Discriminating Human Brain Connectomes

Vikram Ravindra
Department of Computer Science
Purdue University

Joint work: Ananth Grama, Petros Drineas, David Gleich, Huda Nassar

Outline

- Can we identify individual specific markers in connectomes?
- Can we identify tasks using connectomes?
- Can we characterize how well people perform tasks?
- Is identity encoded in functional network?
- Ongoing work: can the algorithmic machinery be generalized to pathologies to identify connectomic markers?

Identifying Individual Specific Markers

- Find compact anatomical markers that characterize identities from neuroimages

- Images acquired in multiple sessions over periods of time may be susceptible to a range of variations.

  - ⋆ Natural variations in brain states

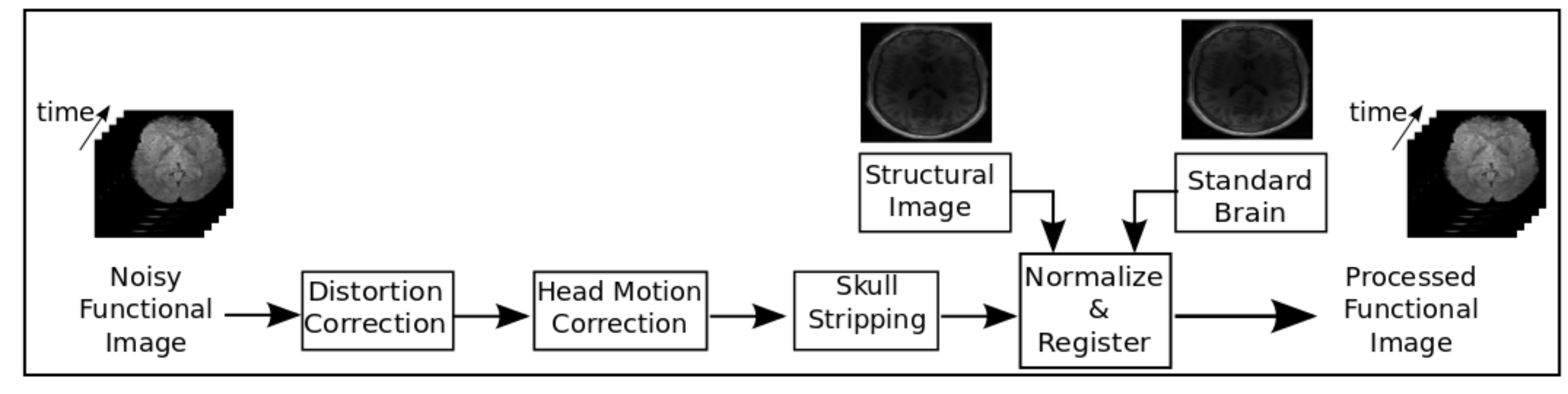
  - ⋆ Technical variations in instrumentation

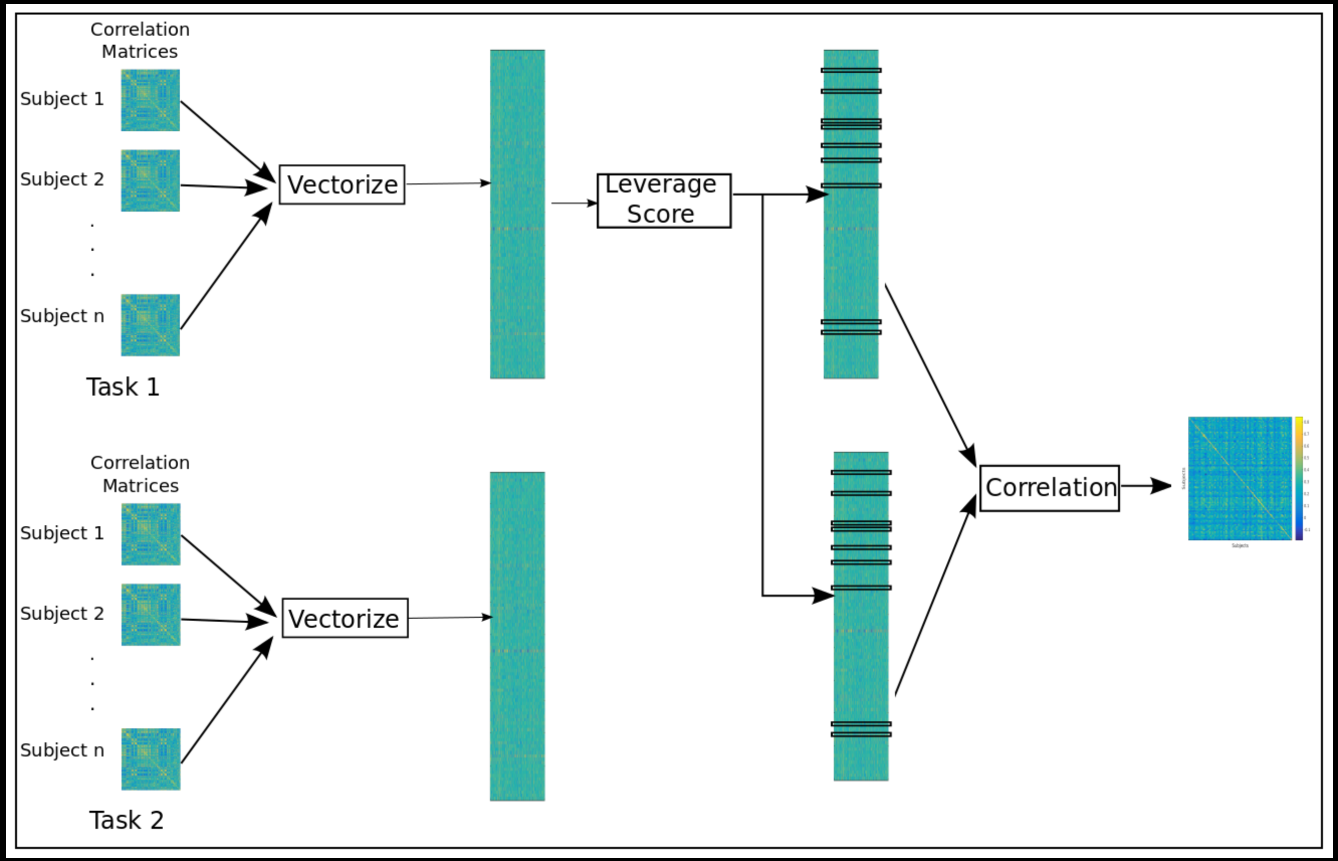
  - ⋆ Mapping stable invariants to anatomical regions

  - ⋆ Anatomical markers must be robust across populations

[Ravindra et al 2018] arXiv:1805.08649
[Ravindra et al 2019] arXiv:1908.03260

Identifying Individual Specific Markers





Statistical Leverage Scores of Matrices

- **Definition**: Given an arbitrary matrix $A \in \mathbb{R}^{m \times n}$, with $n > d$. Let $U \in \mathbb{R}^{m \times n}$ consisting of d left singular vectors of A, and let $U_{i,\star}$ denote the i-th row of U. Then, the statistical leverage scores of rows of A is given by

$$l_i = || U_{i,\star} ||_2^2$$

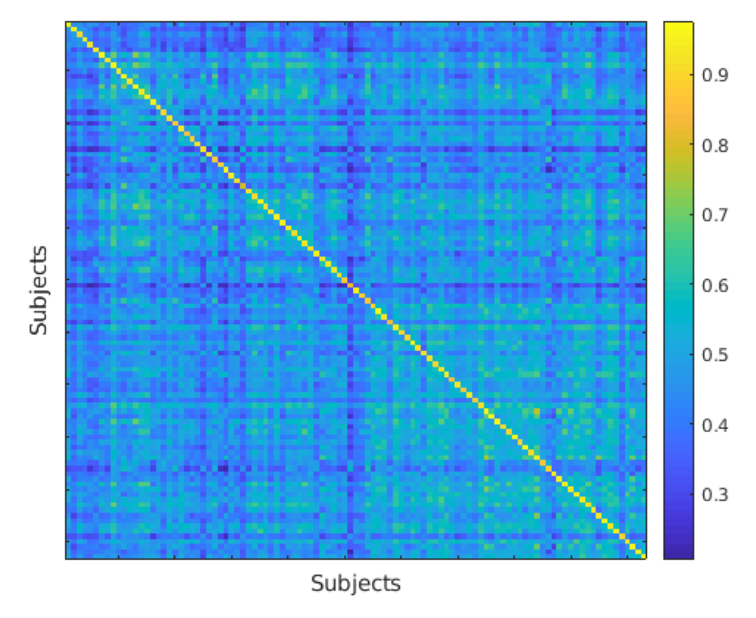
for $i \in \{1, \ldots, n\}$

Identifying Individual Specific Markers

- **Approach**: Use concepts from matrix sampling — *leverage score sampling* to select critical row sets that best represent the feature space.

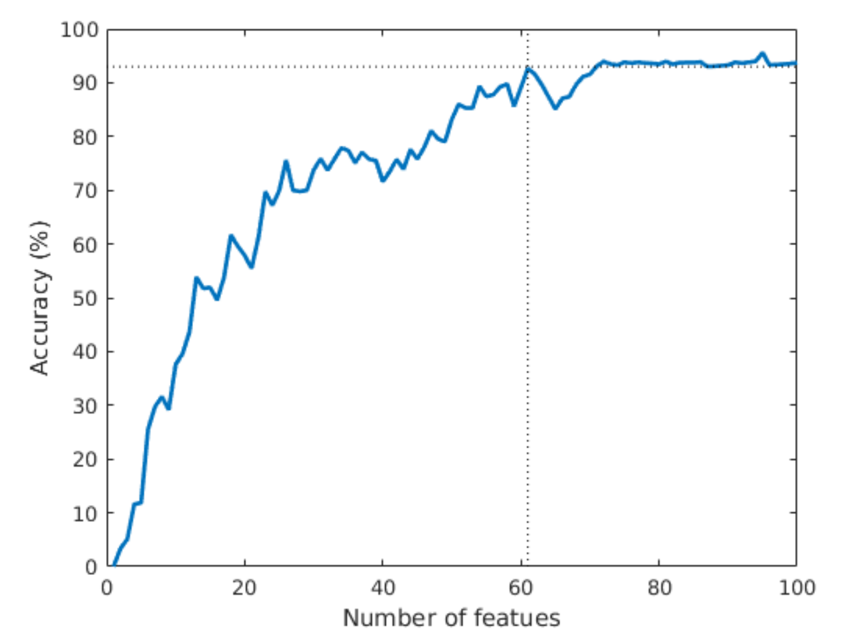
- **Definition**: Given a matrix $A \in \mathbb{R}^{m \times n}$ with $m \gg n$, let $U \in \mathbb{R}^{m \times n}$ be an orthonormal matrix that spans the column space of A ($U^T U = I$ and $UU^T = P_A$ — a projection matrix that spans the column-space of A).
  We sample rows in proportion to their leverage scores, i.e.,
  $$p_i = \frac{||U_{i,\star}||_2^2}{\sum_i ||U_{i,\star}||_2^2} = \frac{1}{n}(P_A)_{i,i} \qquad \forall i \in \{1 \dots m\}\,.$$

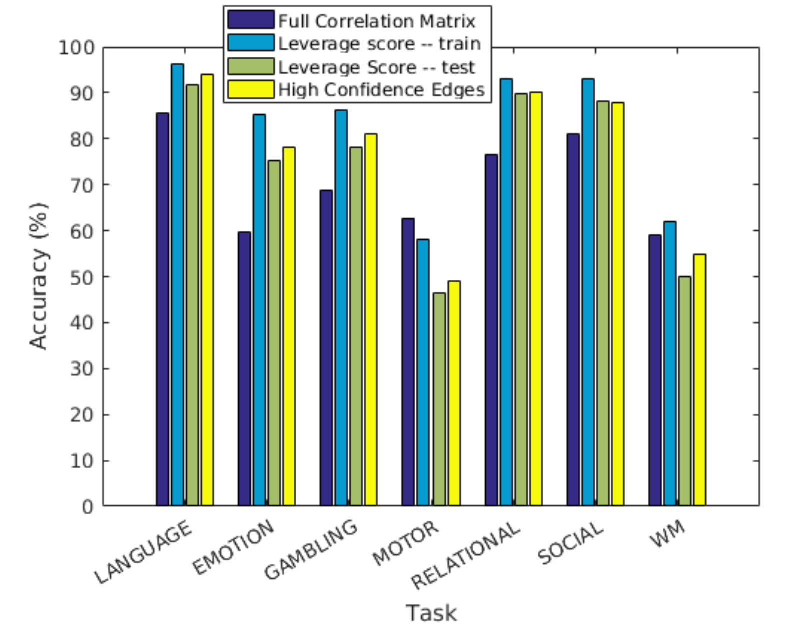
Identifying Individual Specific Markers



High diagonal values show that images of same subjects on different days are more similar than images drawn from different subjects

We need only 60 features to accurately predict Identities of subjects





Features selected by leverage score sampling accurately predicts identities for a range of tasks

Identifying Individual Specific Markers

Principal Component	Training Accuracy(%	Testing Accuracy(%
All	88.65 ± 1.76	88.98 ± 1.72
2:end	94.30 ± 1.35	71.76 ± 8.76
11:end	96.74 ± 1.00	69.61 ± 8.94
21:end	95.03 ± 1.90	69.44 ± 8.99
31:end	71.97 ± 6.08	68.95 ± 9.07
41:end	72.77 ± 1.74	65.70 ± 9.59
Leverage Score	96.23 ± 2.24	93.11 ± 3.61

Comparison of Leverage Score sampling with low-rank approximations

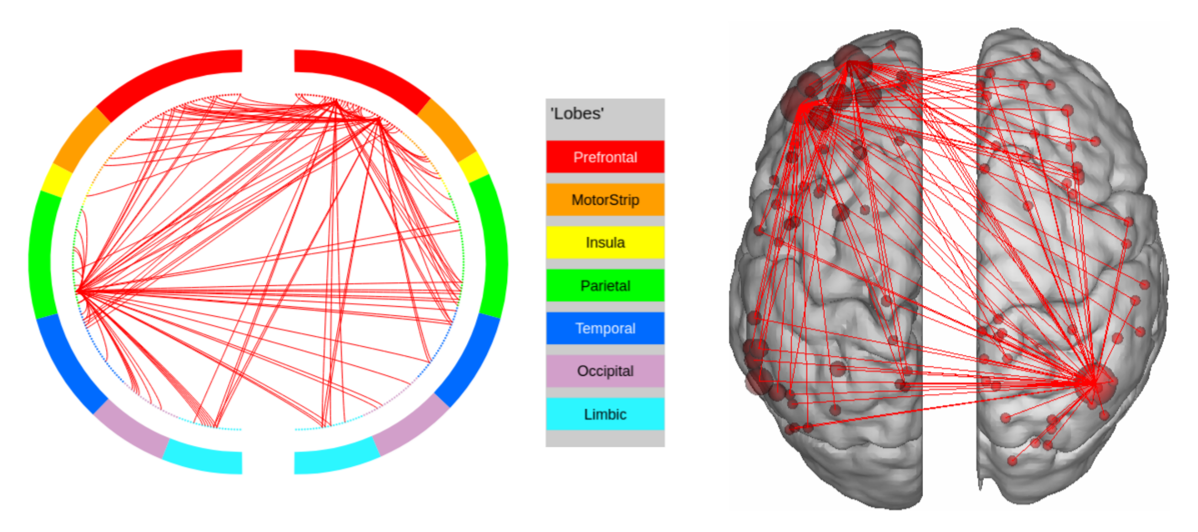
Training/ Test Split	Training Accuracy(%)	Testing Accuracy(%)
80/20	96.23 ± 2.24	93.11 ± 3.61
50/50	96.30 ± 2.59	92.94 ± 3.82
30/70	96.81 ± 3.07	90.23 ± 4.30
20/80	97.01 ± 3.22	87.60 ± 5.27
10/90	97.72 ± 2.65	81.86 ± 7.15

A 50/50 split suffices for the markers to generalize to the test set

Anatomical Markers of Identity

- We retain features in the original matrix. Hence, they are directly interpretable (thus making them applicable to identifying markers with various neurodegenerative diseases)

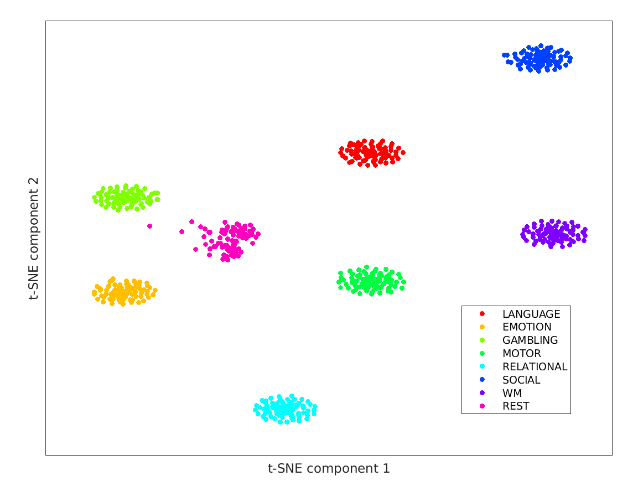
- We require fewer than 100 markers from over 64k candidates demonstrating high sensitivity.

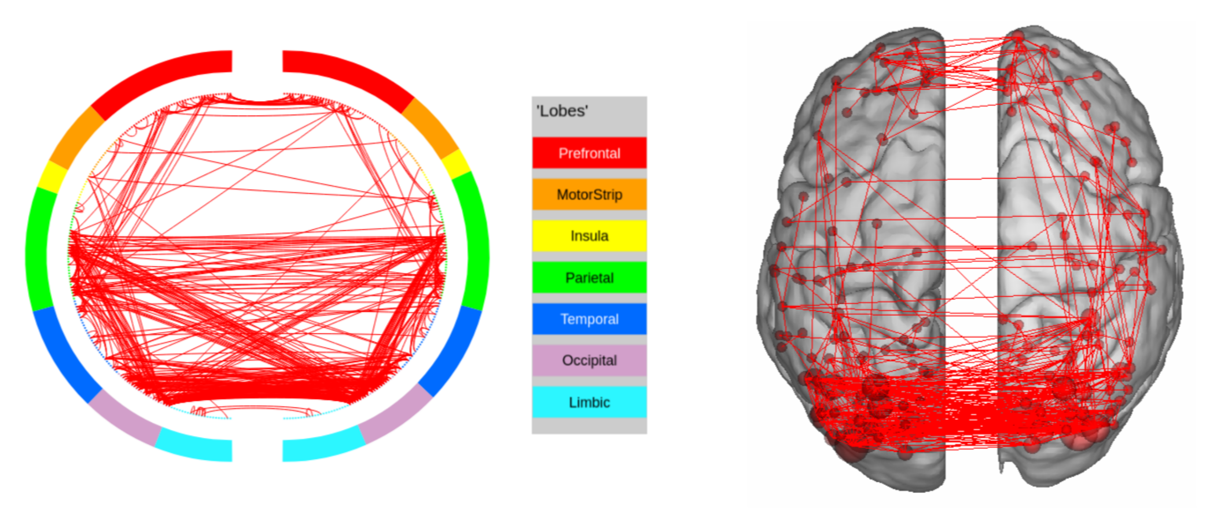


Identifying Tasks from Connectomes

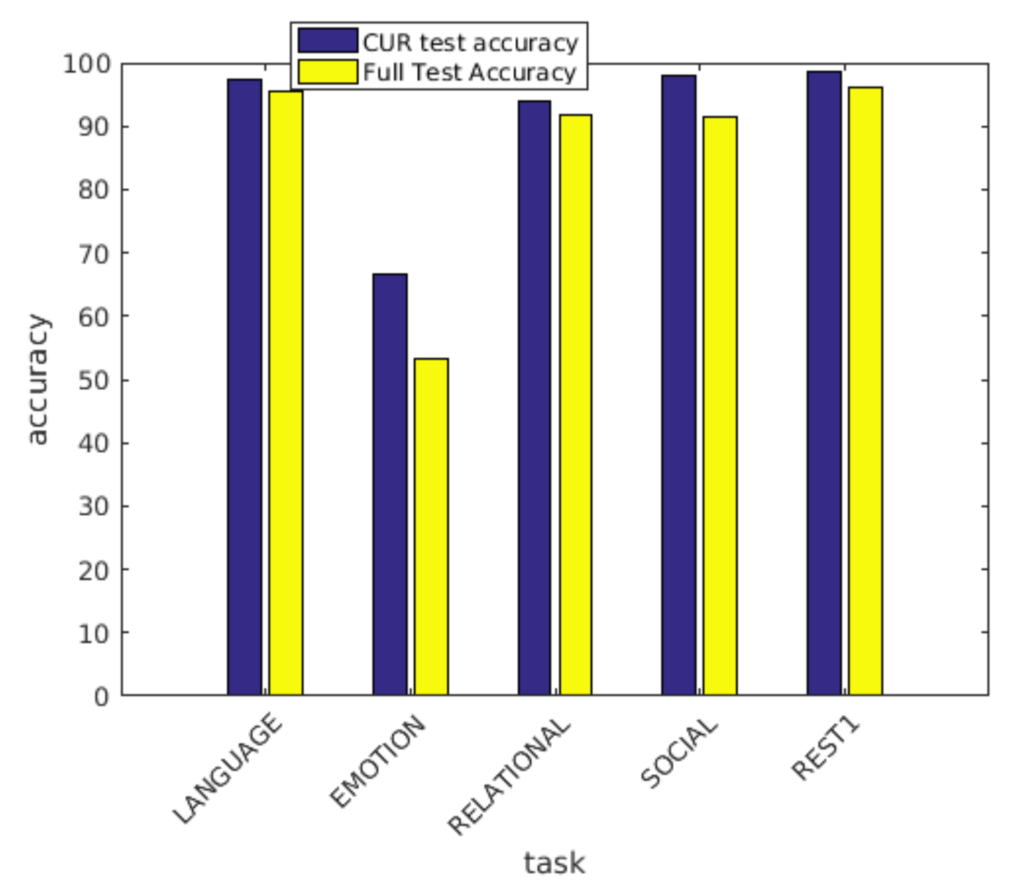
- A t-distribution stochastic neighbor embedding (t-SNE) of all connectomes show that tasks are separable into clusters.

- We used leverage-score sampling to find markers that distinguish the tasks performed by the same subject.

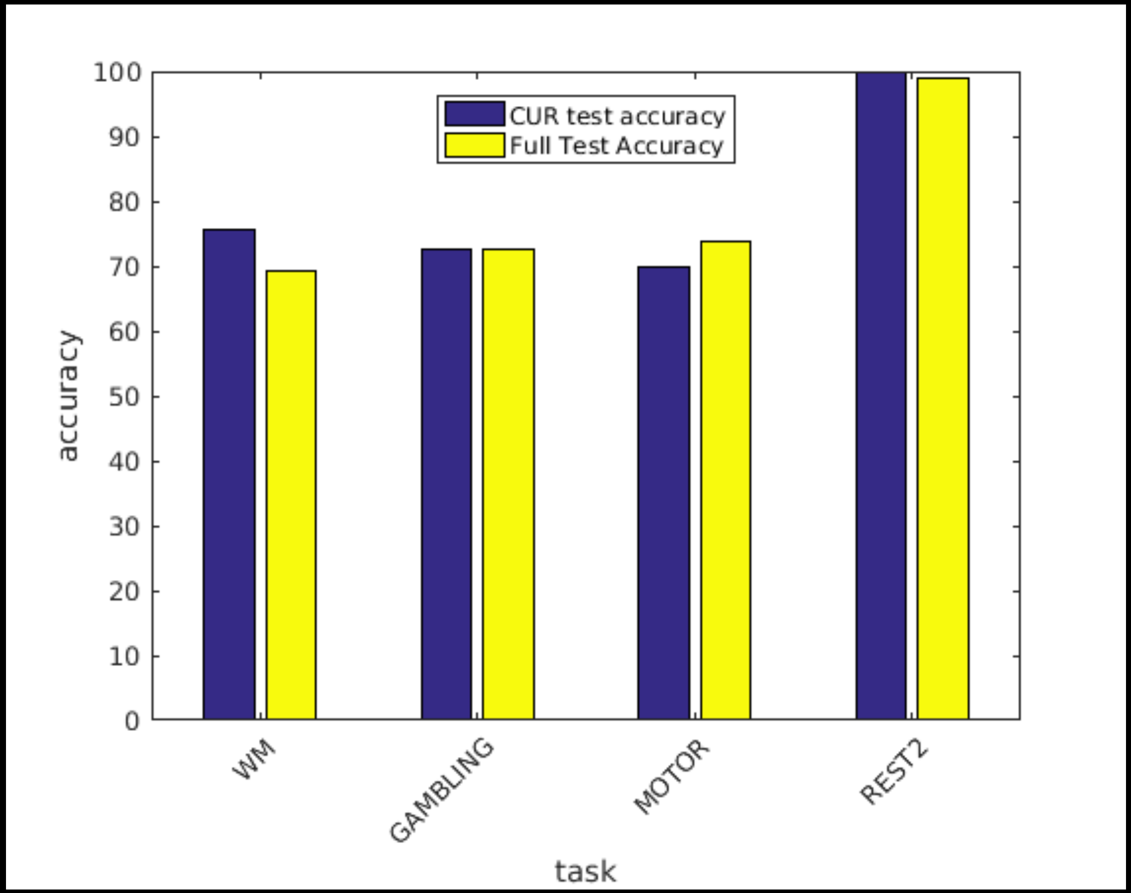




Identifying tasks from Connectomes



Task prediction accuracy on Day 1



Task prediction accuracy on Day 2

Predicting Task Efficacy

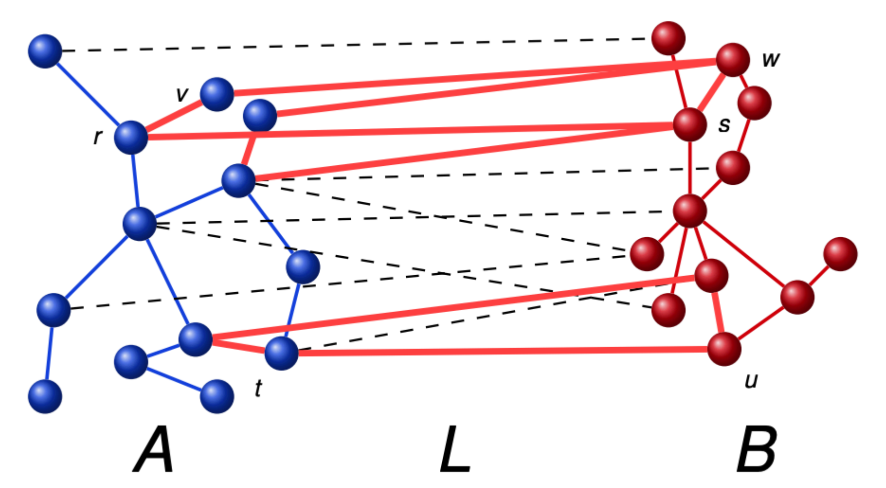
- We use markers selected by our leverage-score sampling to predict the efficiency with which subjects perform tasks.

- We use support vector regression to learn performance metrics

Task	Train Error %	Test Error %
Language	0.33 ±0.11	1.52 ± 0.20
Emotion	0.28 ±0.07	0.60 ± 0.37
Relational	0.44 ± 0.04	2.74 ± 0.34
Working Mem	0.57 ± 0.12	1.93 ± 0.41

Individual Specific Markers in Functional Connectomes

- Functional network pipelines align (by registering to a common coordinate) and normalize (by registering to high resolution image of same subject) using structure
- Previously identified signatures may be coded in function or structure
- Can we identify markers in the functional connectomes that characterize identity to similarly high accuracy?



Individual Specific Markers in Functional Connectomes

- Register each session to the first image (do not align to common coordinate)

- Restrict to the cortex and compute voxel level correlations across time steps to identify statistically significant co-firing regions

- This results in a sparse representation of the functional connectomes as a weighted graph with edge weights representing preferred distances

- Identifying markers translates to node correspondences across graphs modeled as a rigid graph alignment problem

[Ravindra et al. 2019] arXiv:1908.03201

Rigid Graph Alignment

- $$F = \max_{X} \underbrace{\alpha \mathbf{L} \bullet \mathbf{X} + \beta \mathbf{A} \bullet \mathbf{X}\mathbf{B}\mathbf{X}^{\mathrm{T}}}_{\text{network alignment}} + \overbrace{\gamma C_A \Omega^T C_B^T \bullet X}^{\text{structural alignment}}$$

$$\sum_i X_{ij} \leq 1 \ \forall j = 1, \ldots, |V_B|$$

$$\sum_j X_{ij} \leq 1 \ \forall i = 1, \ldots, |V_A| \quad X_{ij} \in \{0,1\}$$

- **Approach**: Use functional and structural information alternately in an Expectation-Maximization (EM) framework

Rigid Graph Alignment

- E - step

  $$\mathbf{X}^{(t)} = \text{argmax}_{\mathbf{\Omega}} \mathbf{F}(\mathbf{X}, \mathbf{\Omega}^{(t)})$$

- M-step

  $$\mathbf{\Omega}^{(t+1)} = \text{argmax}_{\mathbf{X}} \mathbf{F}(\mathbf{X}^{(t)}, \mathbf{\Omega})$$

- Repeat

  $\mathbf{L} = \text{get\_prior}(C_A, C_B)$

  $\mathbf{X} = \text{align}(A, B, \mathbf{L})$

  $B = \mathbf{X} B \mathbf{X}^T$

  $\mathbf{\Omega} = \text{transform\_coordinates}(C_A, C_B, \mathbf{X})$

  $C_B = \mathbf{X} C_B \mathbf{\Omega}$

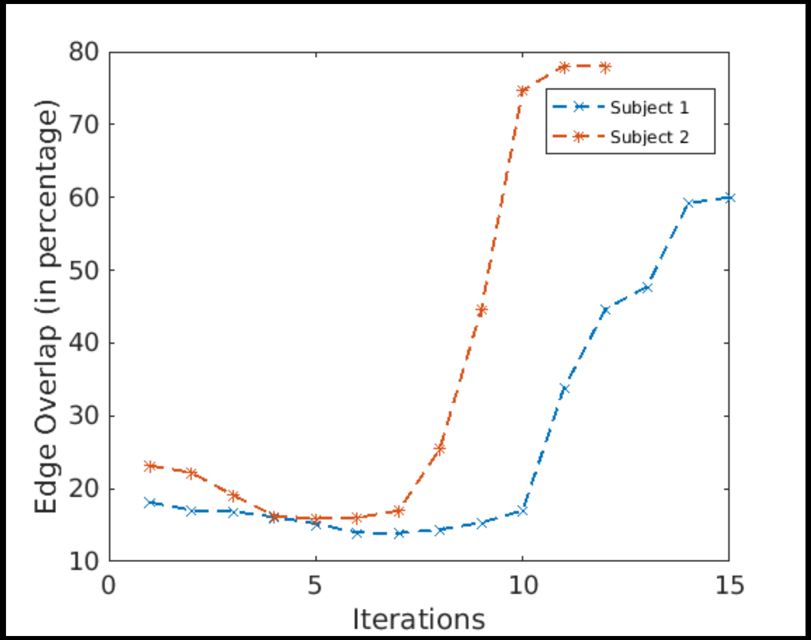
  until *converged*

$$\mathsf{L}_{i,j} = \begin{cases} \textbf{exp}(-||C_{A_i} - C_{B_j}||_2^2) & ||C_{A_i} - C_{B_j}||_2^2 \le d_i^k \\ 0 & \textbf{otherwise} \end{cases}$$

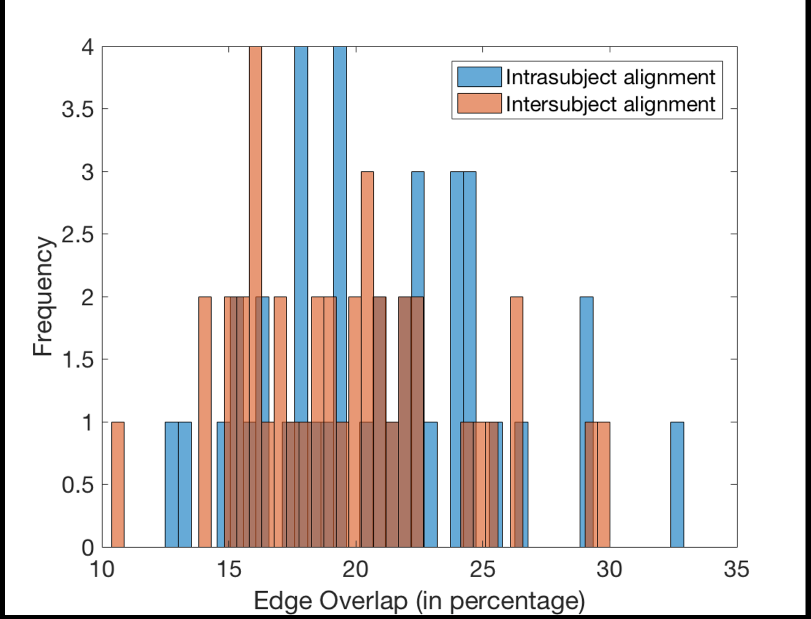
Exact Alignment is NP hard!

Use Kabsch's algorithm

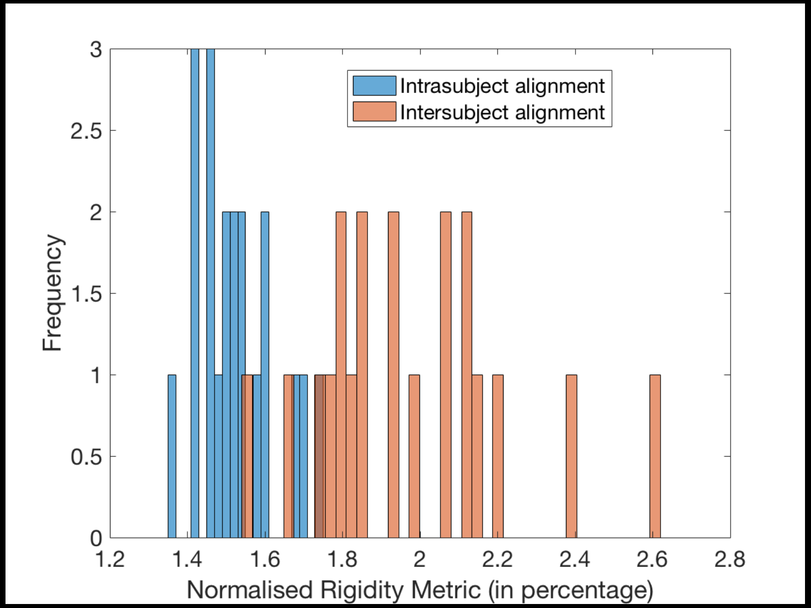
Individual Specific Markers in Functional Connectomes



The number of corresponding edges increase when we use both function and structural information



Network Alignment algorithms cannot separate inter-subject and intra-subject alignments

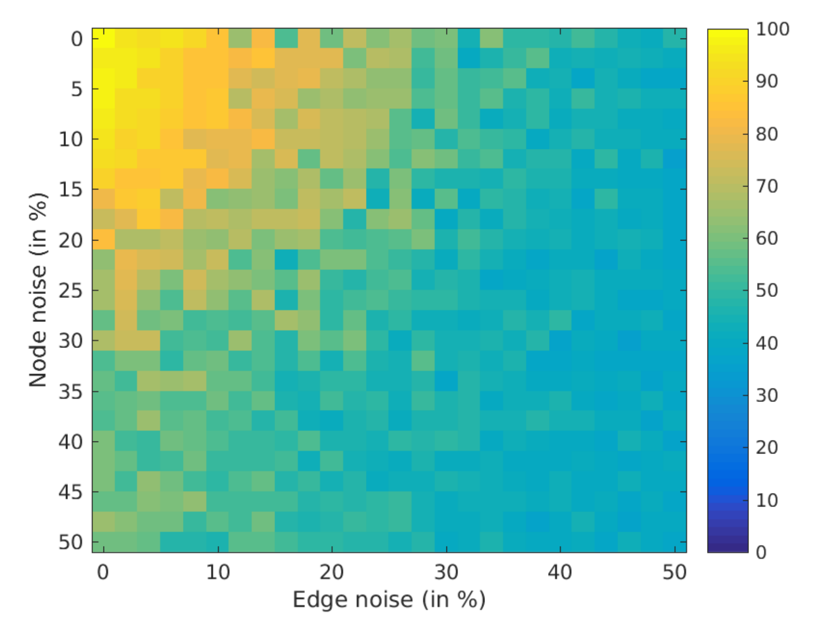


Rigid Graph Alignment can separate inter-subject and intra-subject alignments

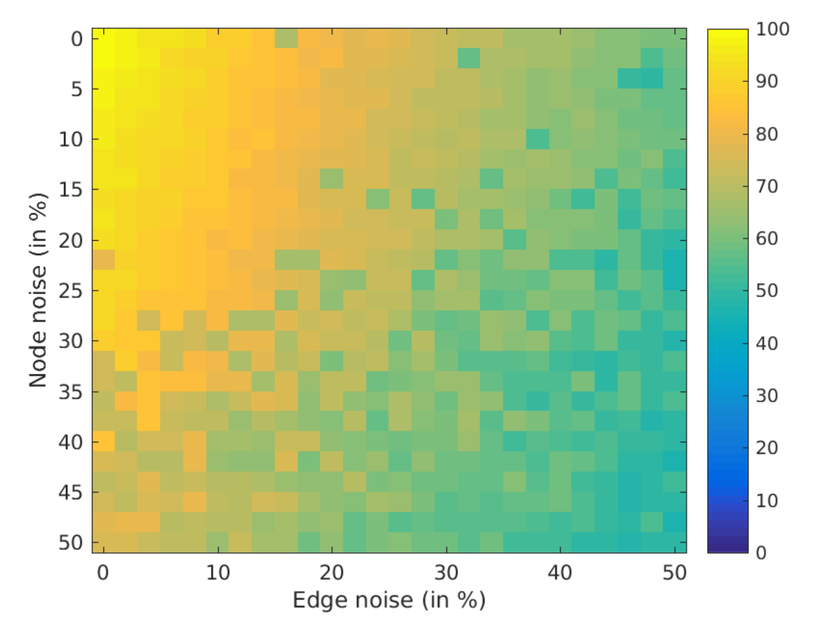
Stability of Rigid Graph Alignment

- We create synthetic graphs using a preferential-attachment model
- Nodes are assigned random coordinates
- We perturb positions of nodes and add/remove edges

Stability of Rigid Graph Alignment

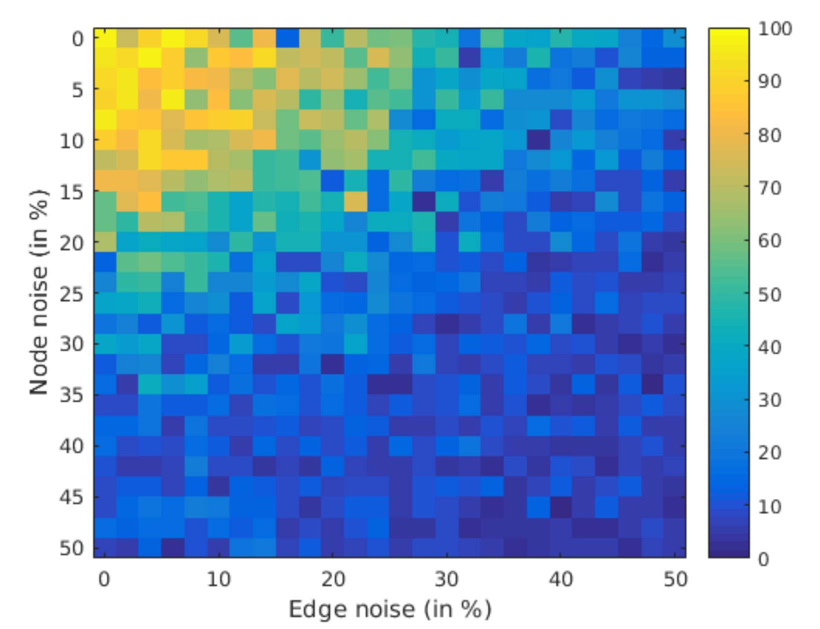


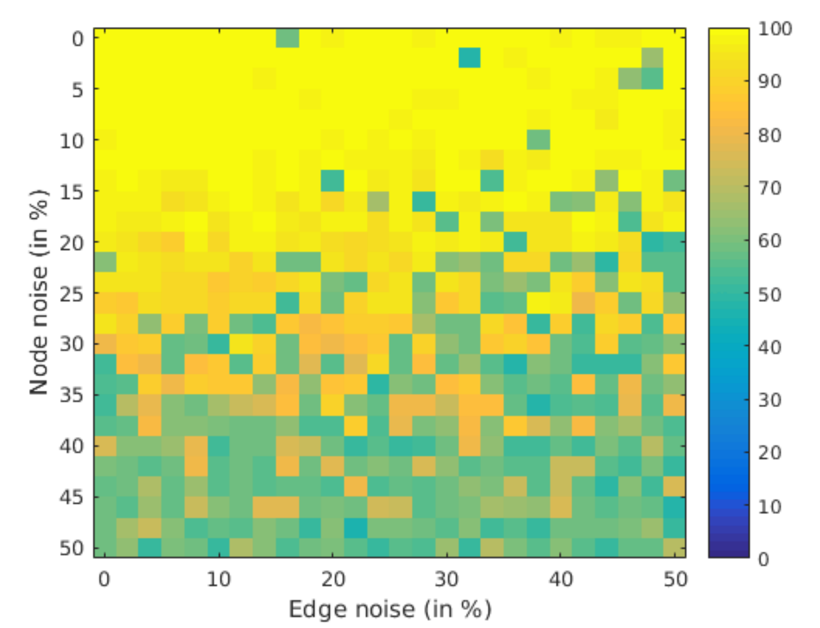
Edge-overlap decreases with node and edge noise in regular network alignment algorithms



Rigid graph alignment is robust to edge and node noises, as evidenced by high edge-overlap values for substantial value of noise

Stability of Rigid Graph Alignment





Number of correct node matches decrease for modest noise with regular network alignment

Number of correct node matches remains unaffected for substantial noise

Future Questions

- What features are actually discriminating? Are Leverage score features the best features or are there other feature sets that are more discriminating and compact?
- What are markers of neurodegenerative disorders?
- How do markers evolve with progression of diseases?
- Can we map image phenotypes to genotypic markers and to drug targets?
- Can we functionally align able bodied and injured brain to reason about how we re-learn tasks after injury?

Thank You!