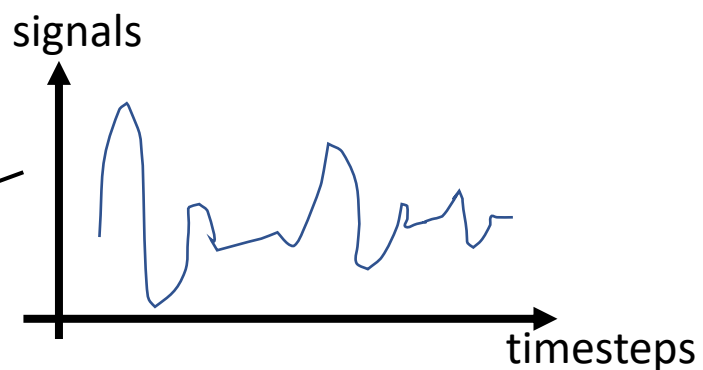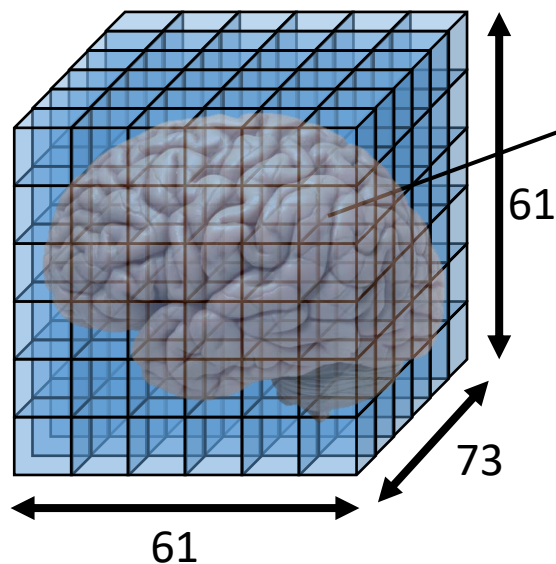# Brain Graph Project

11/18/2022

Weikang Qiu

Dataset Overview

- 1770 brain voxel-level time series data (subjects) of size (61, 73, 61, 375)
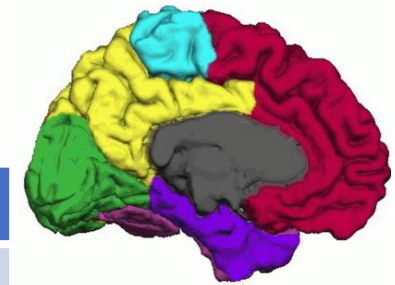
3D pixel

Cube size

Timestamps

signals

timesteps

61

73

61

Atlas & Parcellation

- Parcellation
- Atlas: maps or rules we follow to perform the parcellation

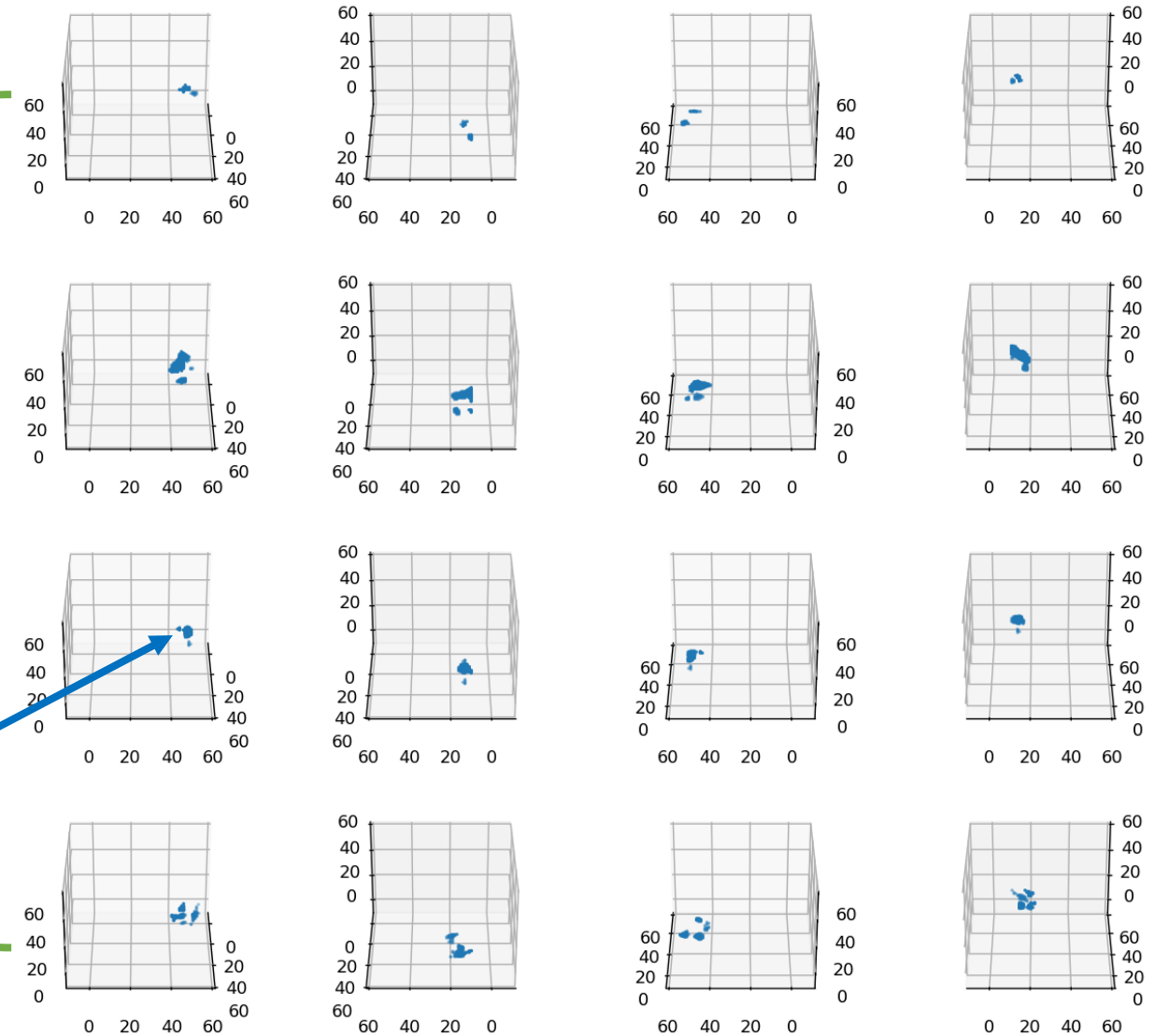| Atlas' Name	Number of Nodes
AAL3v1	164
Scheafer400	400
shen268	268
shen368	368

Parcellation Disjoint

- Voxels in the same regions are disjoint

Each row represents a region of the parcellation

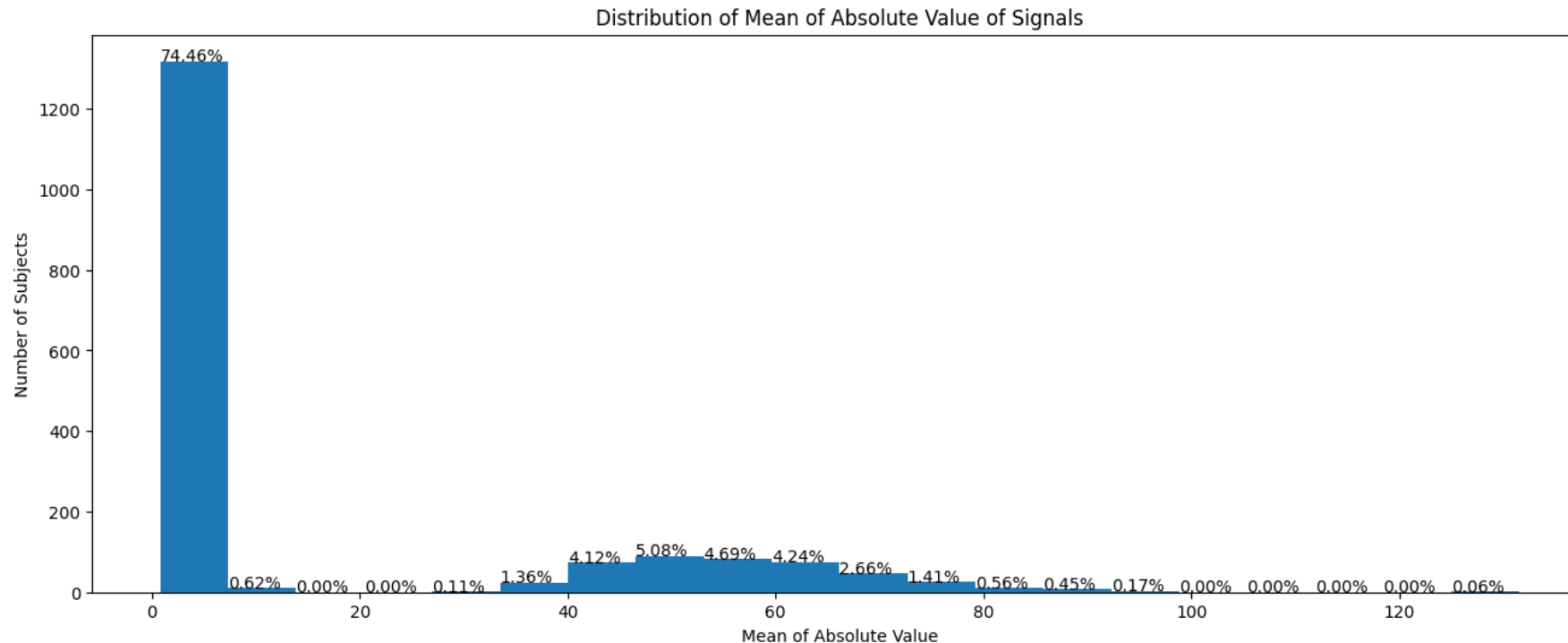points are voxels in the same region

Selected from Shen368 atlas

Signal Data Distribution

- Subjects $\{x_i\} = \{x_1, x_2, \cdots, x_n\}, x_i \in \mathbb{R}^{61 \times 73 \times 61 \times 375}$
  - $\text{MEAN}(|x_i|) \in \mathbb{R}$ (mean of absolute value)
- Two clusters
  - 74.5% are around zero.
  - 24.9% are between 30-80

Distribution of Mean of Absolute Value of Signals

Preliminary Model: Pooling

- **Assign Matrix**

$$S \in [0,1]^{n_l \times n_{l+1}}$$

Intuitively, each entry (i,j) of assignment matrix denotes a "**soft**" assignment of the node i to a cluster j in the next coarsened layer.

- **Pooling nodes**
  - $X^{(l+1)} = S^T X^{(l)} \in \mathbb{R}^{n_{l+1} \times d}$
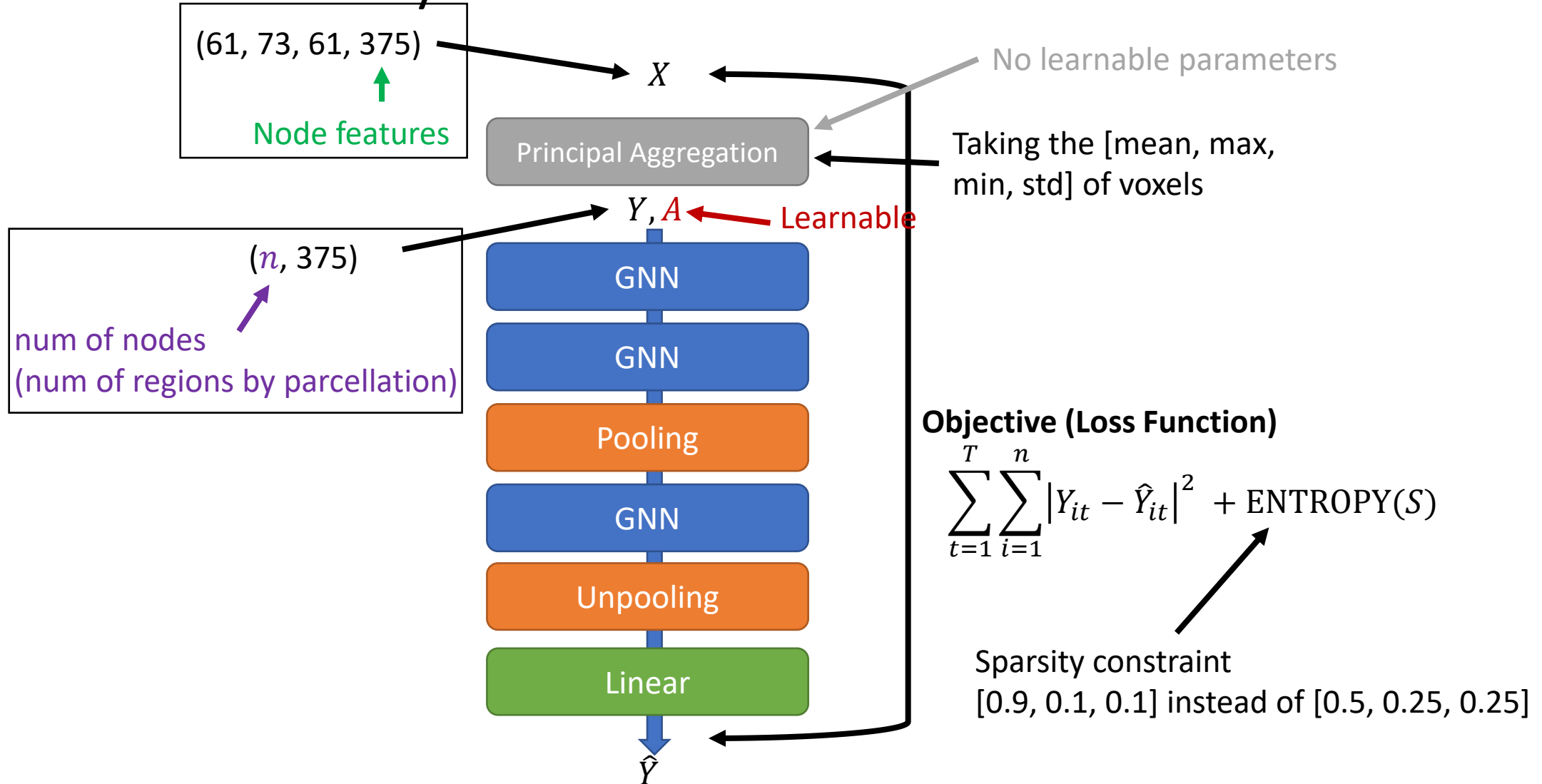
  - $[s_{11}, s_{12}, s_{13}] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}^{(l)} = s_{11} x_1^{(l)} + s_{12} x_2^{(l)} + s_{13} x_3^{(l)} = x_1^{(l+1)}$

- **Pooling edges** (adjacency matrix)
  - $A^{(l+1)} = S^T A^{(l)} S \in \mathbb{R}^{n_{l+1} \times n_{l+1}}$
  - Summing edges between two sets of nodes in a pair of clusters

Preliminary Model: Architecture

(61, 73, 61, 375)

Node features

X

No learnable parameters

Principal Aggregation

Taking the [mean, max, min, std] of voxels

Y, A — Learnable

(n, 375)

num of nodes
(num of regions by parcellation)

GNN

GNN

Pooling

GNN

Unpooling

Linear

$\hat{Y}$

Objective (Loss Function)

$$\sum_{t=1}^{T} \sum_{i=1}^{n} \left| Y_{it} - \hat{Y}_{it} \right|^2 + \mathrm{ENTROPY}(S)$$

Sparsity constraint
[0.9, 0.1, 0.1] instead of [0.5, 0.25, 0.25]

Results: Training Loss

Different Architectures

	GCN	Sage (mean)	Sage (max)	Sage (add)
N_layer = 1	**0.0260**	0.0408	0.0583	0.0414
N_layer = 2	0.0791	0.0692	0.0673	0.0498
N_layer = 3	0.0791	0.0791	Out of Memory	0.0791

Different Initial Aggregation

	GCN
Mean Aggregation	**0.0229**
Principal Aggregation	0.0260

Different Atlas

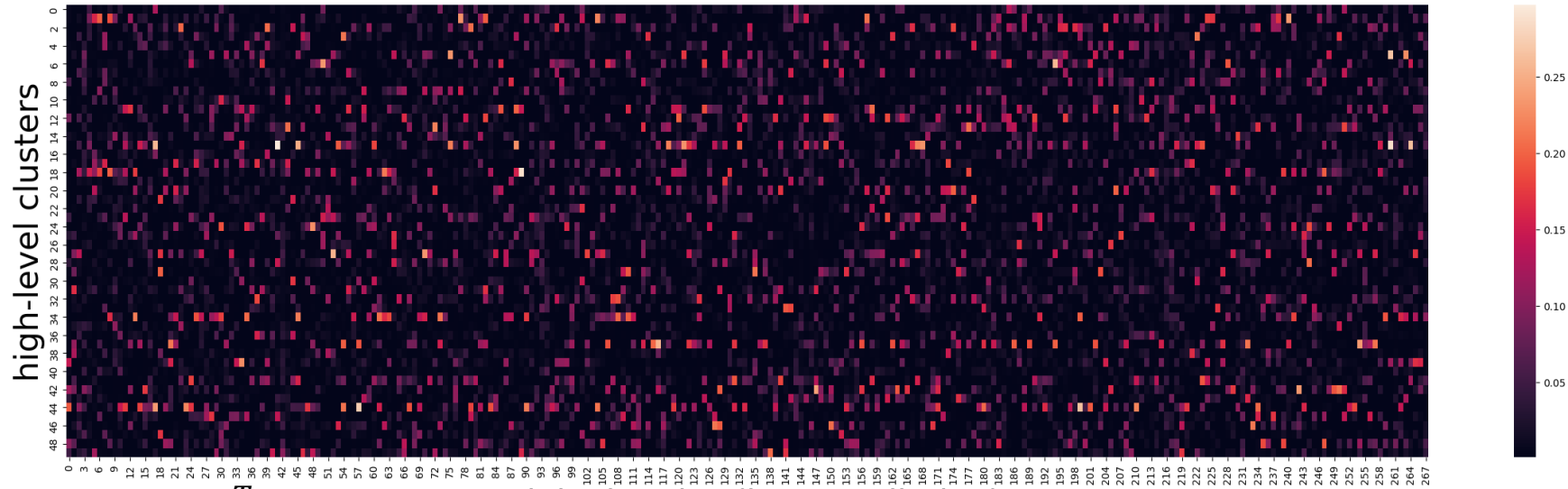| Atlas	GCN
Shen268	0.02596
Shen368	0.05481
Scheafer400	0.05364
AAL3v1	**0.02271**

Results: Regression Results
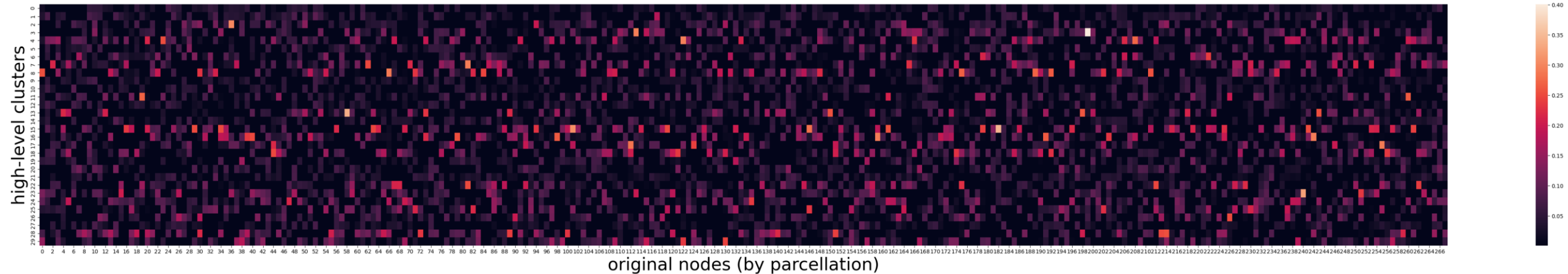
- Predictions are smoother
- A general issue in regression task
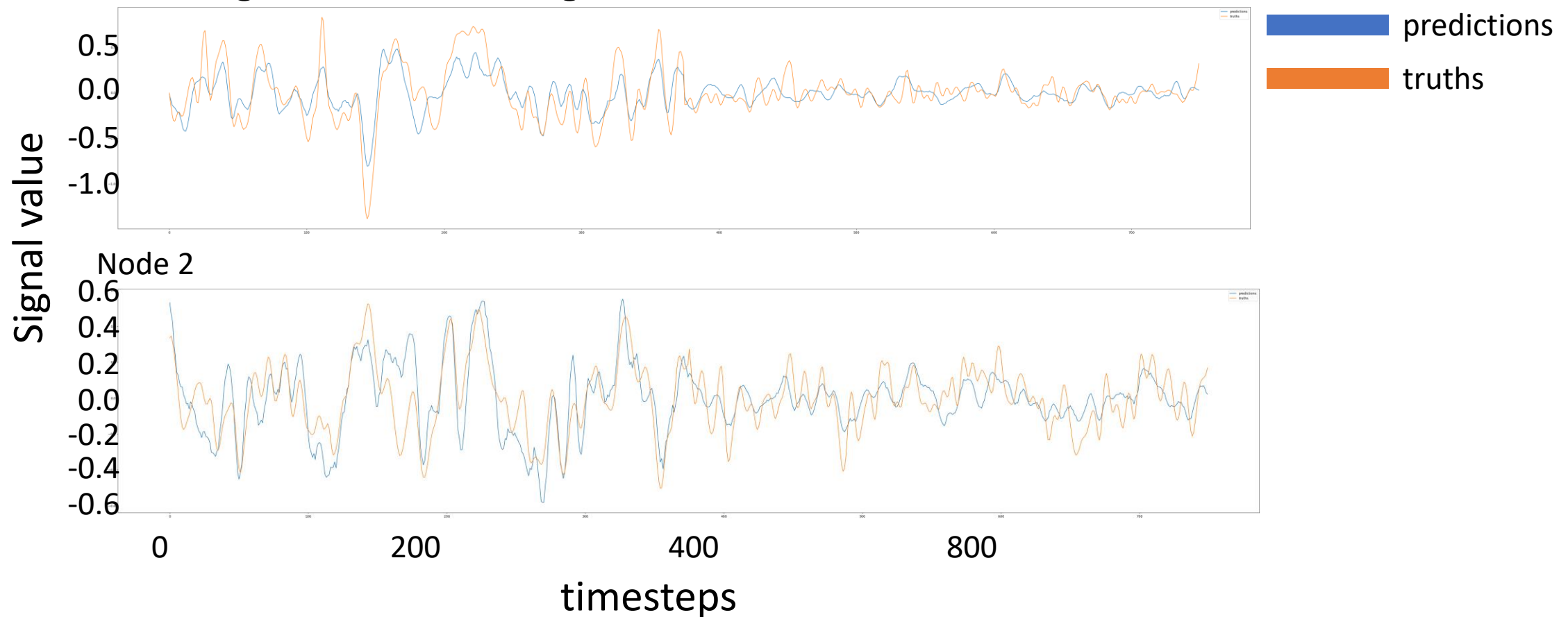
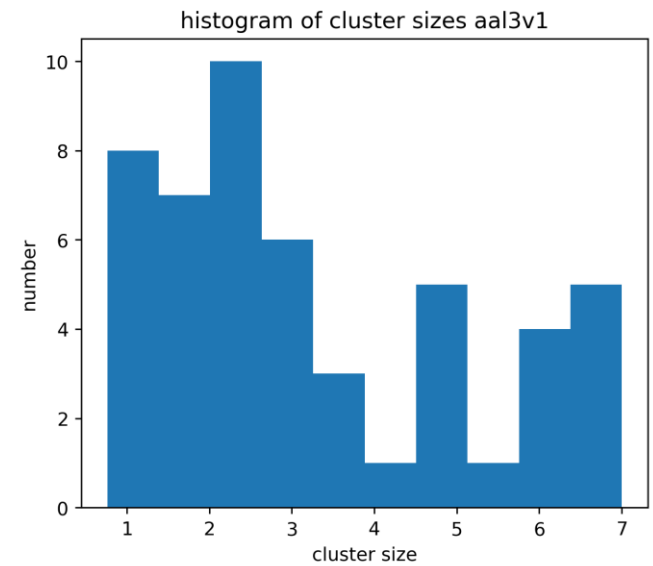# Results: Assignment Matrix

Number of Clusters = 50, i.e. S_{50}^T

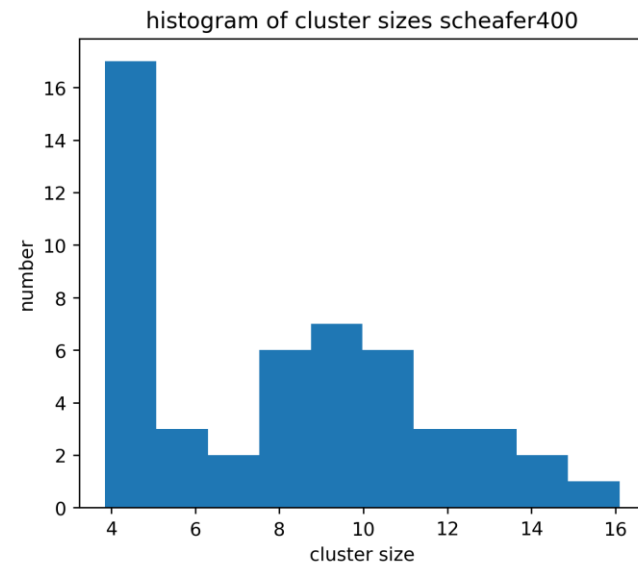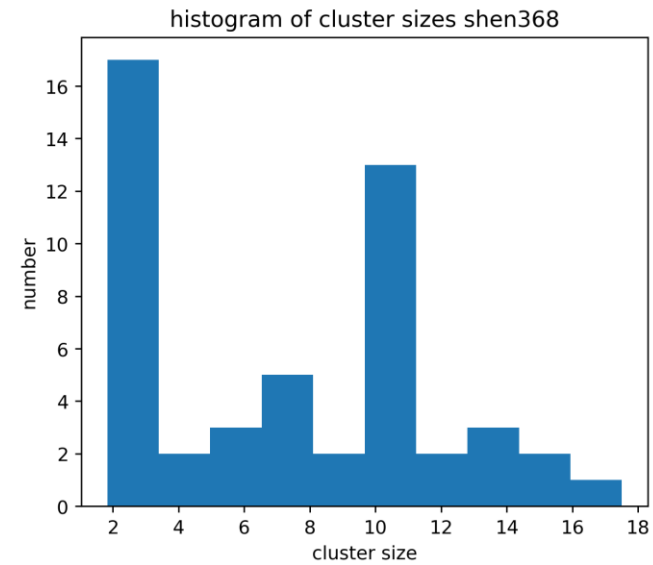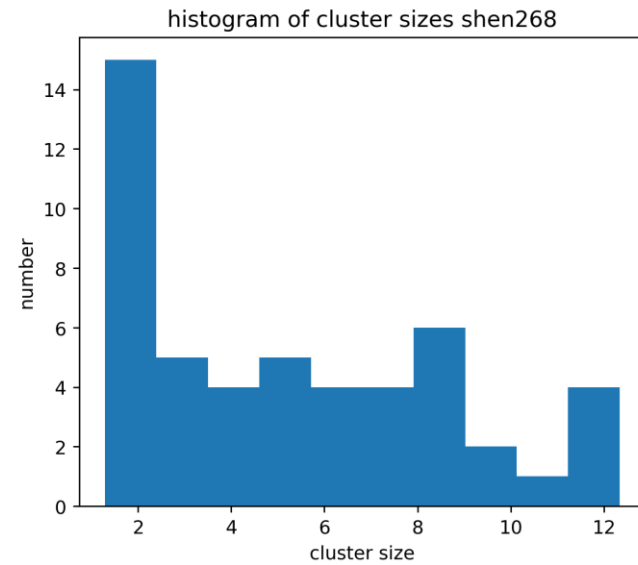high-level clusters

original nodes (by parcellation)

Number of Clusters = 30, S_{30}^T

high-level clusters

original nodes (by parcellation)
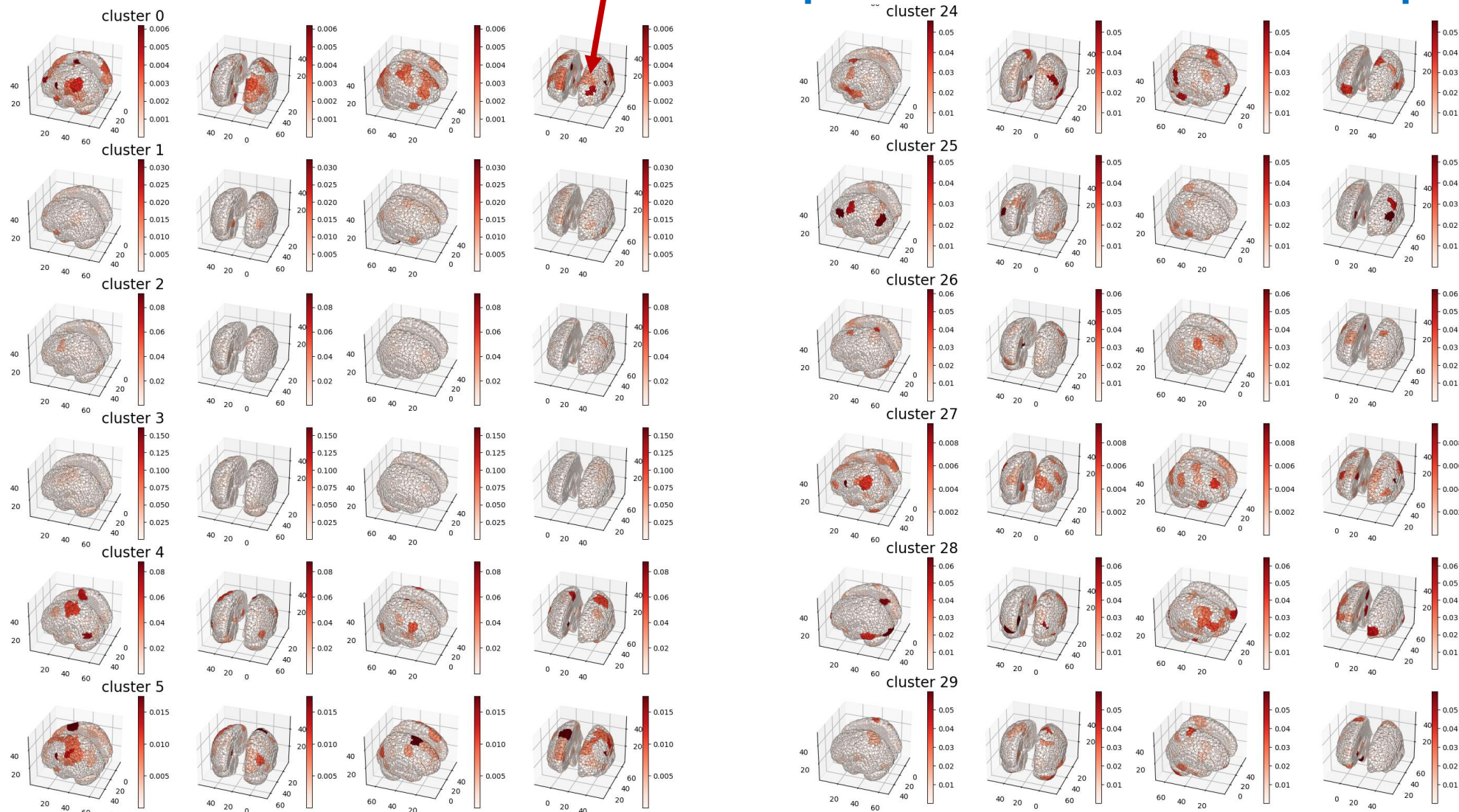
Results: Cluster Size

Results: Clustering
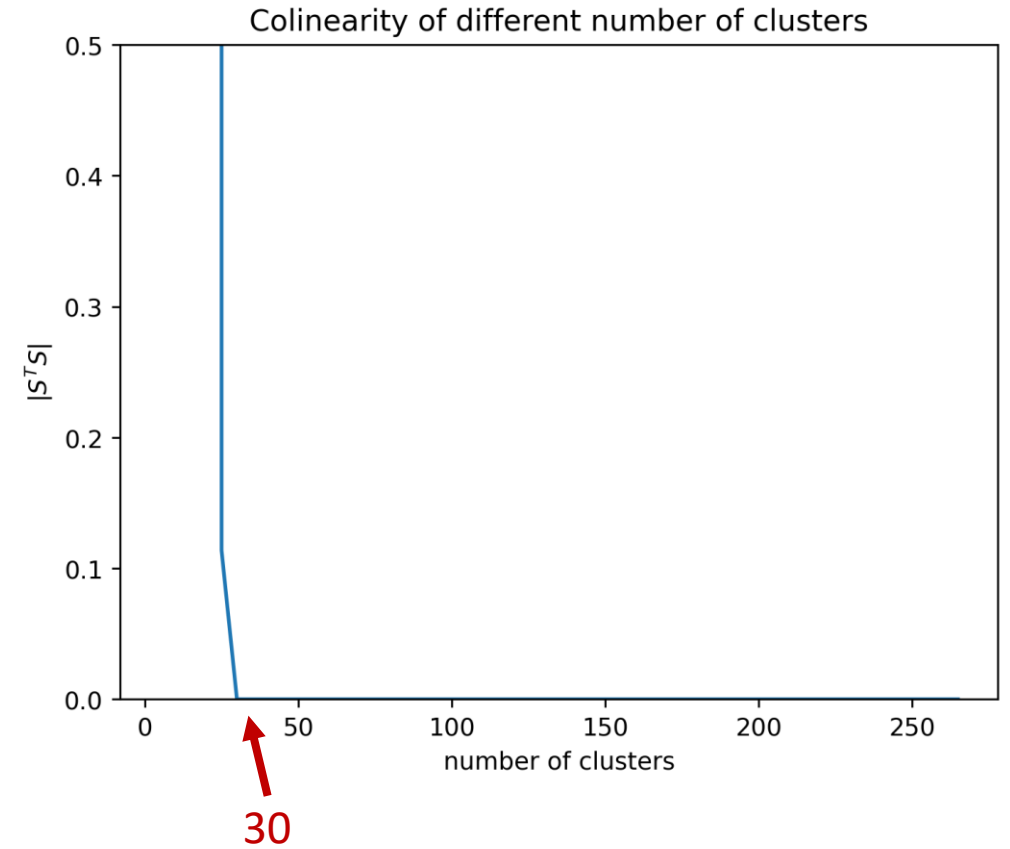
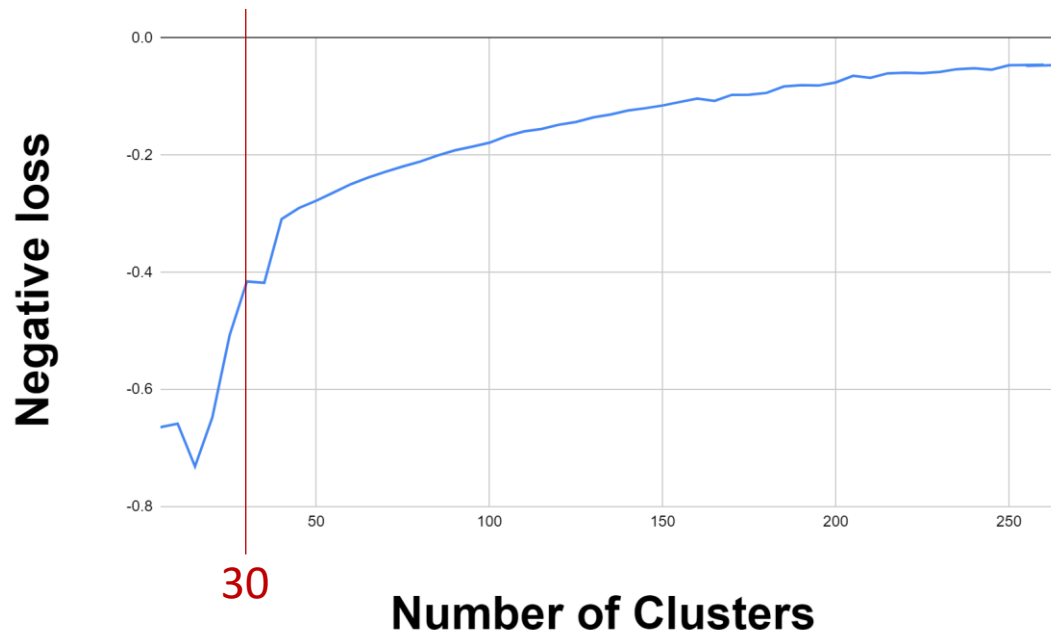Color strength represents assignment magnitude

Each column represents a different view angle in the 3D space

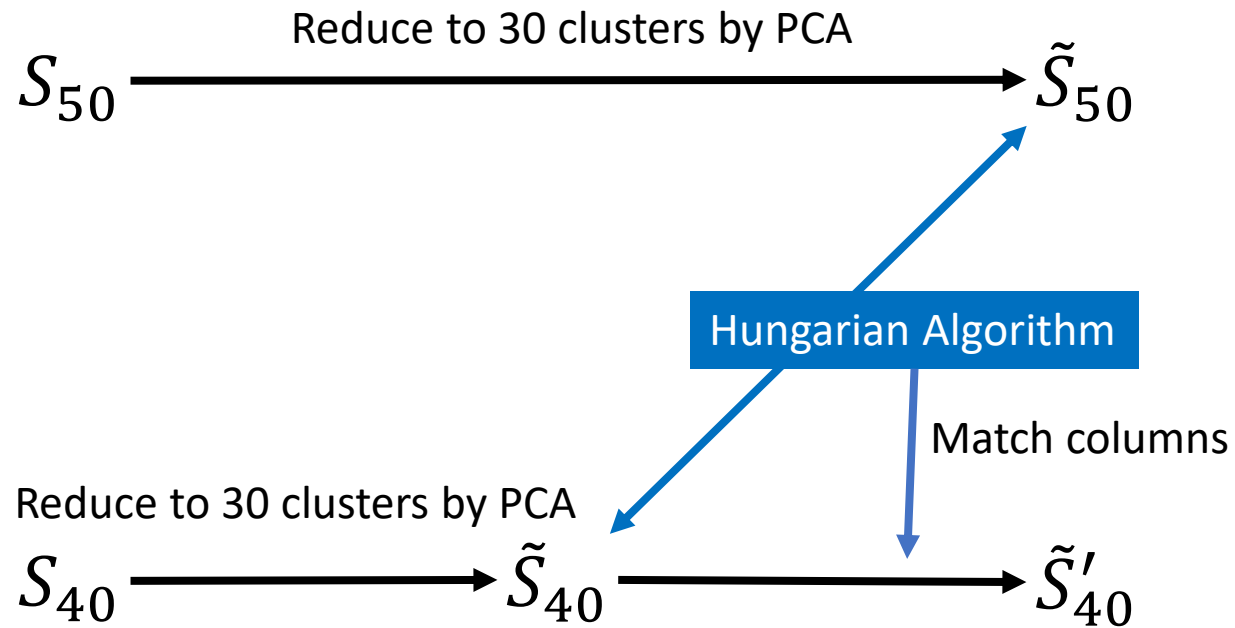Each row represents assignment magnitude of one cluster

Study of Number of Clusters

- Metric:
  - **Gram determinant** of matrix S: $|S^T S|$
  - Larger collinearity, smaller $|S^T S|$

Study of Number of Clusters

S_{50} $\xrightarrow{\text{Reduce to 30 clusters by PCA}}$ $\tilde{S}_{50}$

Hungarian Algorithm

Match columns

S_{40} $\xrightarrow{\text{Reduce to 30 clusters by PCA}}$ $\tilde{S}_{40}$ $\longrightarrow$ $\tilde{S}'_{40}$

Distance:
$\text{dist}(\tilde{S}_{50}, \tilde{S}_{40})$: 31.04
$\text{dist}(\tilde{S}_{50}, \tilde{S}'_{40})$: 29.36

$$\text{dist}(A, B) = \frac{1}{268} \sum_{i=1}^{268} \sum_{j=1}^{30} (A_{ij} - B_{ij})^2$$