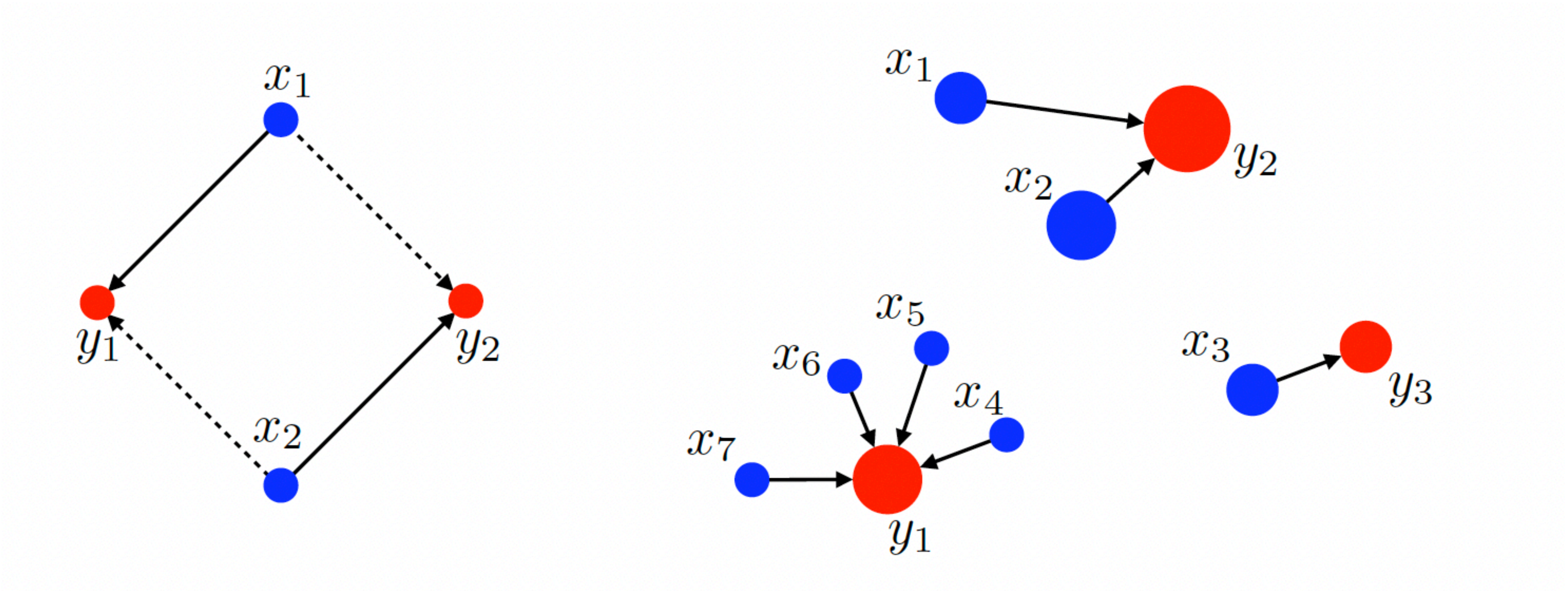
Estimating connectomes in a missing form

1. Time series-based approach
2. Transforming distribution of ROIs across atlases



Monge [1781]

Optimal Transport



A mapping between locations x and y

$$T : \{x_1, \ldots, x_n\} \rightarrow \{y_1, \ldots, y_n\}$$

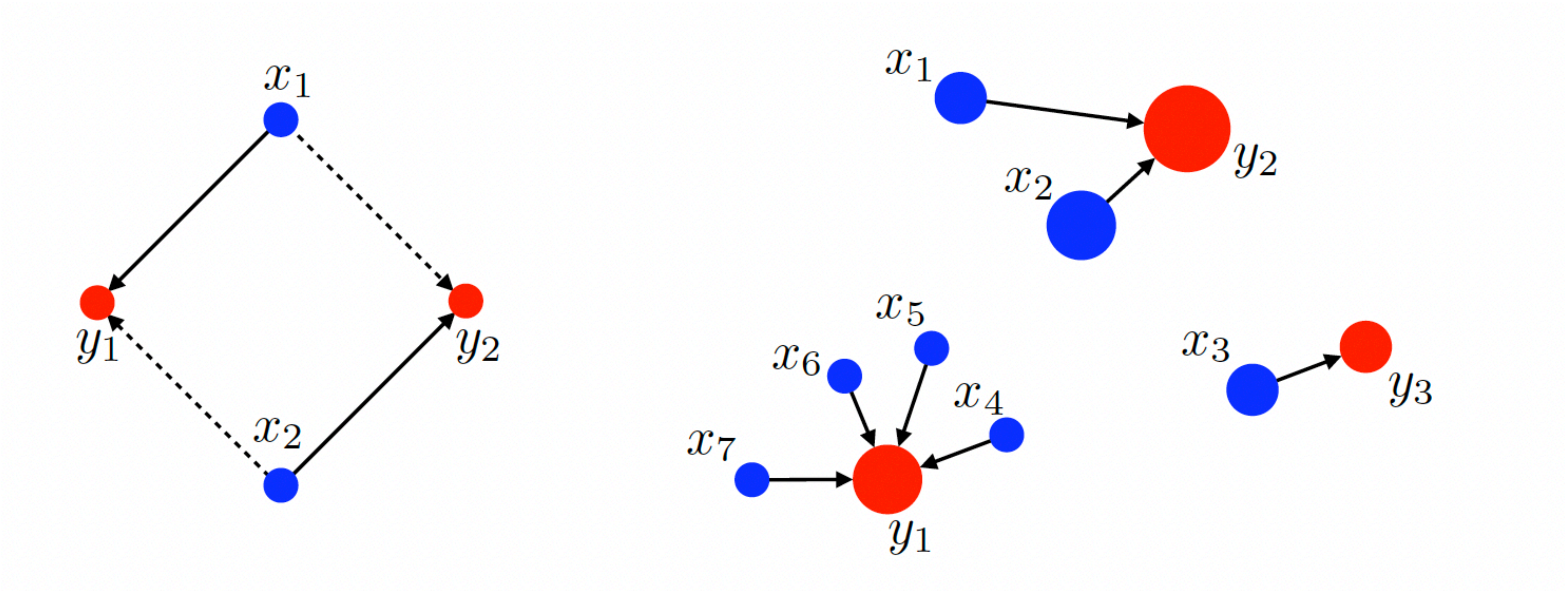
must verify

$$b_j = \sum_{i:T(x_i)=y_j} a_i$$

The only criterion here is to make sure we transfer all mass into some location y_j

Monge [1781]

Optimal Transport



This map should minimize some transportation cost, which is parameterized by a cost function C

$$\min_T \left\{ \sum_i C(x_i, T(x_i)) : T_\sharp \alpha = \beta \right\},$$

Optimal Transport

Monge [1781]

Kantorovich [1942]

Admissible Couplings

Kantorovich Relaxation [1942]

$$\mathbf{U}(\mathbf{a},\mathbf{b}) \overset{\text{def.}}{=} \left\{\mathbf{P} \in \mathbb{R}_+^{n\times m} \;:\; \mathbf{P}\mathbb{1}_m = \mathbf{a} \quad \text{and} \quad \mathbf{P}^{\mathrm{T}}\mathbb{1}_n = \mathbf{b}\right\},$$

$$\mathbf{P}\mathbb{1}_m = \left(\sum_j \mathbf{P}_{i,j}\right)_i \in \mathbb{R}^n \quad \text{and} \quad \mathbf{P}^{\mathrm{T}}\mathbb{1}_n = \left(\sum_i \mathbf{P}_{i,j}\right)_j \in \mathbb{R}^m.$$

Optimal Transport

Monge [1781]

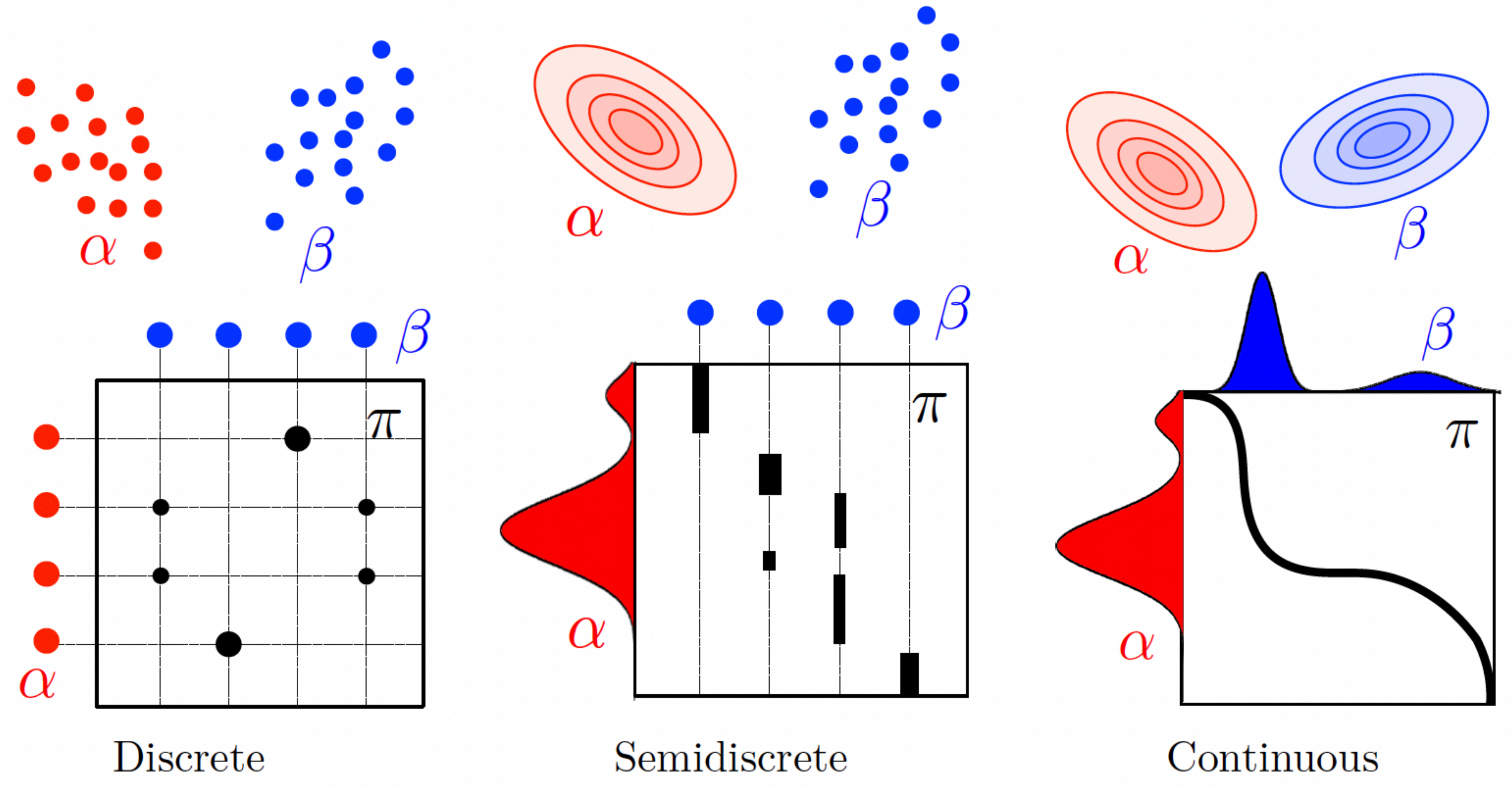
Kantorovich [1942]

$$A = \begin{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ (1 & 1 & \cdots & 1) \\ & & \vdots & \\ (1 & 1 & \cdots & 1) \end{pmatrix}_1 & \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ (1 & 1 & \cdots & 1) \\ & & \cdots & \end{pmatrix}_2 & \cdots & \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \cdots(1 & 1 & \cdots & 1) \\ \cdots(1 & 1 & \cdots & 1) \end{pmatrix}_n \end{pmatrix}$$

$$L_c(\mu_t, \nu_t) = \min_T C^T T \text{ s.t, } A\underline{T} = \begin{bmatrix} \mu_t \\ \nu_t \end{bmatrix}.$$

Kantorovich's optimal transport problem now reads

$$L_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) \overset{\text{def.}}{=} \min_{\mathbf{P} \in \mathbf{U}(\mathbf{a},\mathbf{b})} \langle \mathbf{C}, \mathbf{P} \rangle \overset{\text{def.}}{=} \sum_{i,j} \mathbf{C}_{i,j} \mathbf{P}_{i,j}.$$



Kantorovich Relaxation is symmetric

$$P \in U(a, b) \Leftrightarrow P^T \in U(b, a)$$

Optimal Transport

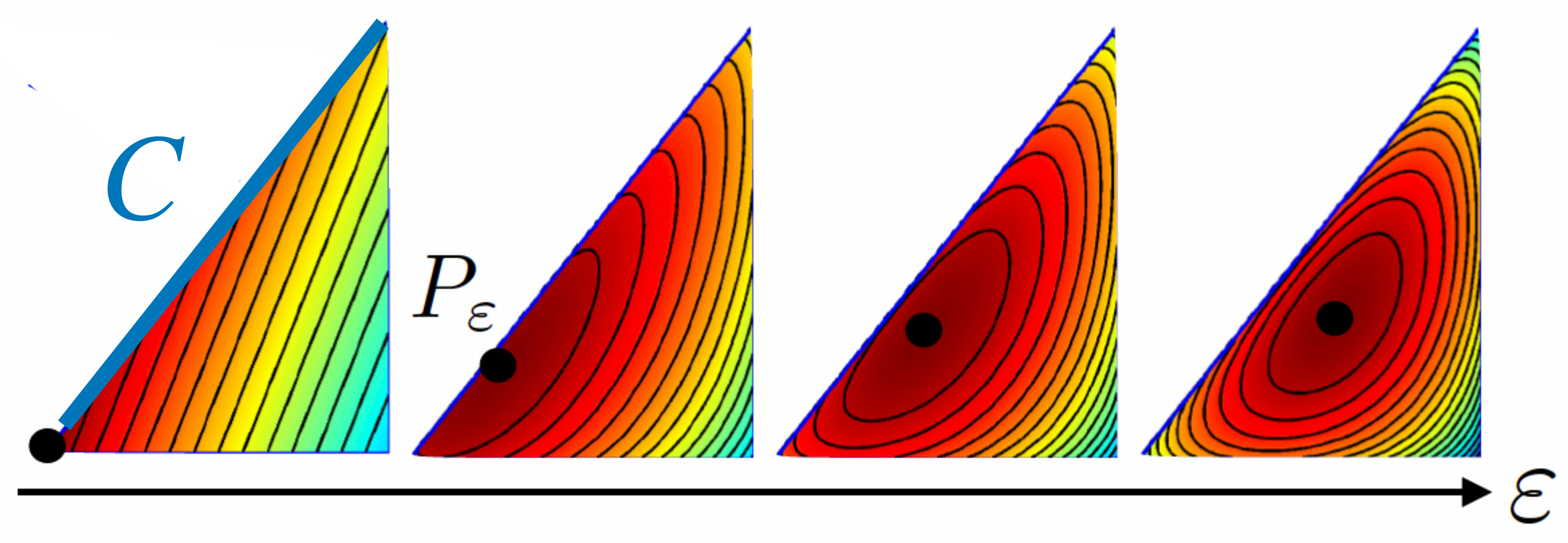
Monge [1781]

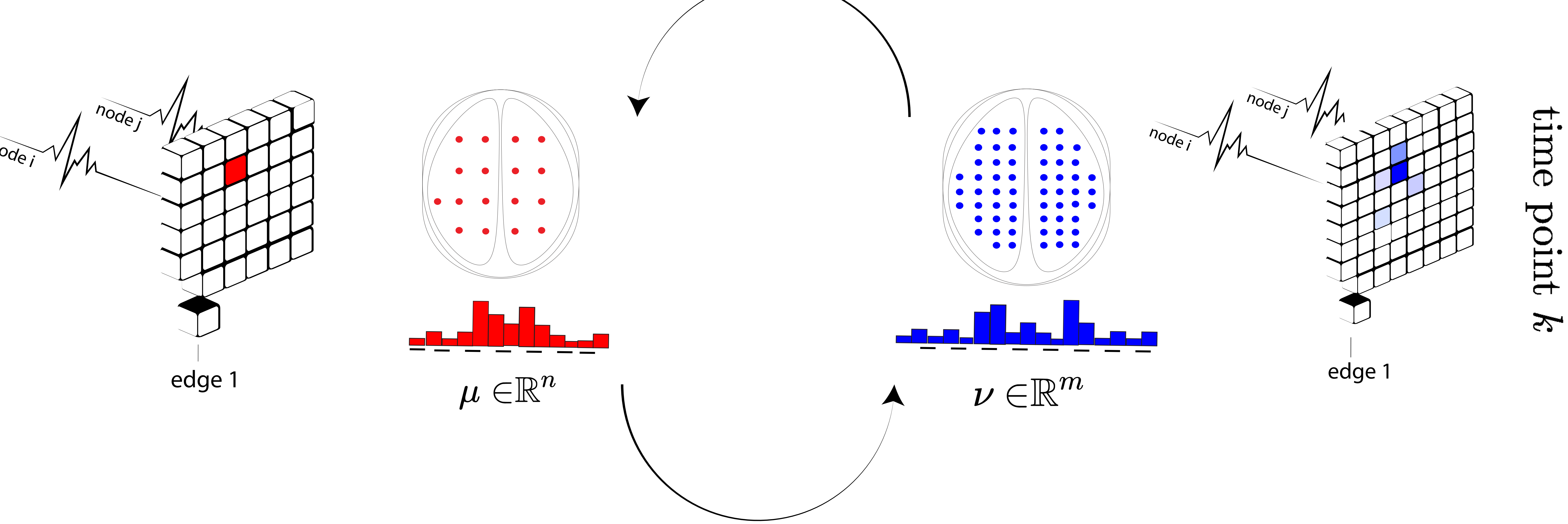
Hitchcock [1941]

Kantorovich [1942]

Entropy regularization: An approximation solution

$$\mathrm{L}_{\mathbf{C}}^{\varepsilon}(\mathbf{a}, \mathbf{b}) \stackrel{\text{def.}}{=} \min_{\mathbf{P} \in \mathbf{U}(\mathbf{a}, \mathbf{b})} \langle \mathbf{P}, \mathbf{C} \rangle - \varepsilon \mathbf{H}(\mathbf{P}).$$





$$L_c(\mu_t, \nu_t) = \min_T C^T T - \epsilon H(T) \text{ s.t, } A\underline{T} = \begin{bmatrix} \mu_t \\ \nu_t \end{bmatrix}.$$

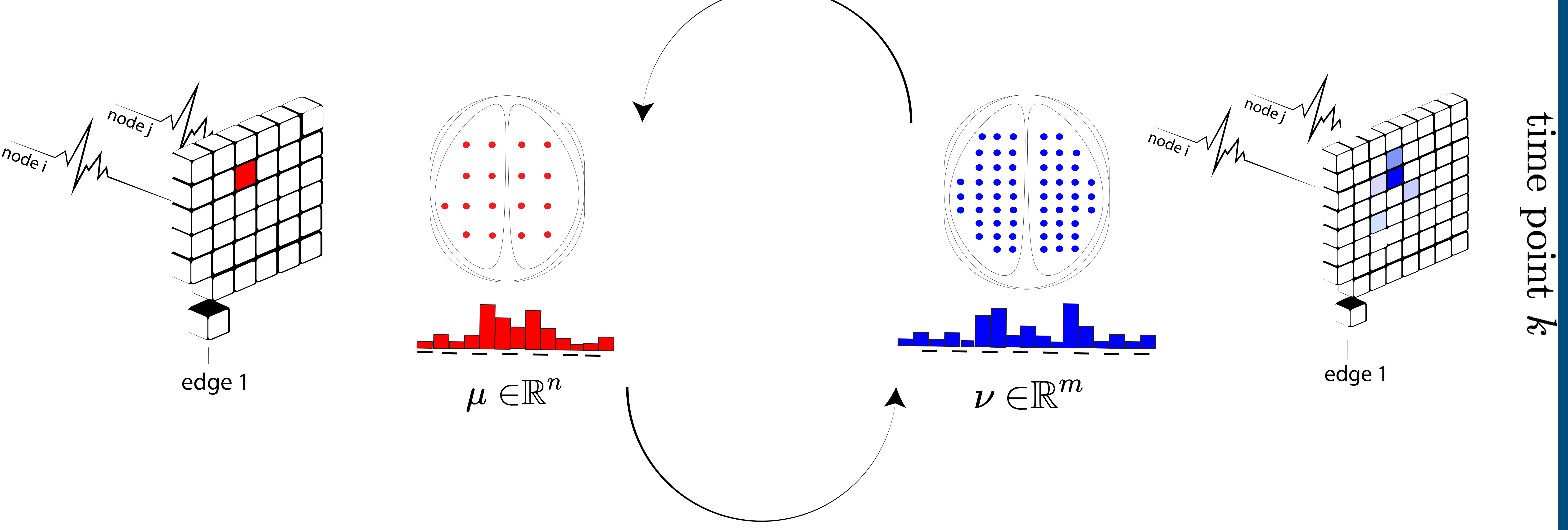
$$A = \begin{matrix} m \\ \\ n \end{matrix} \begin{pmatrix} \overset{1}{\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}} & \overset{2}{\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}} & \dots & \overset{n}{\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}} \\ (1 \; 1 \; \dots \; 1) & (1 \; 1 \; \dots \; 1) & \dots & (1 \; 1 \; \dots \; 1) \\ \vdots & & & \\ (1 \; 1 \; \dots \; 1) & & \dots & \dots (1 \; 1 \; \dots \; 1) \end{pmatrix}$$

$$C = \begin{pmatrix} C_{1,1} & .. & C_{1,m} \\ .. & & .. \\ C_{n,1} & .. & C_{n,m} \end{pmatrix} \in \mathbb{R}^{n \times m} \qquad C_{i,j} = \text{Functional distance}$$

Cross Atlas Remapping via Optimal Transport (CAROT)

A data-driven method to measure the distance and find a policy to transform connectomes

1. Translating each time frame to a vector
2. Cost matrix
3. Loss function



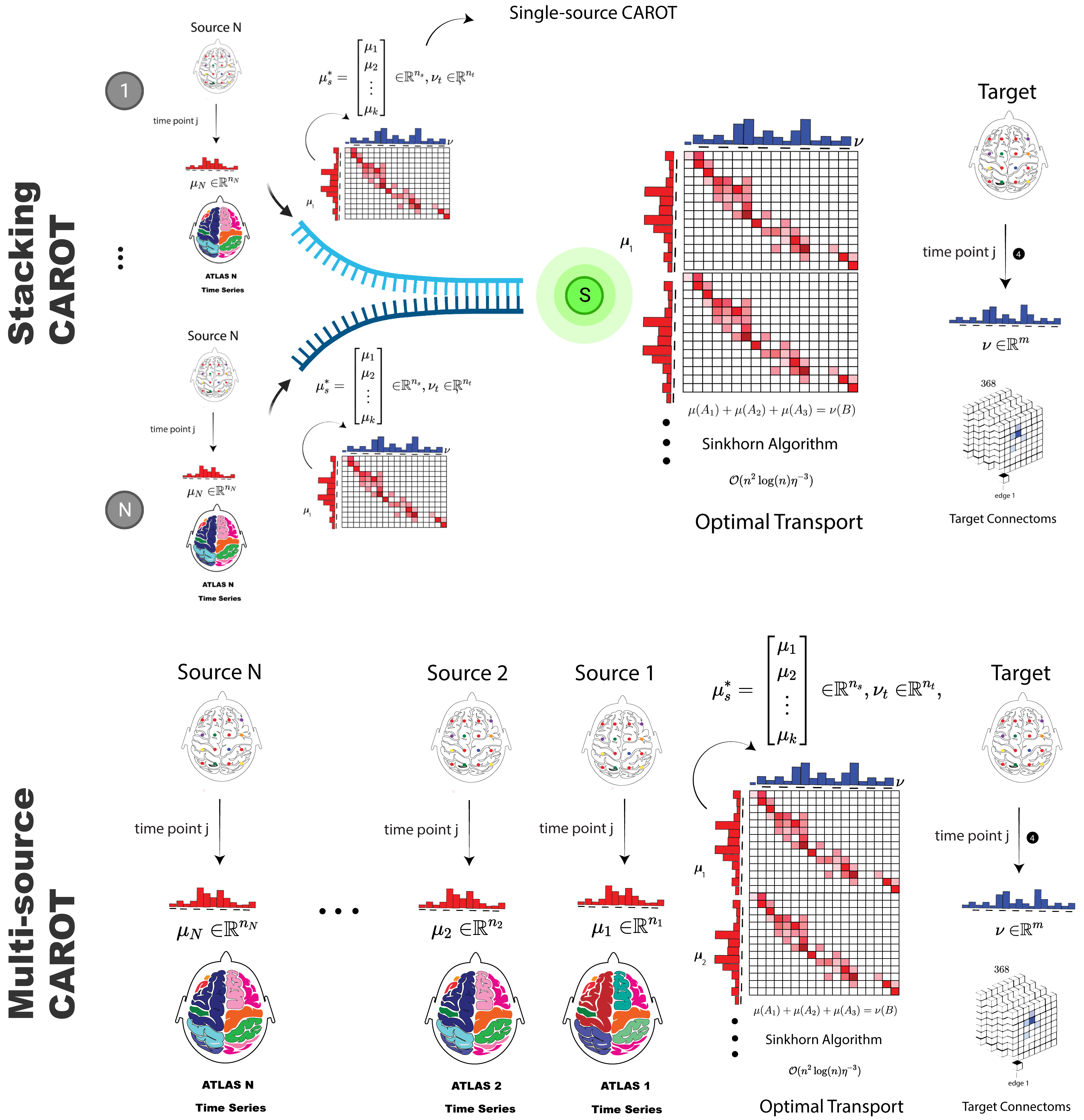
test data point $\nu = \mu T$

What if multiple parcellations for each individual are available?

Cross Atlas Remapping via Optimal Transport (CAROT)

Test data point available in the source atlas

1. Applying the trained policies T
2. Some of large scale projects release data in multiple atlases
3. A need for an advanced version

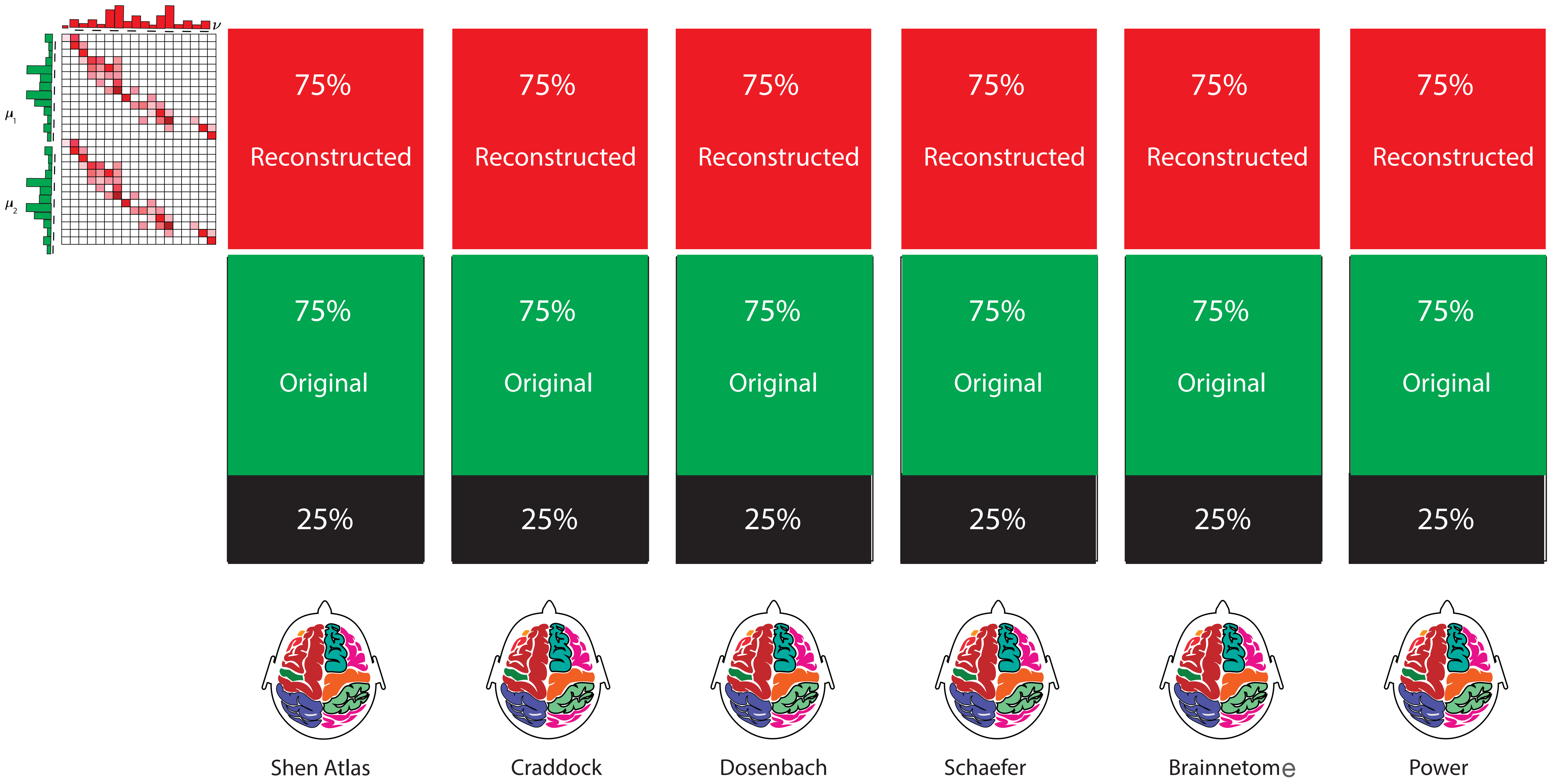


Stacking multiple optimal transport

An advanced version when multiple parcellations are available

1. Incorporating multiple time series
2. Bigger cost matrix
3. Bigger policy

The Human Connectome project is used for training mappings, intrinsic analysis, and for some downstream analysis

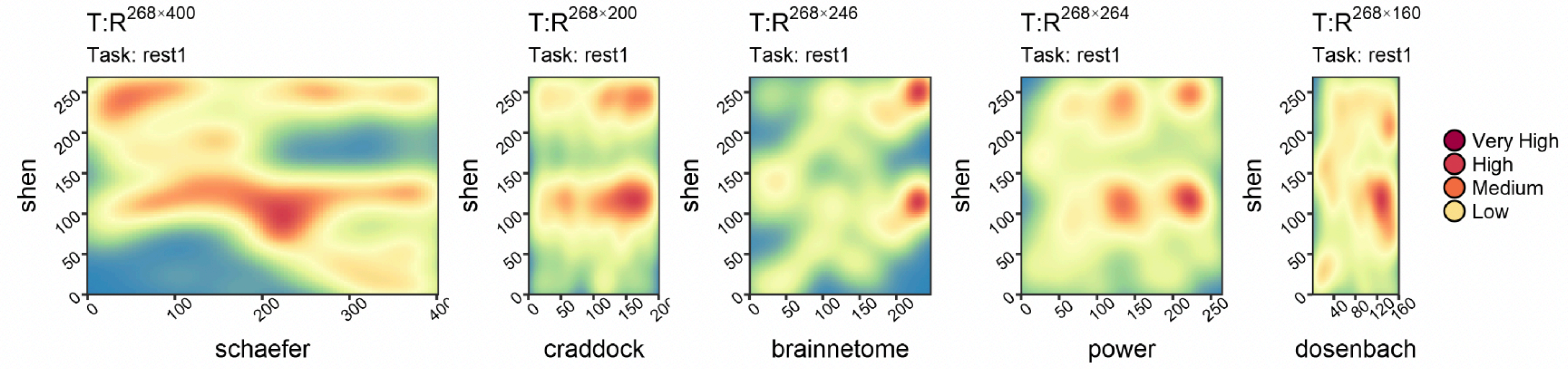


$$\binom{6}{2} = 15 \text{ transportation policies}$$

Experiments:

Human connectome projects

1. Train-test split
2. 25% for policy training
3. 75% for testing
4. 10 fold CV

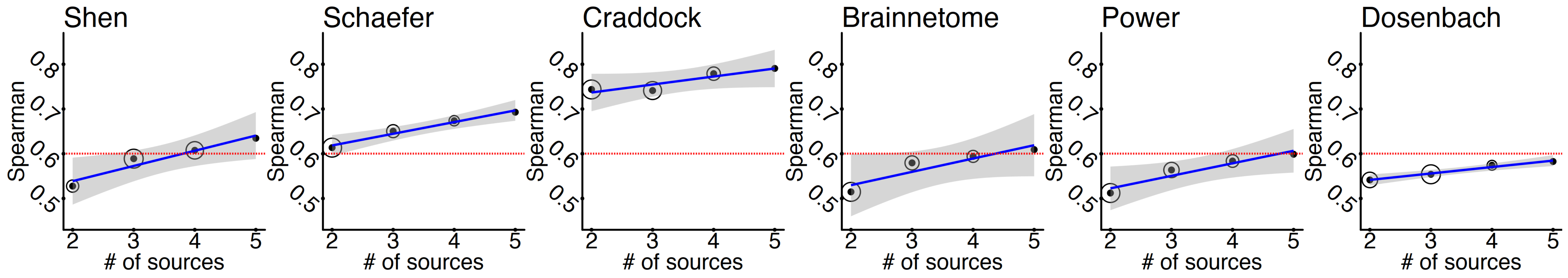


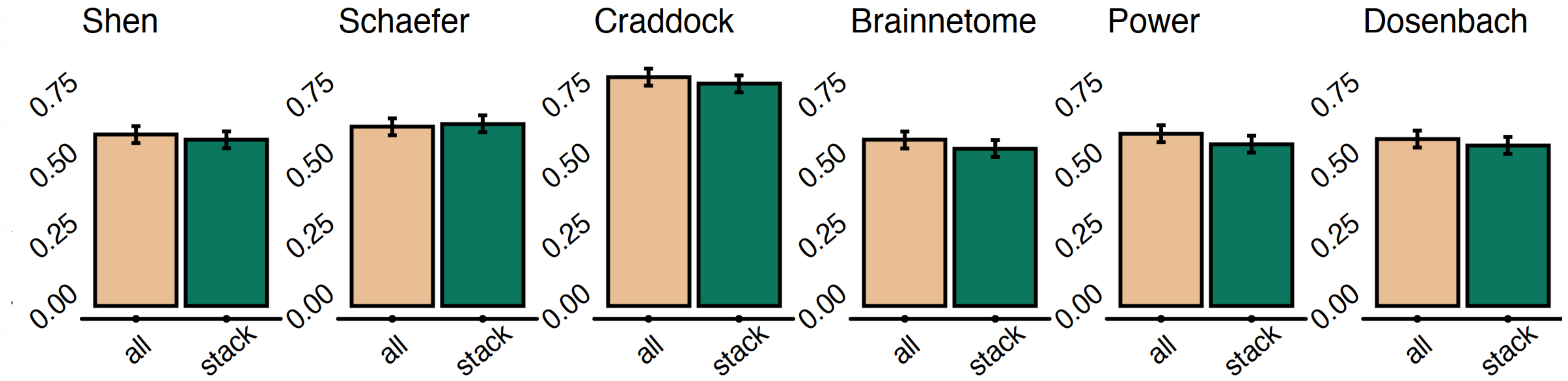
- Red spots represent higher transportation and blow spots belong to zero transportation.
- You can see that some spots are more intense than others indicating higher transformation between regions.
- This emphasizes some of the structural differences between atlases:
  - The horizontal line between Schaefer and Shen is belonging to areas that are missing in Schaefer

Policies

How does a policy look like

1. Topological differences are clear
2. Schaefer doesn’t include some areas





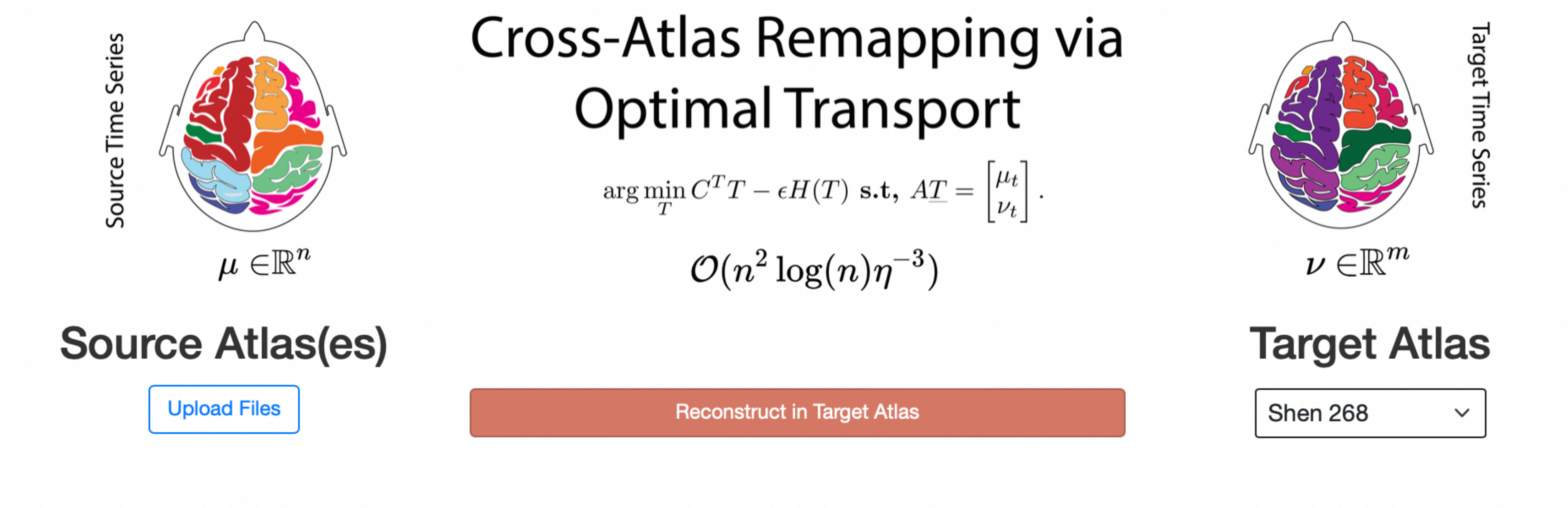
- There are differences among various runs and targets:
  - Similar atlases reproduced more similar connectomes
- We can predict behavior (e.g., fluid intelligence) and can identify individuals across different runs.
- The correlation as a function of a number of sources.

Experimental results

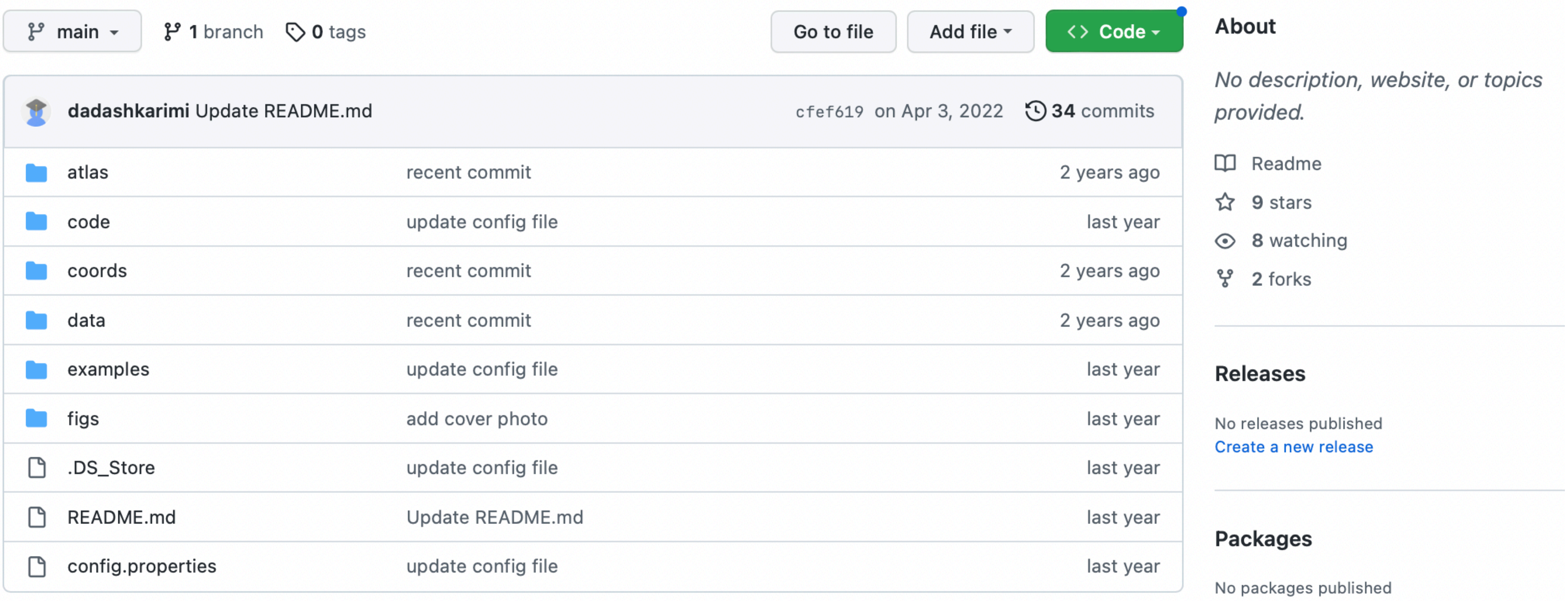
HCP dataset, resting scan connectomes

1. Intrinsic evaluation; correlation with original counterparts
2. Downstream analysis, results on predicting IQ

carotproject.com



https://github.com/dadashkarimi/carot



Software

GitHub and live demo

1. Live demo for some atlases
2. GitHub repository for all types of data

Summary

CAROT encourages open science in connectomics

1. CAROT helps overcome multiple atlas problem
2. CAROT brings good quality

- In sum, CAROT allows a connectome generated from one atlas to map to a different atlas without needing raw data.
- These reconstructed connectomes are similar to the original connectomes created from the raw data.
- Using CAROT accelerates the use of big data, and makes replication efforts easier.

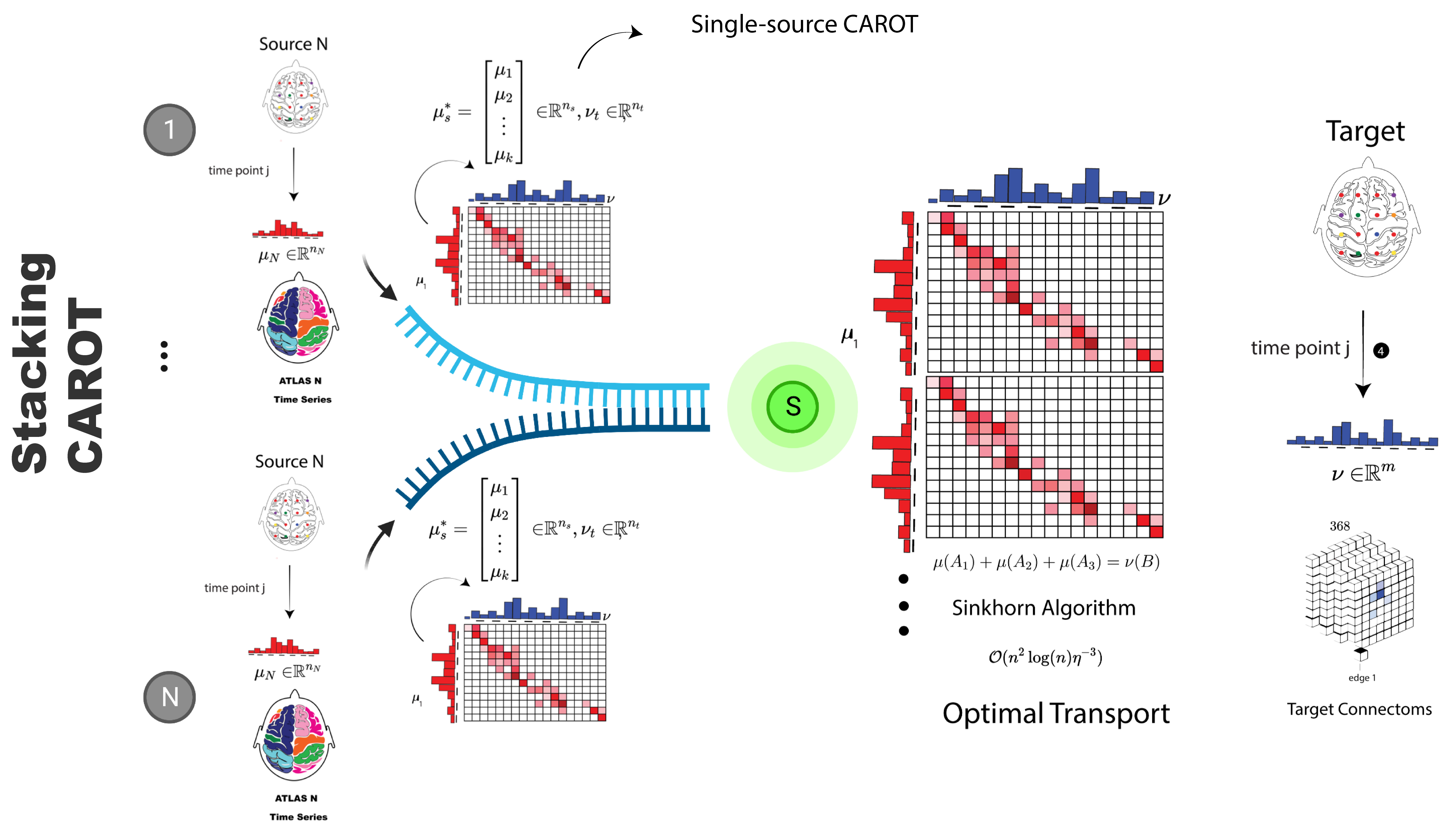
Thank you so much: MINDS lab and IID lab

- Dustin Scheinost
- Amin Karbasi
- Qinghao Liang
- Matthew Rosenblatt
- Stephanie Noble
- Raimundo Rodriguez
- Brendan Adkinson
- Huili Sun
- Jean Ye
- Maya Foster
- Chris Camp
- Michael Farruggia
- Link Tejavibulya
- Wei Dai
- Raina Vin
- AJ Simon
- Camille Duan
- Rongtao Jiang
- Angeliki Pollatou
- Shuo Chen
- Marie Mccusker

Stacking multiple optimal transport

An advanced version when multiple parcellations are available

1. Incorporating multiple time series
2. Bigger cost matrix
3. Bigger policy



$$L_c(\mu^*_t, \nu^*_t) = \min_T C^T T - \epsilon H(T) \text{ s.t, } A\underline{T} = \begin{bmatrix}\mu^*_t \\ \nu^*_t\end{bmatrix}.$$

$$\mu_s^* = \begin{bmatrix}\mu_1 \\ \mu_2 \\ \vdots \\ \mu_k\end{bmatrix} \in \mathbb{R}^{n_s}, \nu_t \in \mathbb{R}^{n_t}, C^* = \begin{pmatrix} C_{1,1} & .. & C_{1,m} \\ .. & .. & .. \\ C_{n_s,1} & .. & C_{n,m}\end{pmatrix} \in \mathbb{R}^{n_s \times m}$$