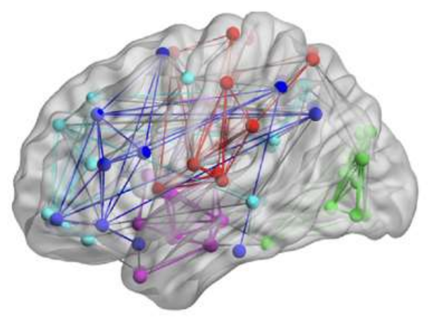


291K Machine Learning

Attention for Brain Network Analysis

Christos Zangos

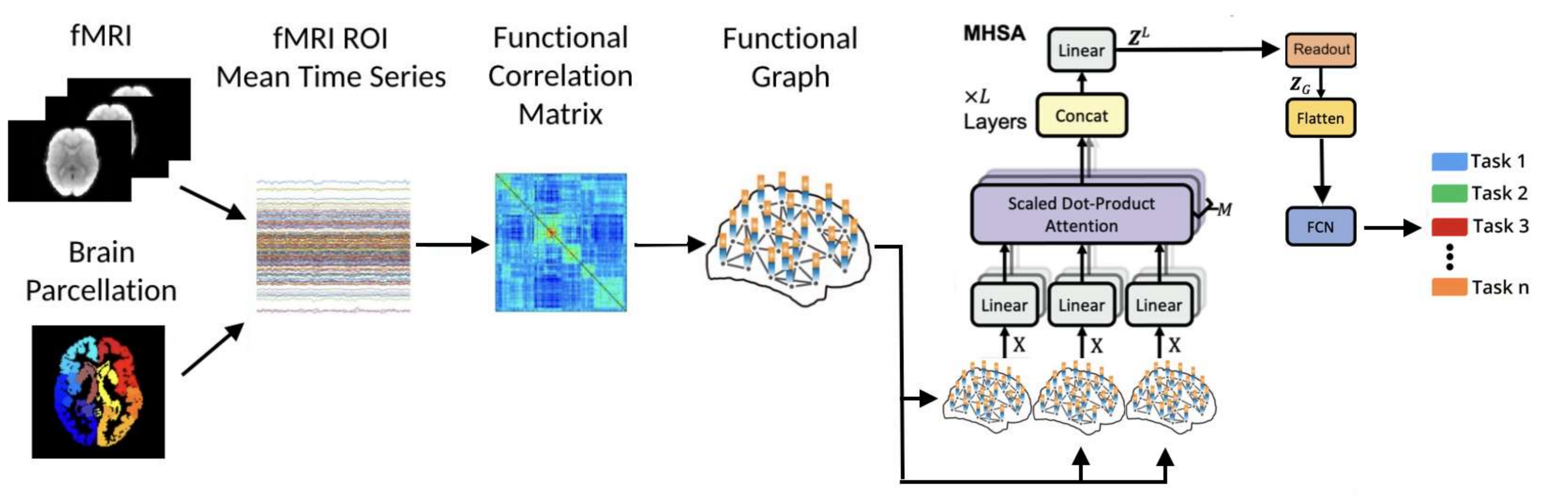
1 Problem Definition

- Brain Network Analysis
  - Classify brain networks based on attention mechanism
- Challenges
  - Scalability
    Too many edges given the number of nodes
  - Brain networks are complete graphs
    Centrality and spatial encodings cannot be applied
- Significance
  - Diagnosis of neurological diseases
  - Monitor the effectiveness of certain therapies
  - Discover novel treatments

2 Brain Network

- Definition of Brain Network
  - $\boldsymbol{X} \in \mathbb{R}^{V \times V}$
  - V is the number of nodes
  - fMRI: most commonly used for brain network construction
    Non-invasive
  - Nodes: Regions Of Interest (ROIs) given an atlas
  - Edges: pairwise correlations between BOLD signal from ROIs

3 Attention Framework



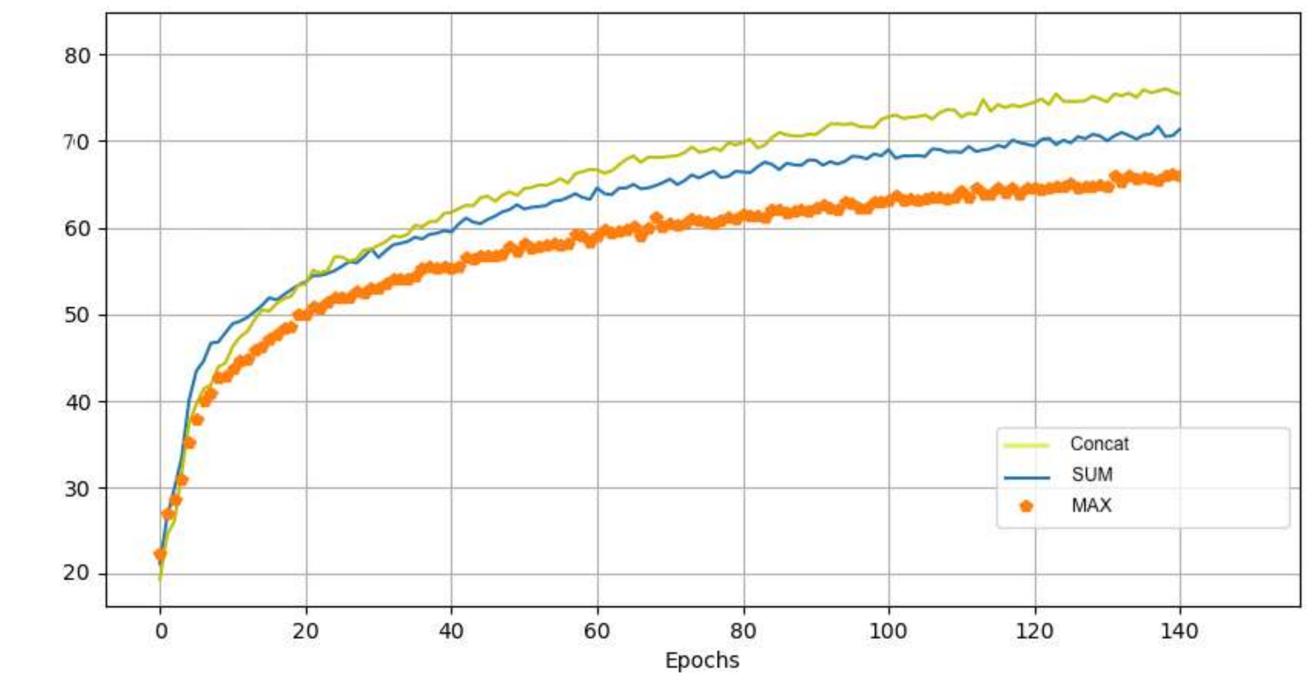
- I use MHSA to generate more expressive node features where $\boldsymbol{Z}^L = MHSA(\boldsymbol{X}) \in \mathbb{R}^{V \times V}$

$$\boldsymbol{Z}^l = (\|_{m=1}^{M} \boldsymbol{h}^{l,m})\boldsymbol{W}_{\mathcal{O}}^l, \boldsymbol{h}^{l,m} = \text{Softmax}\left(\frac{\boldsymbol{W}_{\mathcal{Q}}^{l,m}\boldsymbol{Z}^{l-1}(\boldsymbol{W}_{\mathcal{K}}^{l,m}\boldsymbol{Z}^{l-1})^\top}{\sqrt{d_{\mathcal{K}}^{l,m}}}\right)\boldsymbol{W}_{\mathcal{V}}^{l,m}\boldsymbol{Z}^{l-1}$$

- $\boldsymbol{Z}^0 = \boldsymbol{X}$
- $\boldsymbol{M}$ is the number of heads
- $\boldsymbol{l}$ is the layer index
- $\boldsymbol{W}_{\mathcal{O}}^l, \boldsymbol{W}_{\mathcal{Q}}^{l,m}, \boldsymbol{W}_{\mathcal{K}}^{l,m}, \boldsymbol{W}_{\mathcal{V}}^{l,m}$ are learnable model parameters
- $d_{\mathcal{K}}^{l,m}$ is the first dimension of $\boldsymbol{W}_{\mathcal{K}}^{l,m}$

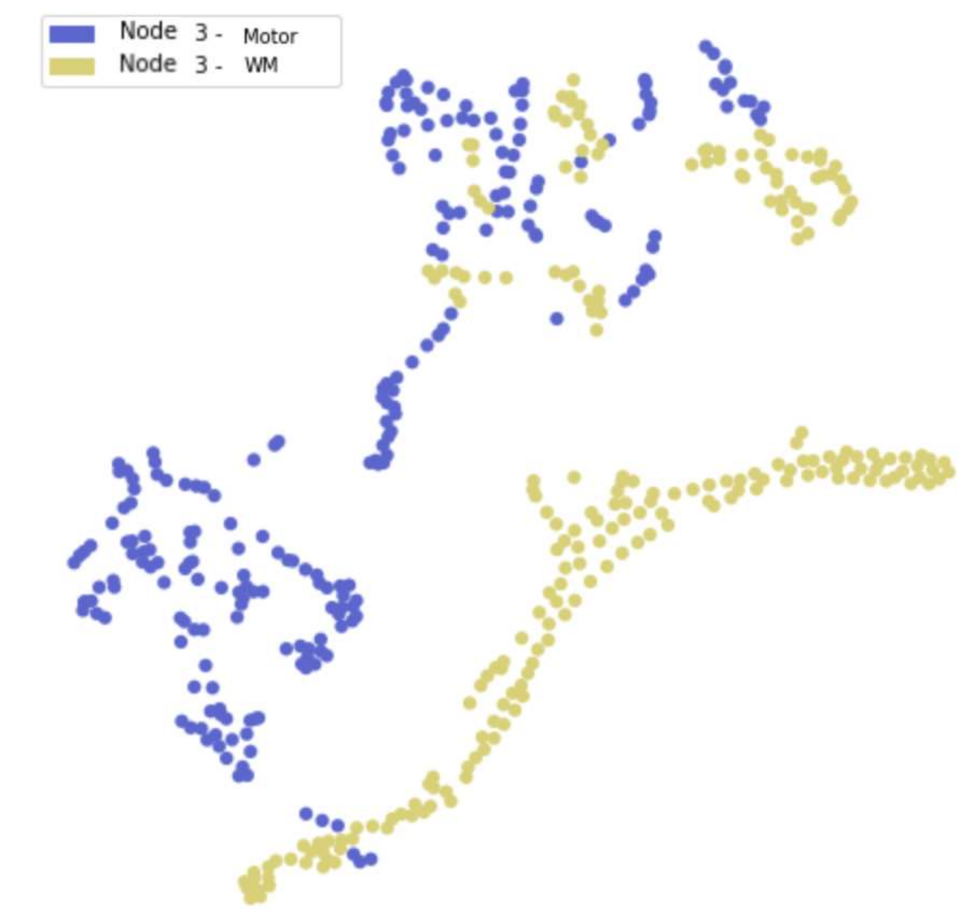
4 Experiment

- **Task-fMRI time series from Human Connectome Project**
  - 340 subjects
  - 360 ROIs *(Glasser Parcellation)*
  - 3 classes
  - 2034 brain graphs
- **Connection profiles as node features**
  - Low-cost
  - Encode structural & positional information
  - Better for GNNs
- **Testing different READOUT functions** (Validation AUC)
  - $\boldsymbol{Z_G} = Concat(ReLU(W\boldsymbol{z}_u^K) \mid u \in V)$
  - $\boldsymbol{Z_G} = Sum(BN(\boldsymbol{z}_u^K) \mid u \in V)$
  - $\boldsymbol{Z_G} = Max(\boldsymbol{z}_u^K \mid u \in V)$

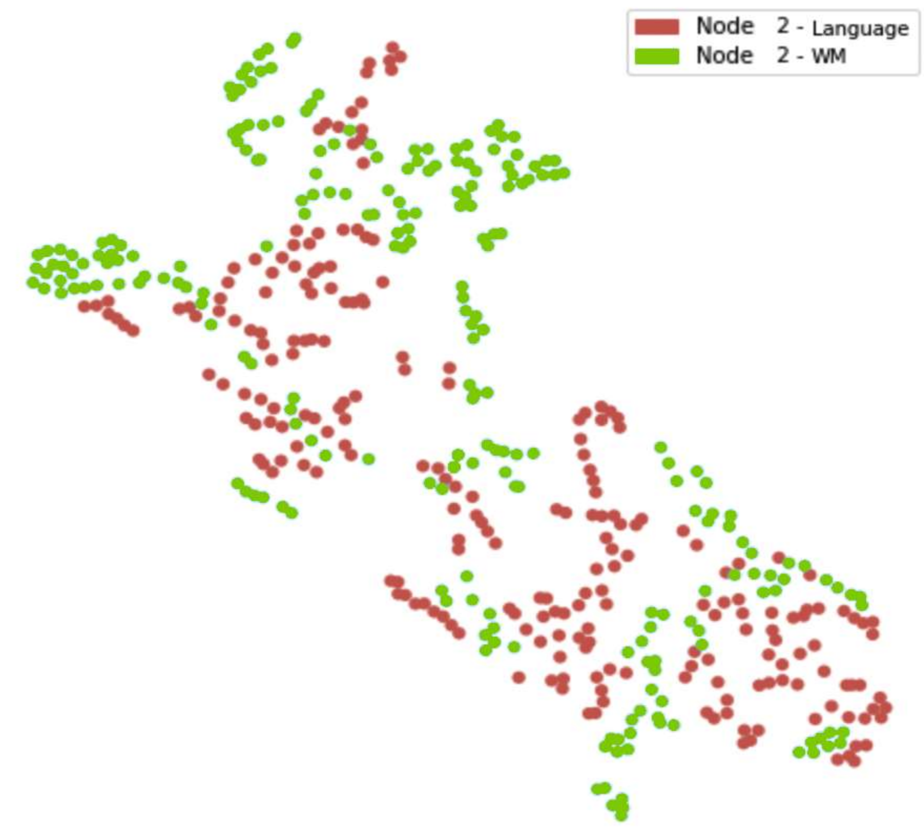


5 Observations on Subject Heterogeneities

- **T-SNE single node attention space for many subjects (M vs WM)**
  - Different tasks have varied representations in the latent space for the same node



- **T-SNE single node attention space for many subjects (WM vs L)**
  - Some tasks have similar embedding patterns across individuals
  - Verified from neuroscientific knowledge for abstract tasks
    Similar ROIs firing



6 Acknowledgements

- This poster was supported by the CS department at UCSB. I gratefully thank Prof. Yu-Xiang Wang, Prof. Lei Li and TA Xuandong Zhao for their great advice and helpful feedback.