Community-Aware Transformer for Autism Prediction in fMRI Connectome

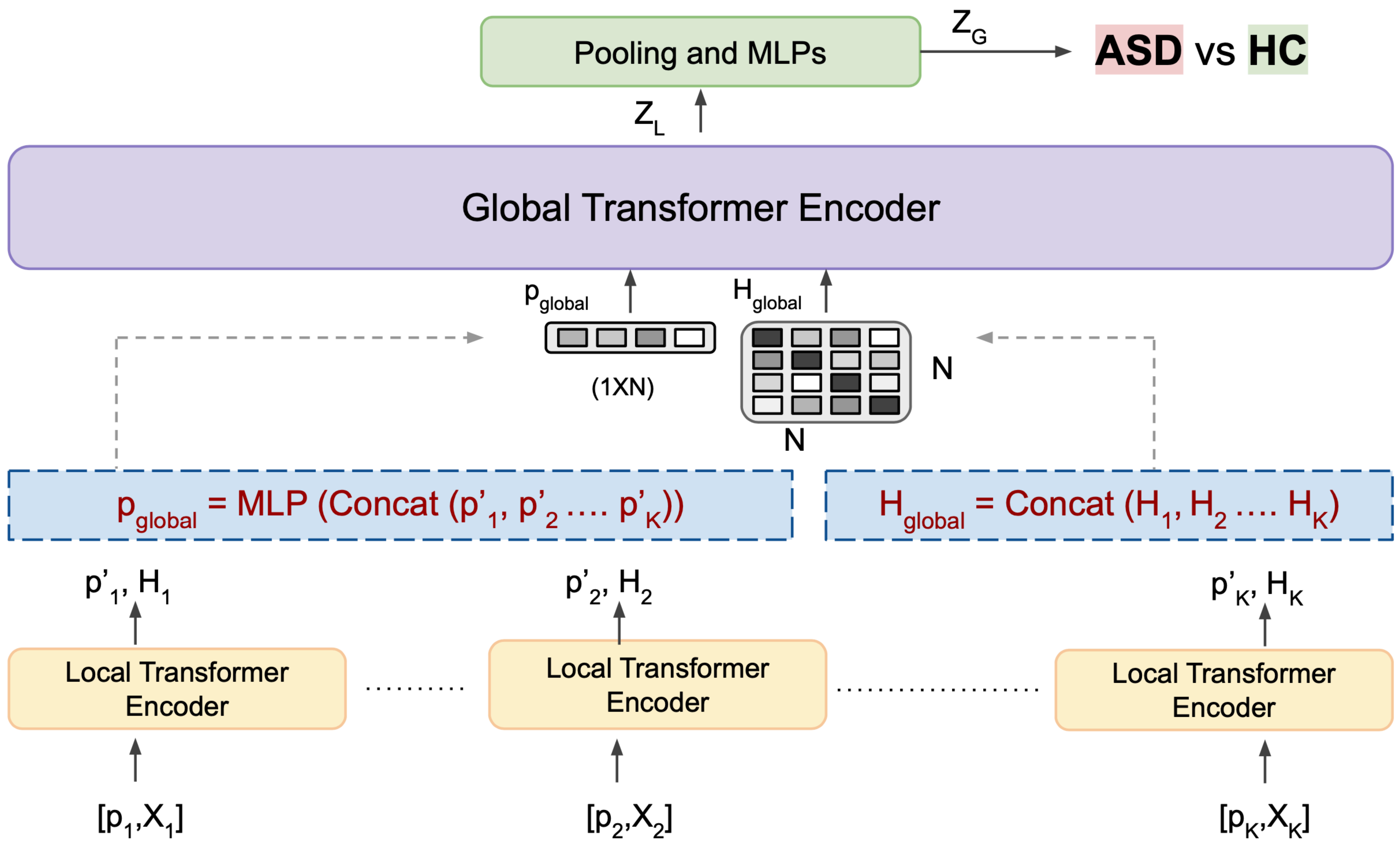
Anushree Bannadabhavi, Soojin Lee,
Wenlong Deng, and Xiaoxiao Li
The University of British Columbia

January 12, 2024

Presented by
Joanne Wardell

Overview: Comm-BrainTF

A hierarchical local-global transformer that learns intra and inter-community aware node embeddings for ASD prediction task



Connectomics

What part of the brain is responsible for ____?

Try to determine source of some disorders

Over simplification of higher cognitive function

Every brain area is part of a larger brain network

What networks are responsible for each function?

Where are the parts of the networks located? How are they connected?

https://www.o8t.com/blog/connectomics,https://en.wikipedia.org/wiki/Connectomics

Motivation - ASD Classification Tasks

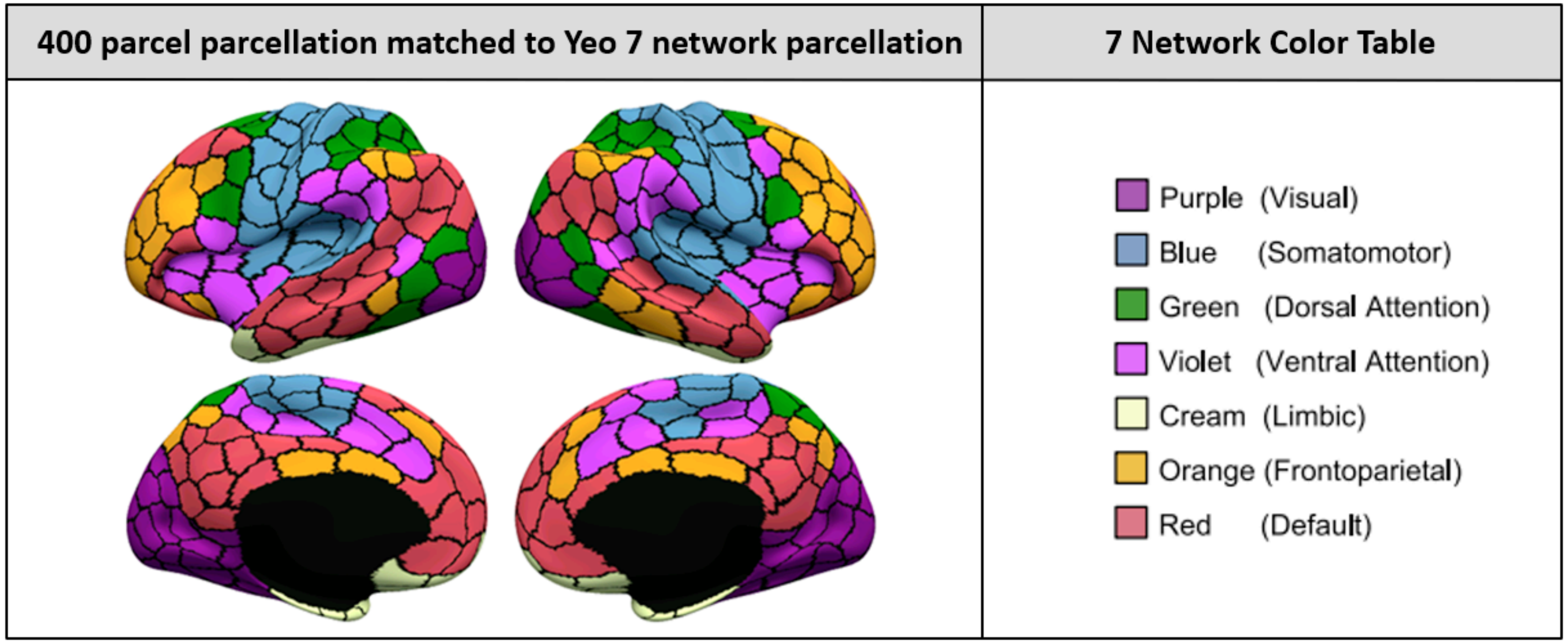
- ASD - Autism Spectrum Disorder
- Exhibit abnormalities in the default mode network (DMN)
  - Hyper and Hypo connectivity with other functional networks
- Classification tasks
  - Age group, Biological Sex, ASD or HC, etc.

ROI Communities

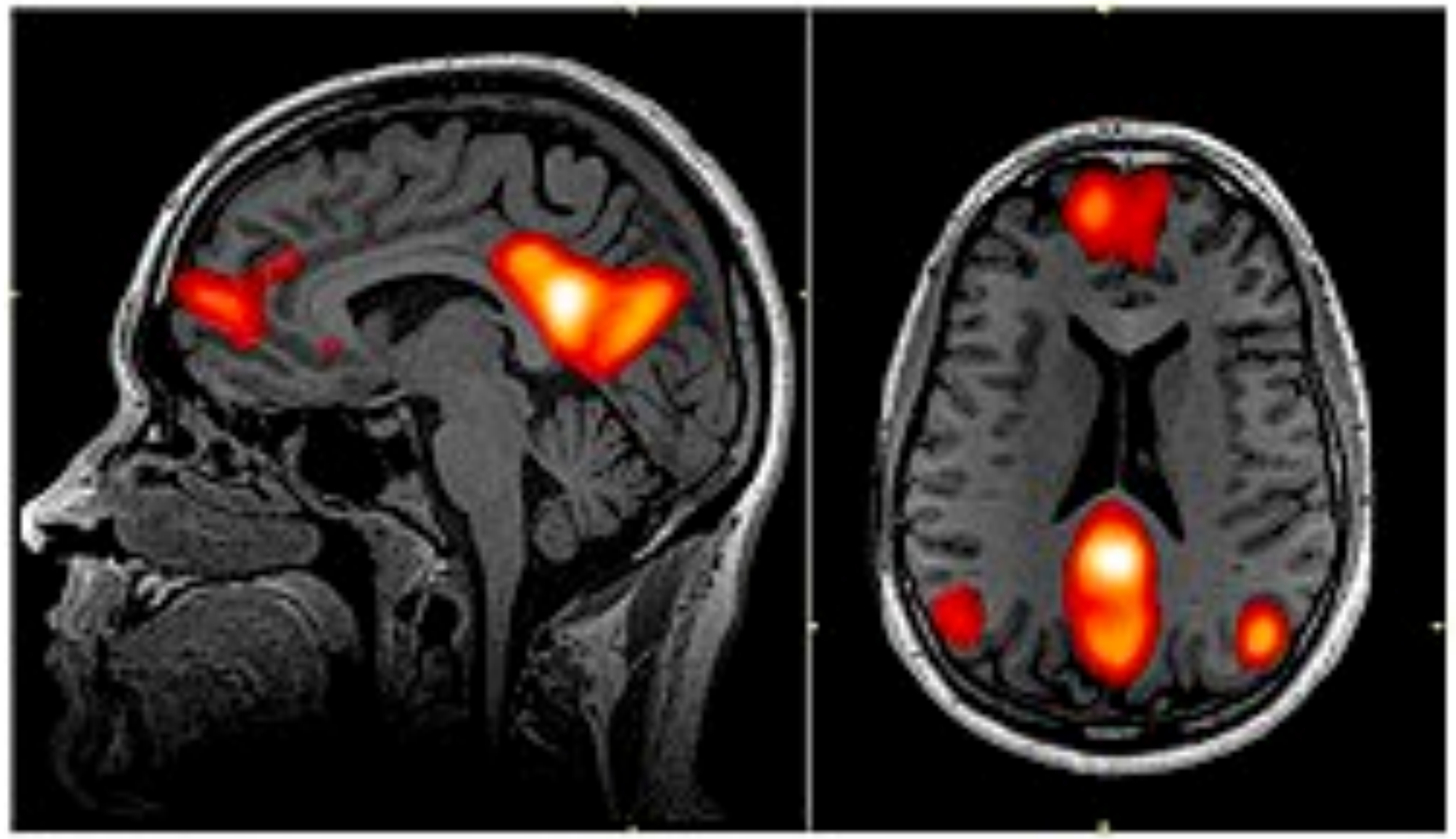
ROI - Region of interest in the brain

ROI Communities -

Groups of ROIs that perform similar functions as an integrated network



Schaefer et al. 2018 Local-Global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI



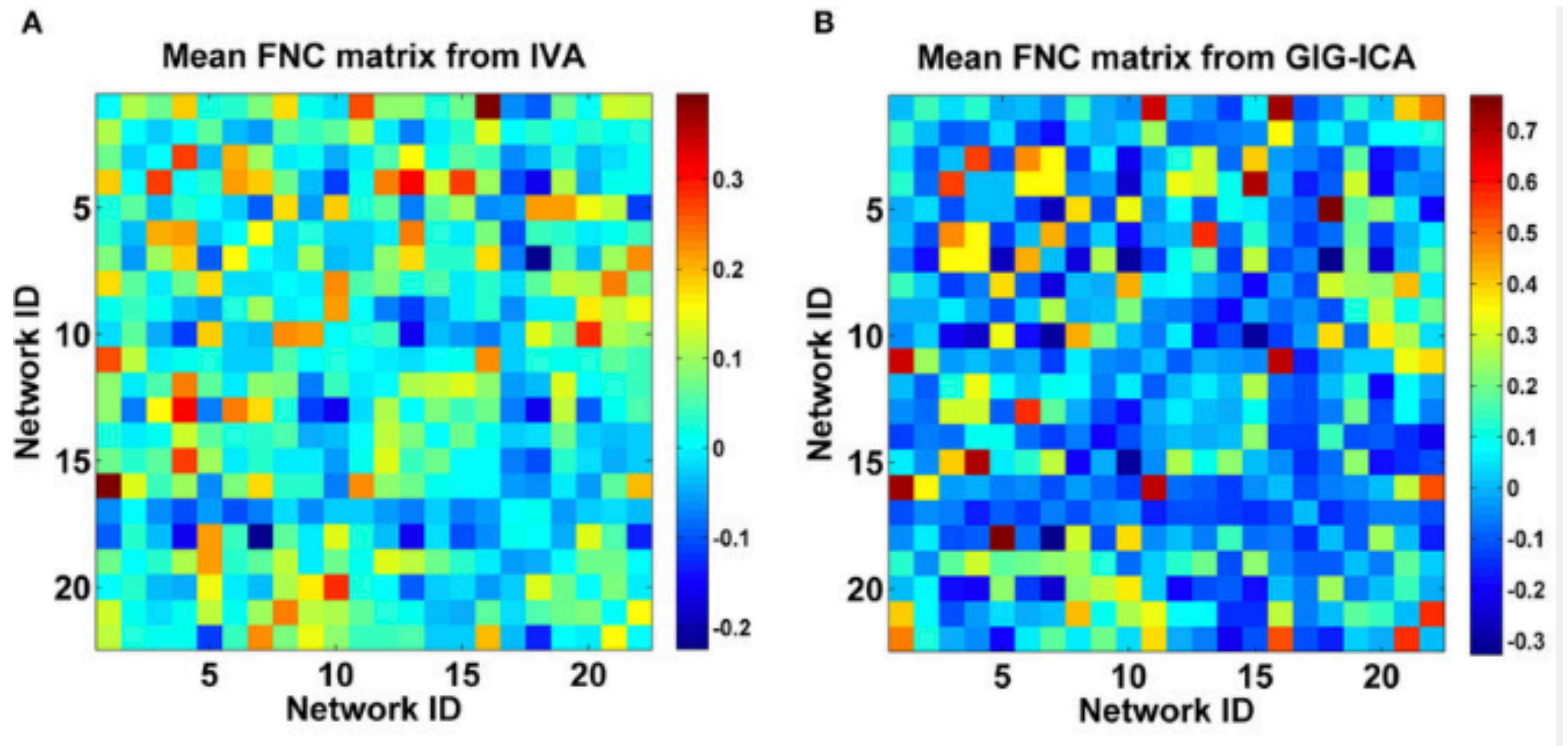
Graner et al. 2013 Functional MRI in the investigation of blast-related traumatic brain injury

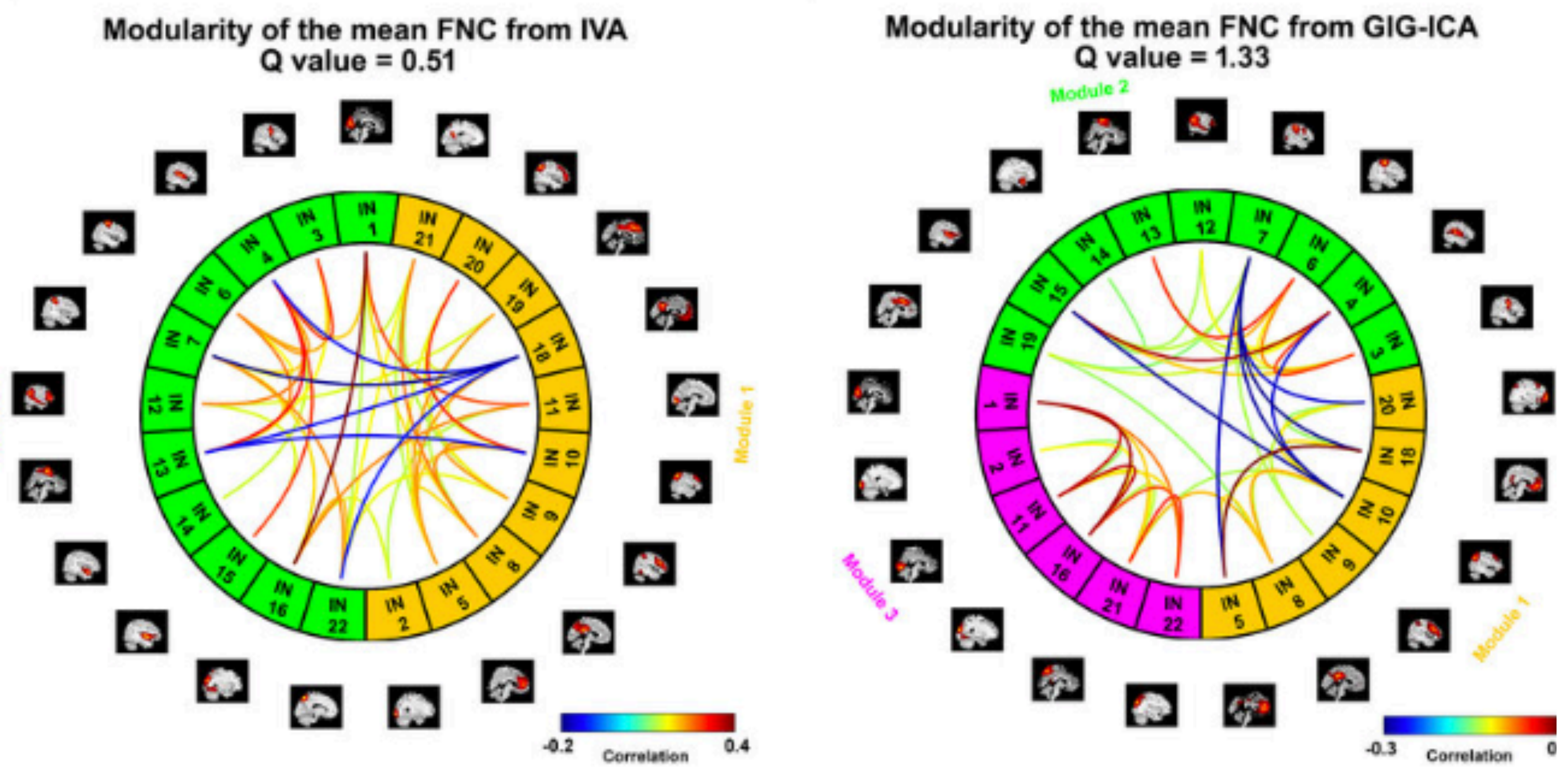
fMRI Connectome

Functional brain activity (measured from fMRI) represented as a graph

Nodes: ROIs

Edges: Relationships or Connections between ROIs





Convolutional Neural Networks

BrainNetCNN - architecture with special edge-to-edge, edge-to-node, and node-to-graph convolutional filters

Kawahara et al. 2017

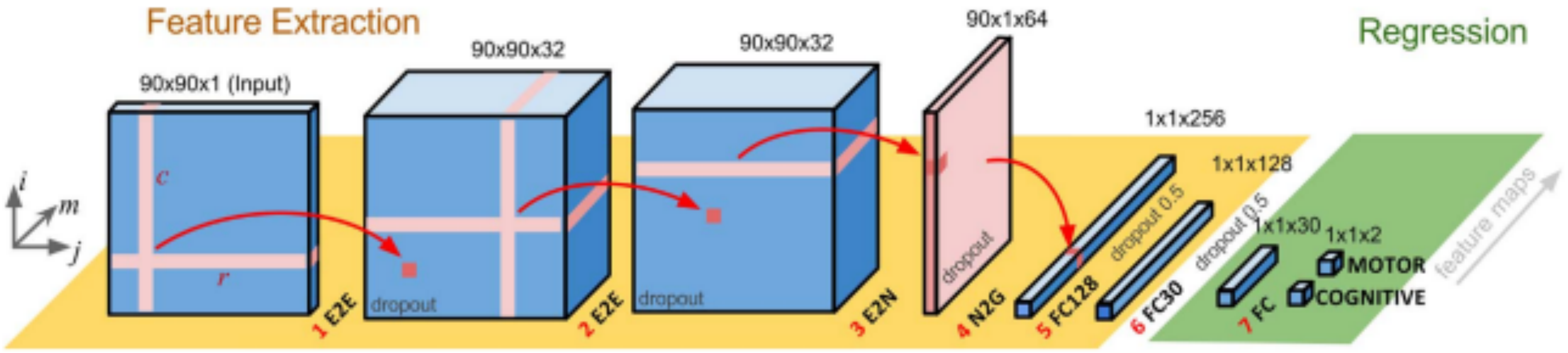


Fig. 1. Schematic representation of the BrainNetCNN architecture. Each block represents the input and/or output of the numbered filter layers. The 3rd dimension of each block (i.e., along vector m) represents the number of feature maps, M, at that stage. The brain network adjacency matrix (leftmost block) is first convolved with one or more (two in this case) E2E filters which weight edges of adjacent brain regions. The response is convolved with an E2N filter which assigns each brain region a weighted sum of its edges. The N2G assigns a single response based on all the weighted nodes. Finally, fully connected (FC) layers reduce the number of features down to two output score predictions.

Graph Neural Networks

BrainGNN - Constructs graph representation from FC matrices

X. Li, Y. Zhou, N. Dvornek et al. 2021

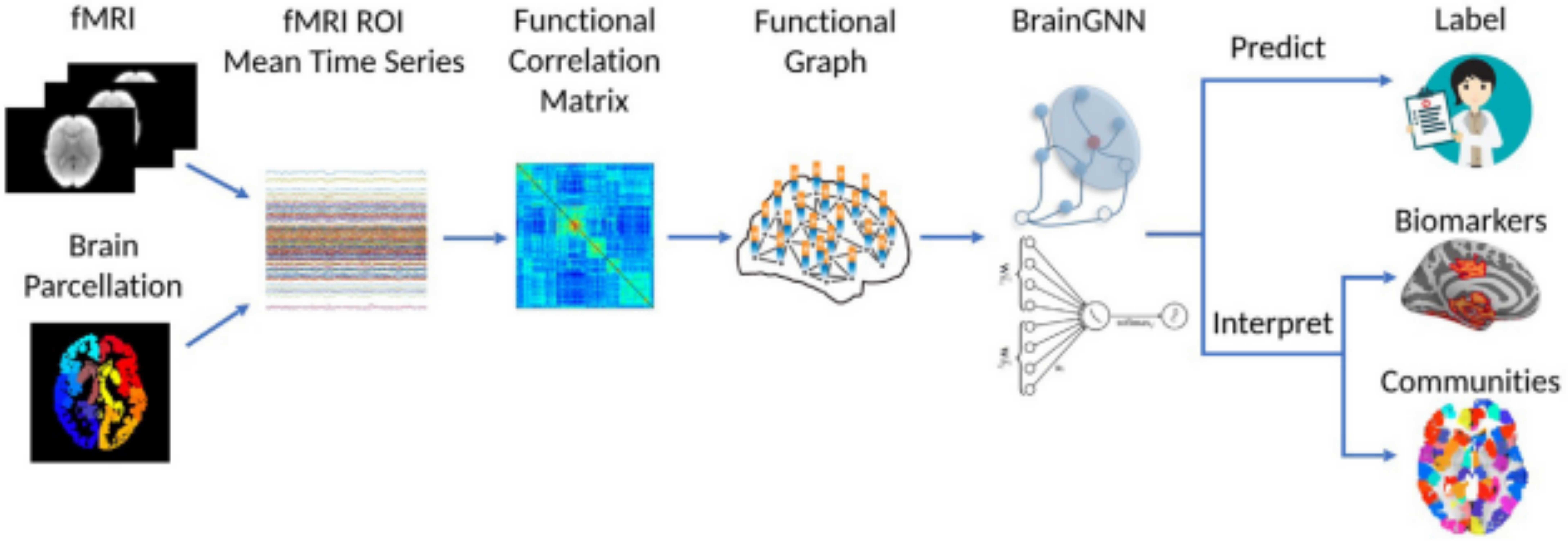
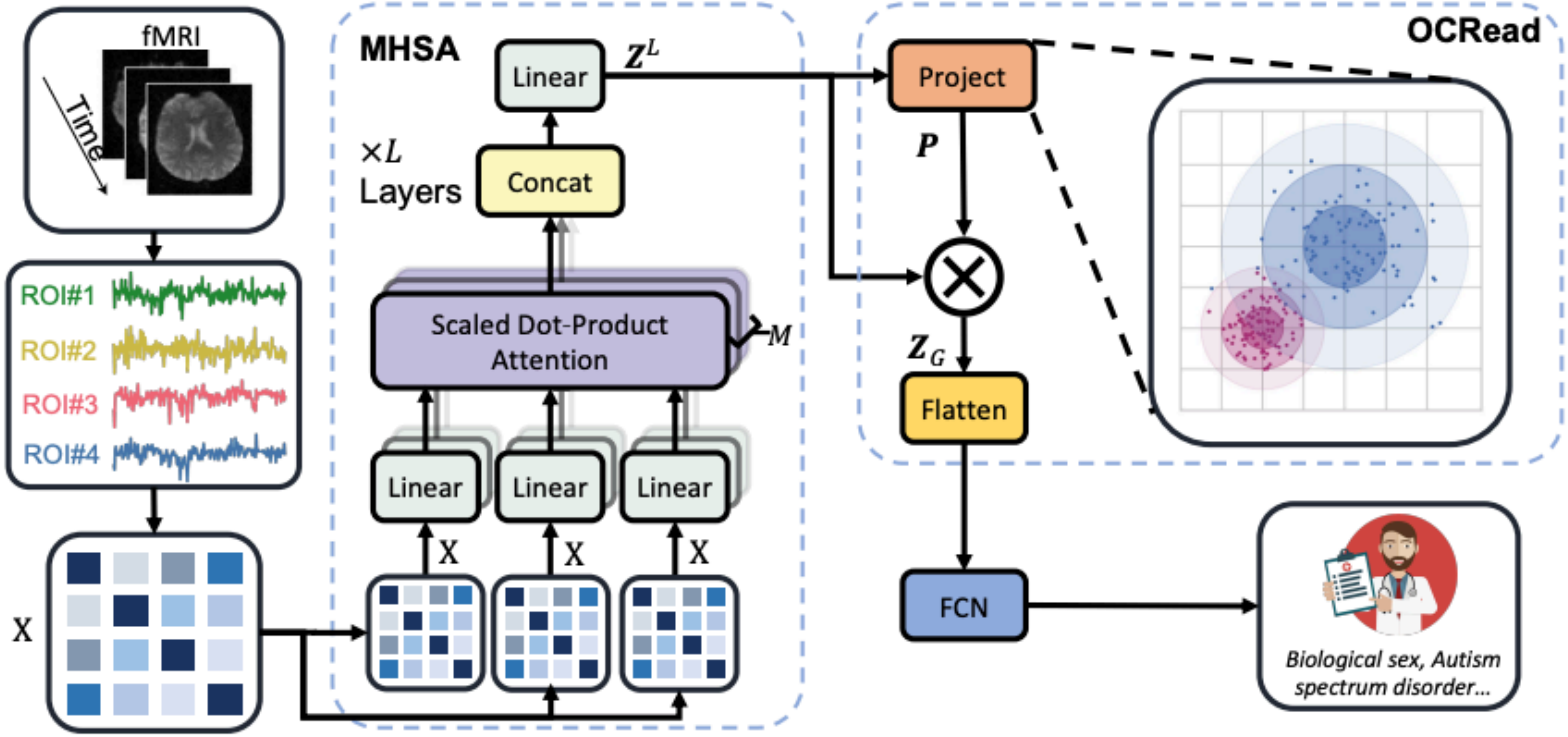


Fig. 1. The overview of the pipeline. fMRI images are parcellated by an atlas and transferred to graphs. Then, the graphs are sent to our proposed BrainGNN, which gives the prediction of specific tasks. Jointly, BrainGNN selects salient brain regions that are informative to the prediction task and clusters brain regions into prediction-related communities.

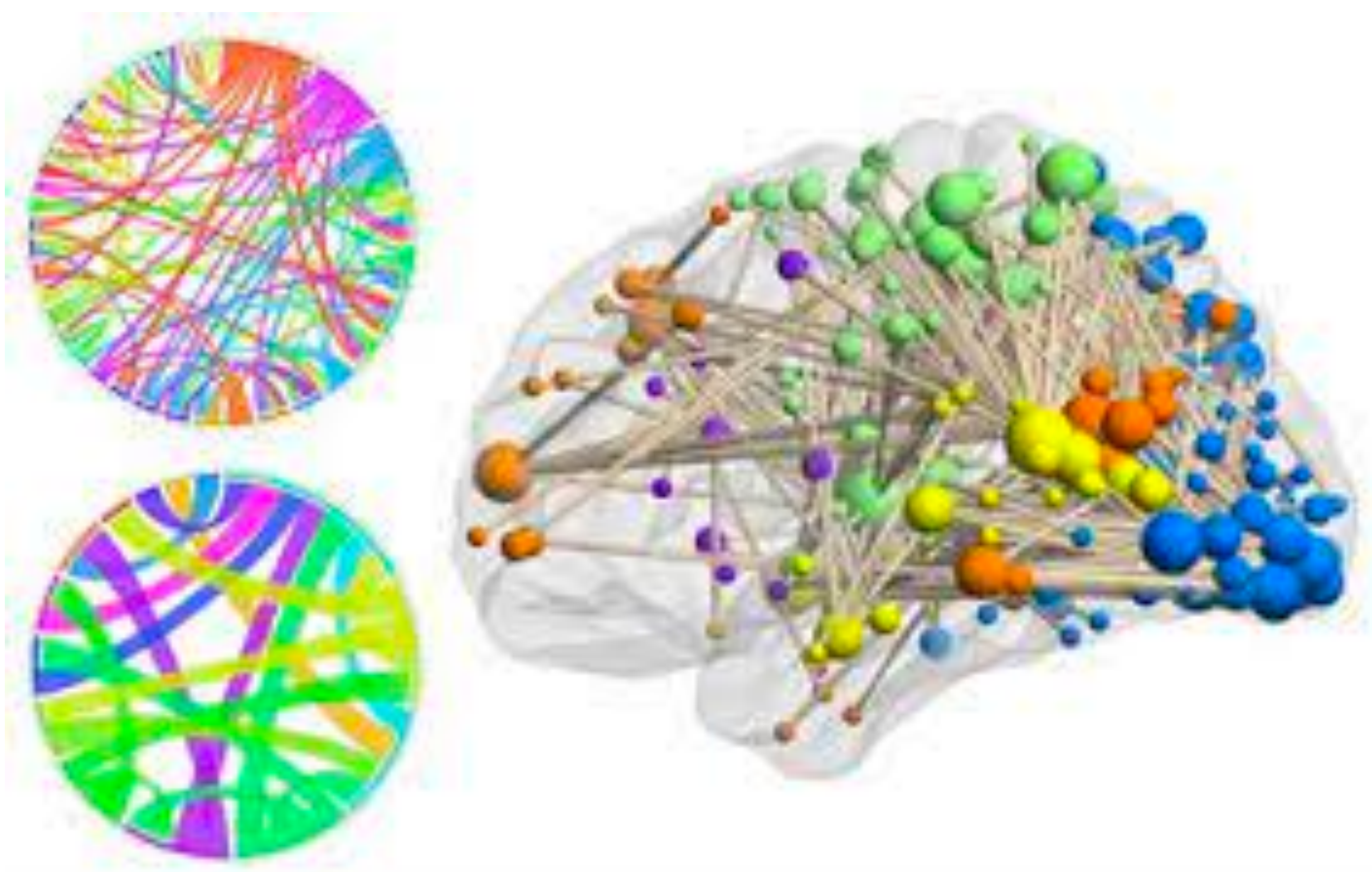
Transformers

Brain Network Transformer - learn node embeddings using a pearson correlation matrix and a readout layer that learns brain clusters

Xuan Kan et al. 2022



Human Brain Connectome is Hierarchical



- ROIs in the same community -> More similar
- ROIs across communities -> Less Similar
- ASD individuals -> hypo & hyper connectivity across networks
- Existing DL models often don't use community labels for ROIs

Pang et al. Functional connectome beyond fMRI: methodology and application, https://www.o8t.com/infinitome/whats-under-the-hood

Limitations in Previous Research

- BrainNetworkTransformer (BNT) outperforms CNN and GNN models for fMRI-based classification...

- Existing DL models -> don’t leverage community hierarchy
  - Treat all ROIs individually

Addressing Previous Limitations

Local Transformer

- Takes FC matrices as input
- Parameters shared across all communities
- Learns personalized prompt tokens

Global Transformer

- Fuses whole brain information
- Receives class tokens and node embeddings from local transformer
- Uses pooling layer to summarize the final prediction

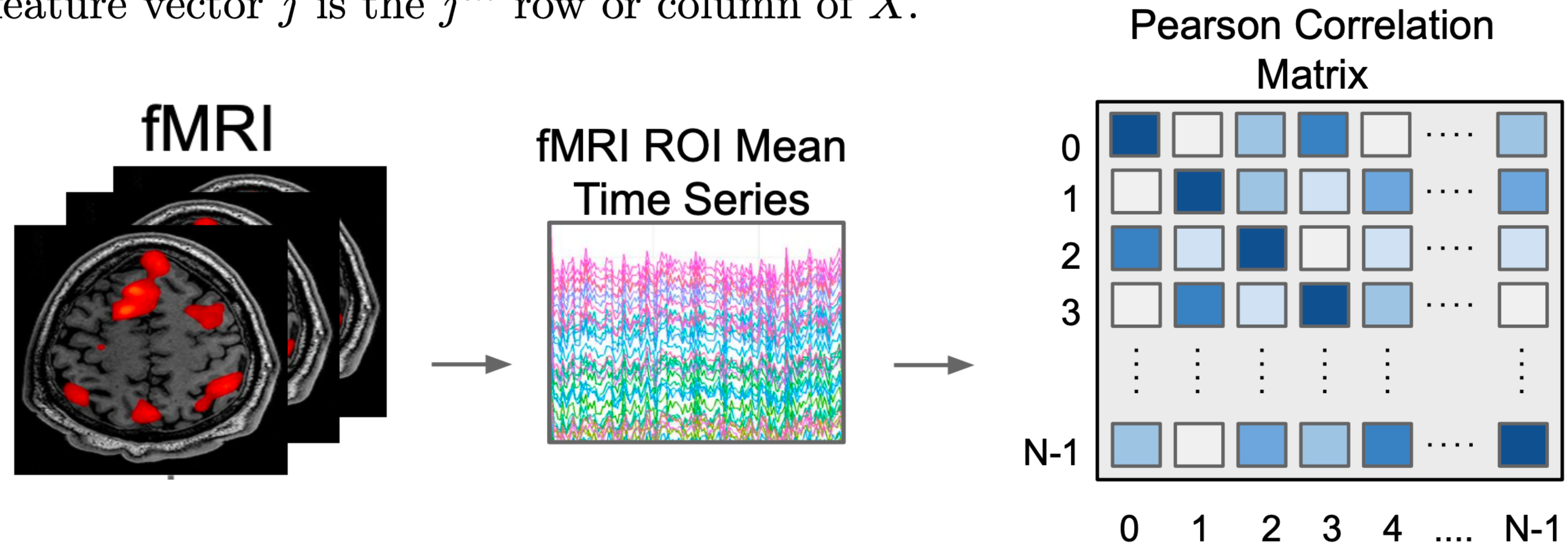
Problem Definition 1/4

1. Parcellate the brain into N ROIs based on a given atlas.

2. Construct the FC matrix using Pearson correlation.

For a brain graph with N nodes, let X denote the symmetric FC matrix. $X \in \mathbb{R}^{N \times N}$

Node feature vector j is the j^{th} row or column of X.



Problem Definition 2/4

3. Rearrange the rows and columns of the FC matrix based on community membership.

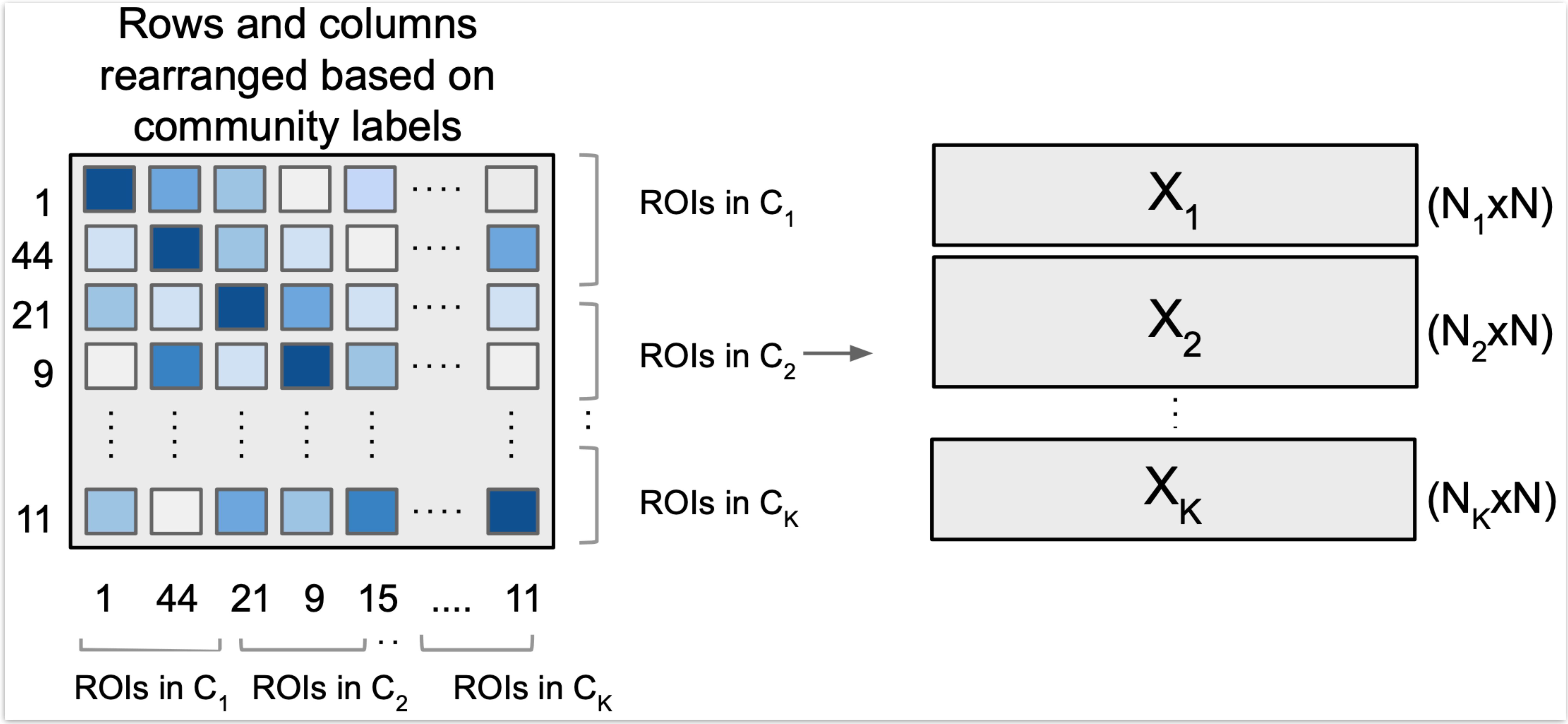
4. Extract K groups of rows or columns as input matrices $\{X_1, X_2, ..., X_K\}$.

   Let K denote the number of functional communities.

   Let k denote the community number.

Each community matrix $X_k \in \mathbb{R}^{N_k \times N}$ has N_k tokens.

Each token is N dimensional.

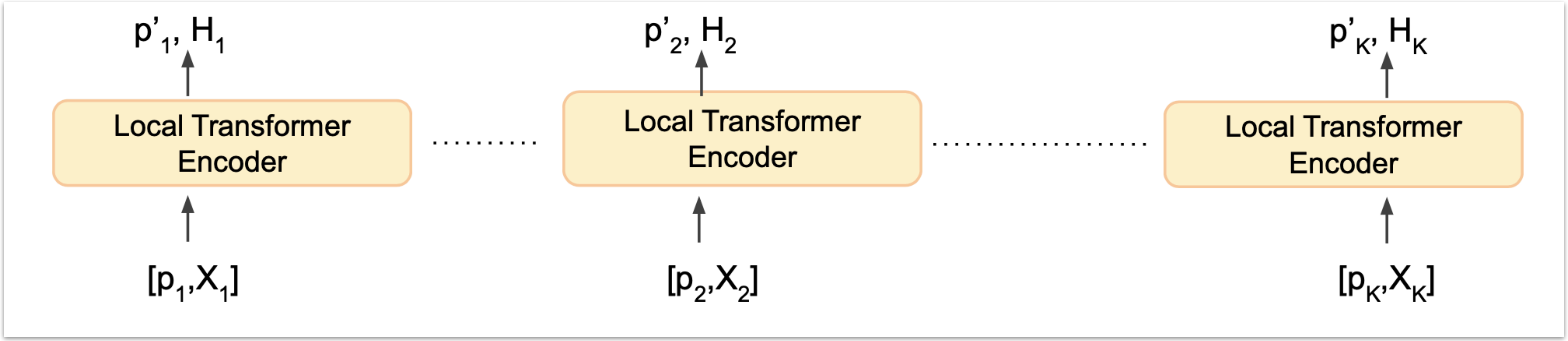


Problem Definition 3/4

5. Input the sequence X_k to the community-k transformer.

6. For each token in the sequence X_k
   transformer outputs an N-dimensional token $H_i, i \in [1, N_k]$.

So for a community matrix X_k, the output token is $H_k \in \mathbb{R}^{N_k \times N}$.

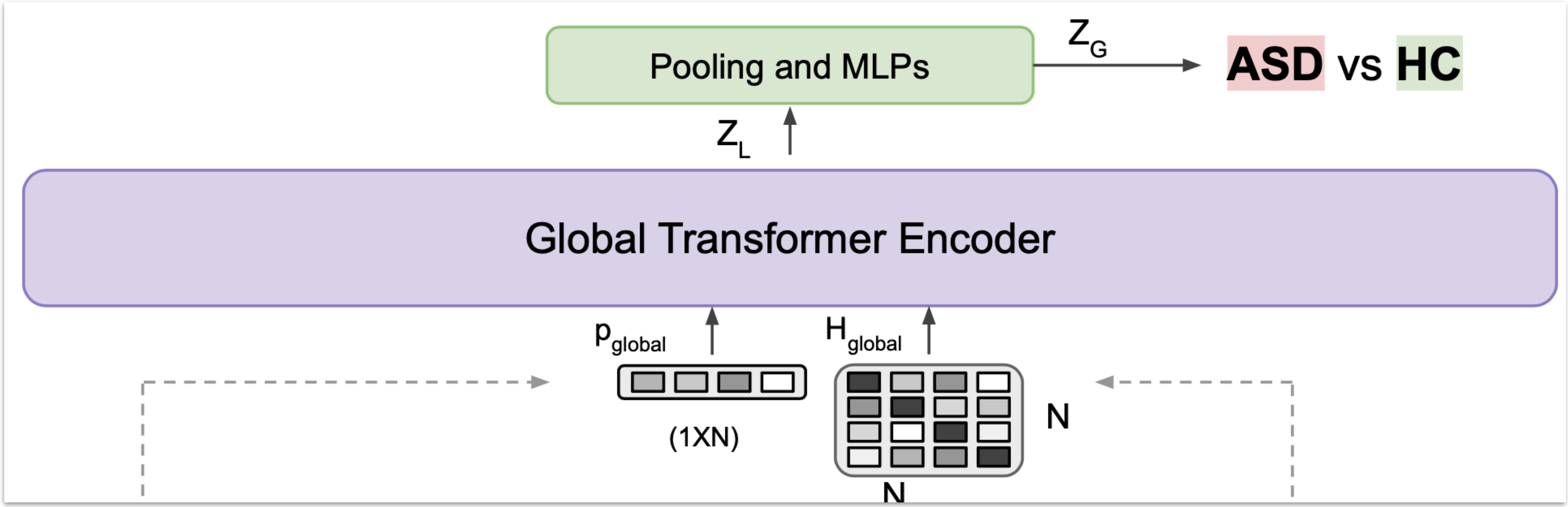


Problem Definition 4/4

7. Input the tokens for all communities to the global transformer.

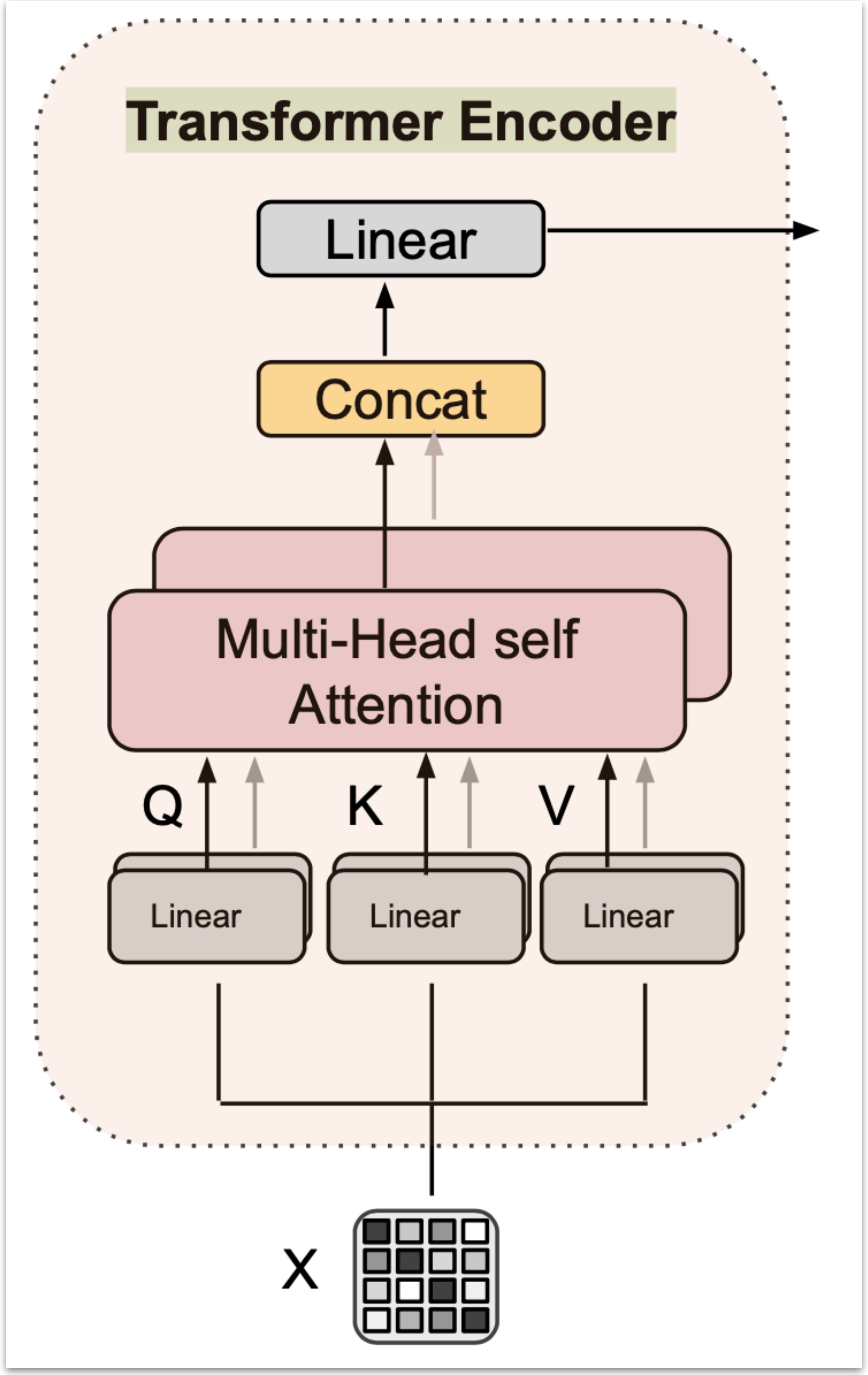
Global transformer learns an embedding $H = [H_1, \ldots, Hk] \mapsto Z_L \in \mathbb{R}^{N \times N}$.

8. Next is a pooling layer and multi-layer perceptrons (MLPs) to predict the output.



Transformer Encoder

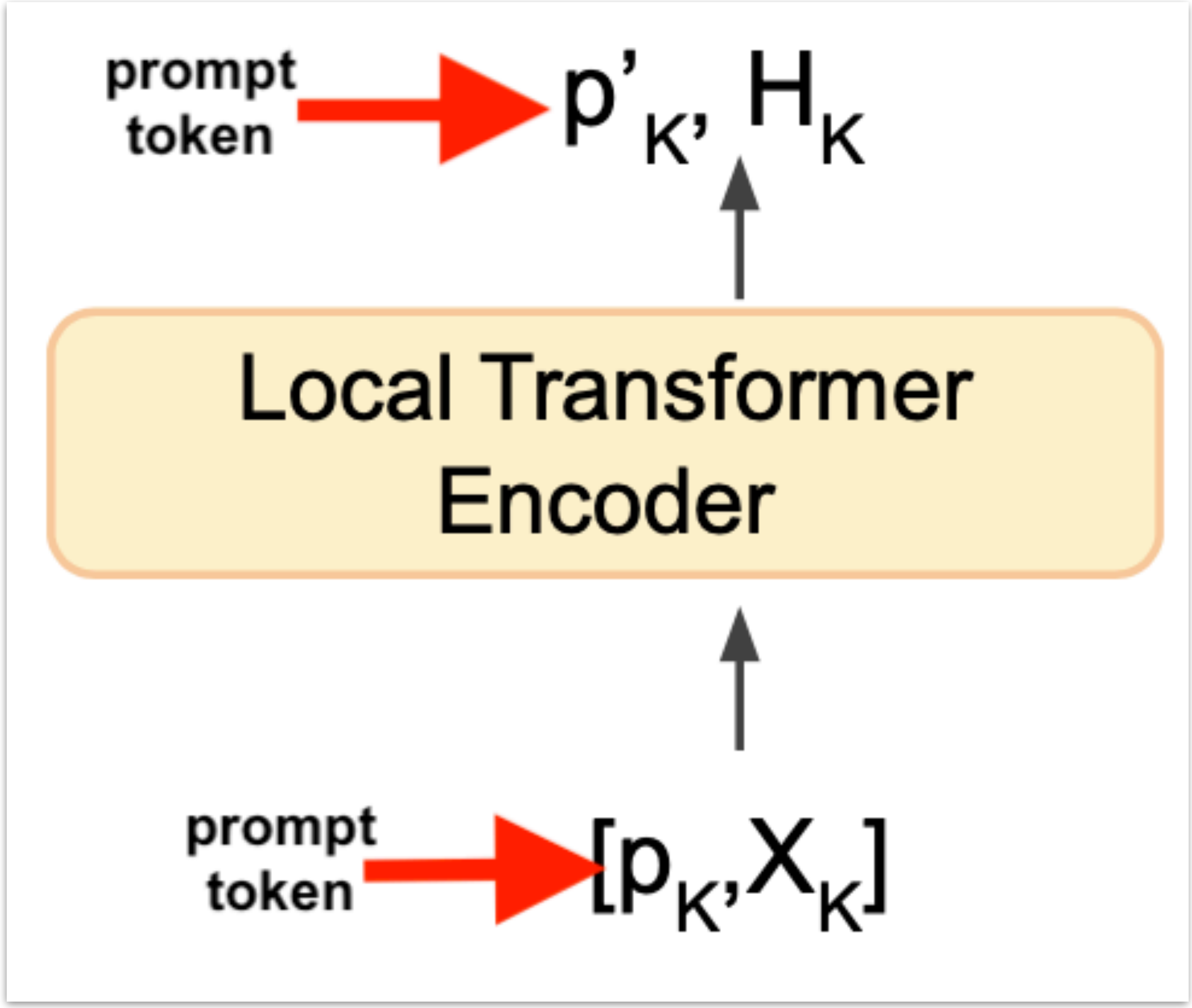
- Used for both local and global transformers
- Input: FC Matrix
    - Global or Local
- Multi-Head Attention Module
    - Captures interdependencies between nodes
- Output: Learned node feature Matrix H_i



Local Transformer

- Learns community brain networks

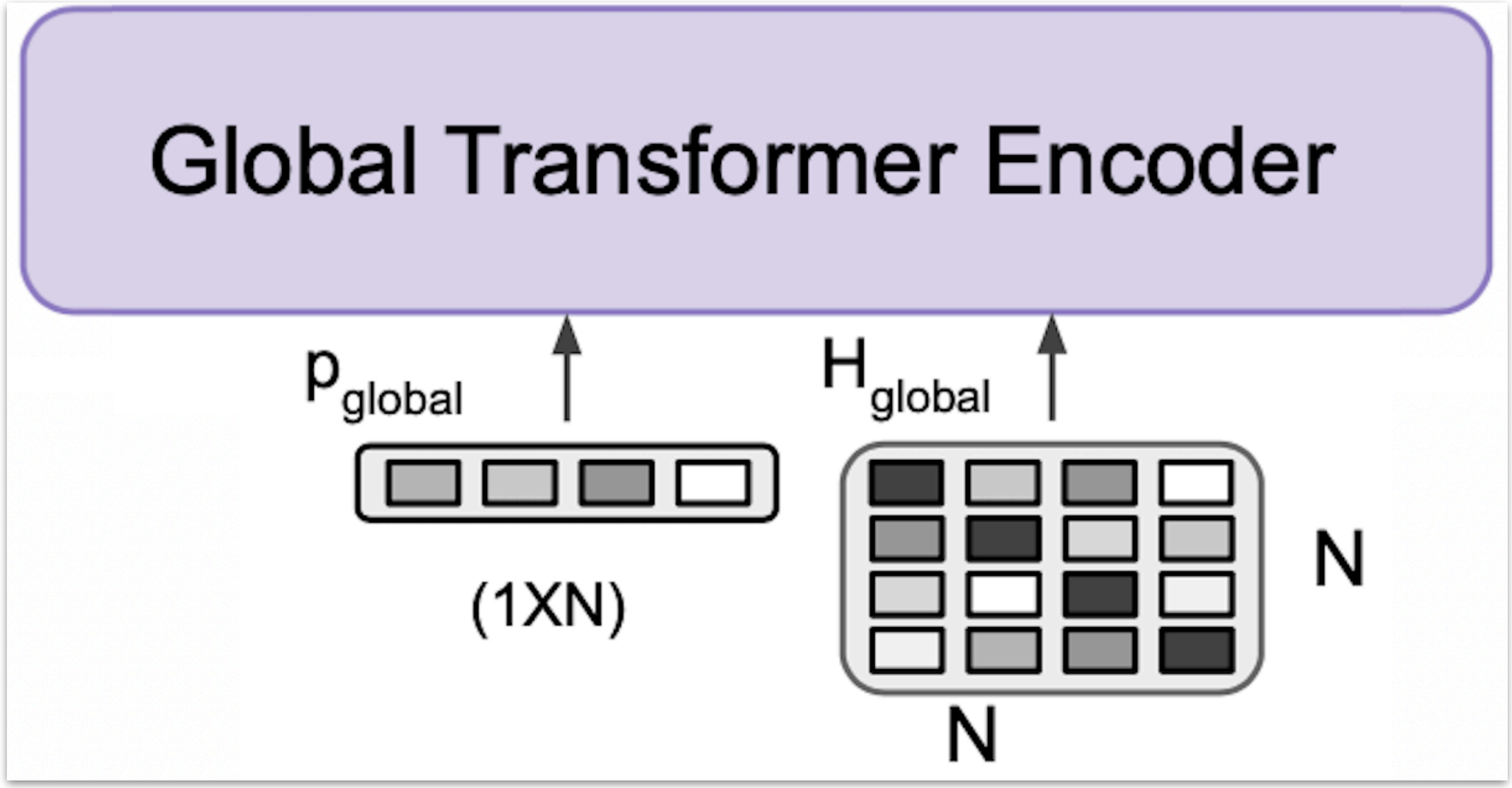
- **Prompt Tokens** for each X_k
  - $\{p_1, p_2, \ldots, p_K\}$, where $p_i \in \mathbb{R}^{1\times N}$
  - Unique, learnable
  - Learn to distinguish b/w node feature matrices of each community
  - Avoids over parameterization
  - Use same local transformer for each community



$$p_i^{'}, H_i = \text{LocalTransformer}([p_i, X_i]) \text{ where, } i \in [1, 2 \ldots K].$$

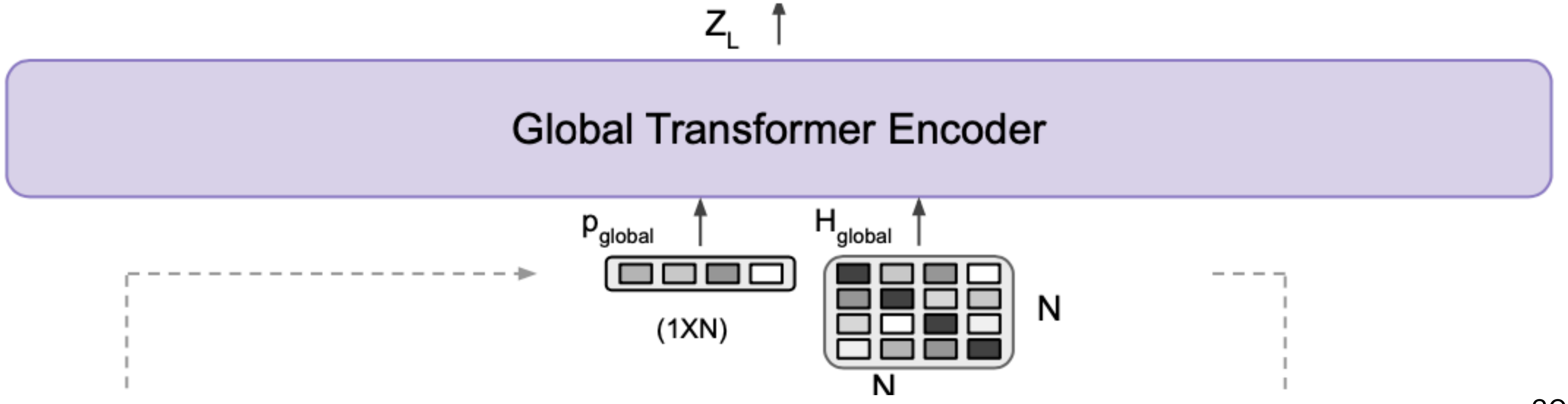
Global Transformer Input

- Learns global brain network
- Combine output of local transformer
  - Community-specific node embeddings
  - Prompt tokens
- Input:
  - $p_{global} = \text{MLP}\left(\text{Concat}\left(p_1^{'}, p_2^{'} \ldots p_K^{'}\right)\right)$
  - $H_{global} = \text{Concat}(H_1, H_2, \ldots, H_K)$



Global Transformer Output

- Runs Multi-Head attention on
  - Prompt input token
  - Learned node feature matrices
- Output:
  - Attention-enhanced node embedding matrix
  - $Z_L = \text{GlobalTransformer}([p_{global}, H_{global}])$

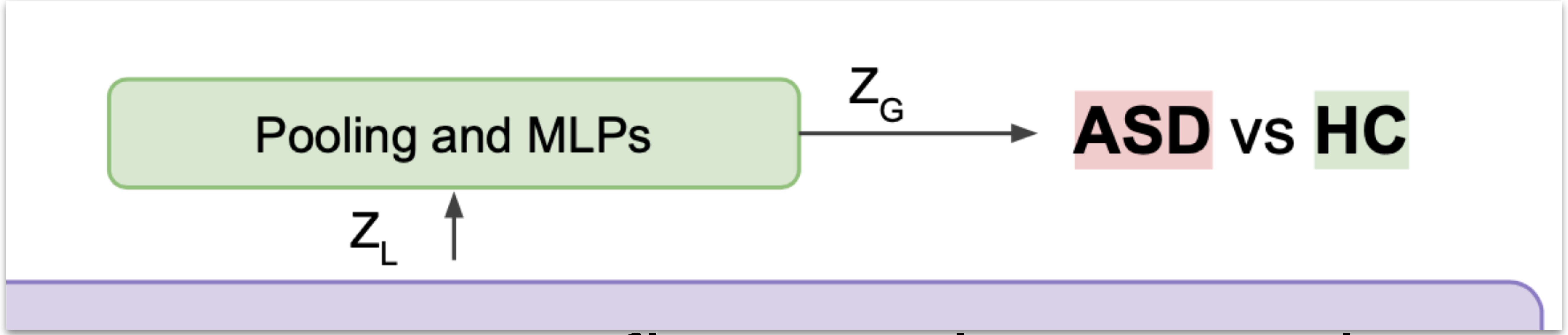


Graph Readout Layer

- Aggregates global embeddings
- Produces high level representation of brain graph

- OCRead Layer
  - Initializes orthonormal cluster centers $E \in \mathbb{R}^{K \times N}$
  - Softly assigns nodes to centers

- Graph Embedding Z_G
  - $Z_G = A^\top Z^L$
  - $A \in \mathbb{R}^{K \times N}$
  - A - Computed by OCRead



Z_G flattened & passed to MLP for graph level prediction

Experiments - Dataset

- ABIDE Dataset
  - resting-state functional MRI (rs-fMRI)
  - 17 international sites
  - Parcellated by Craddock 200 atlas
  - ROIs belong to either of the 8 functional communities
  - 1009 subjects, 51.14% ASD

Experiments - Settings

- All models implemented in PyTorch
- Trained on NVIDIA V100 with 8GB memory
- # attention heads = # communities
  - For local and global
- Adam optimizer
  - Learning rate 10e-4
  - Weight decay 10e-4
- train/validation/test ratio: 70:10:20
- Batch size: 64
- Epochs: 50

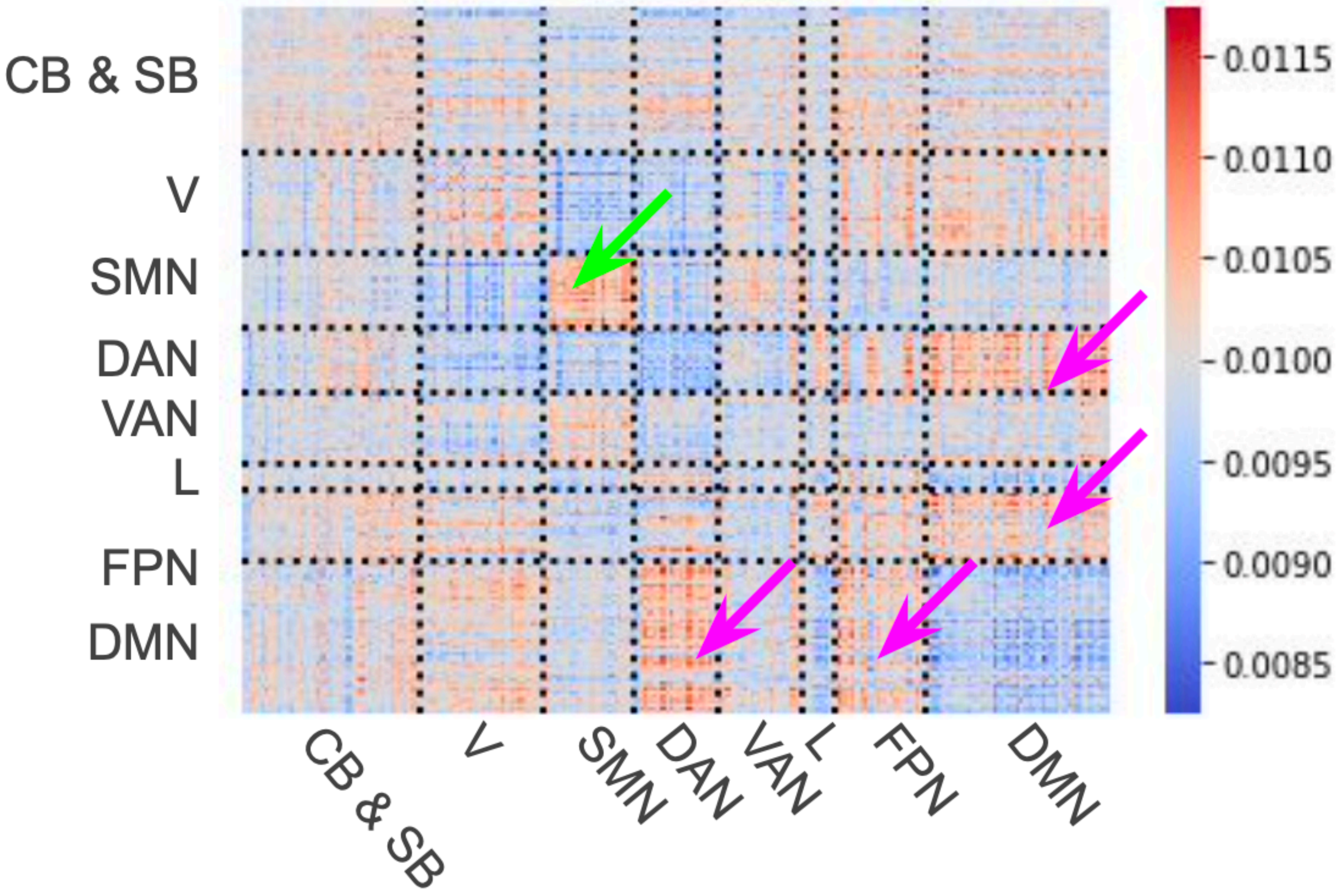
Results

Table 1. Performance comparison with baselines (Mean ± standard deviation)

Model	AUROC	Accuracy	Sensitivity	Specificity
BrainNetTF[15]	78.3 ± 4.4	68.1 ± 3.1	78.1 ± 10.0	58.9 ± 12.0
BrainNetCNN[17]	74.1 ± 5.1	67.5 ± 3.1	65.3 ± 4.3	**69.6 ± 4.1**
FBNETGNN[16]	72.5 ± 8.3	64.9 ± 8.9	60.9 ± 11.3	67.5 ± 13.1
Com-BrainTF	**79.6 ± 3.8**	**72.5 ± 4.4**	**80.1 ± 5.8**	65.7 ± 6.4

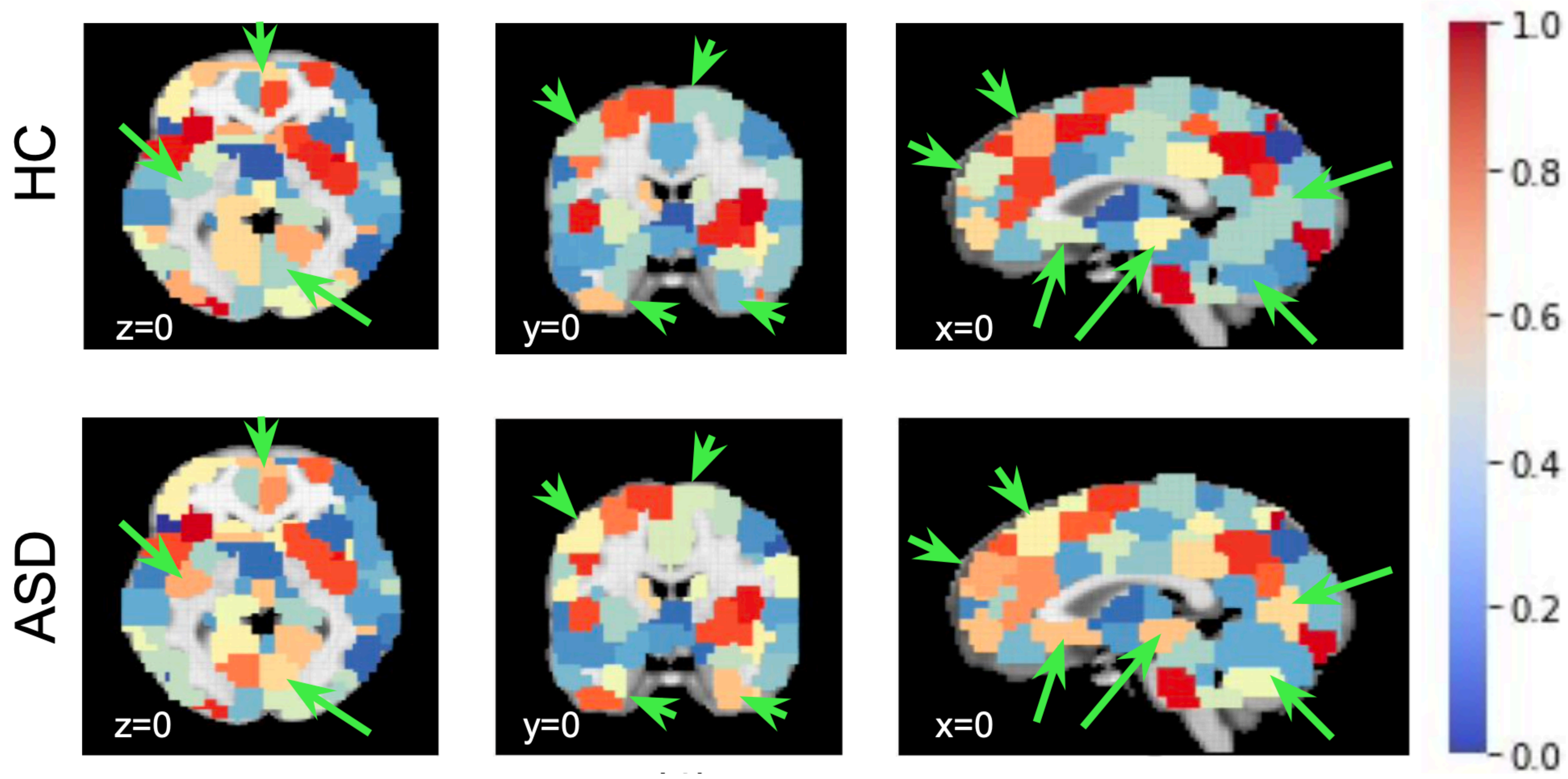
Global Attn Matrix

- Global Attention Matrix of Comm-BrainTF (Avg'd)
  - Highlights prominent communities that influence ASD prediction
- SMN Region (green arrow)
  - Attention scores high
  - Important for ASD pred.
- DMN, DAN, FPN (pink arrows)
  - Attn high
  - Conn. b/w networks important



Prompt Attn Vector

- First row of attn matrix
  - Attn of ROIs to prompt
- Generated using avg'd prompt vectors over correctly classified data
  - Red->Blue is attn scores
- Differences in HC vs ASD
  - Regions important to ASD classification
  - DMN and SMN



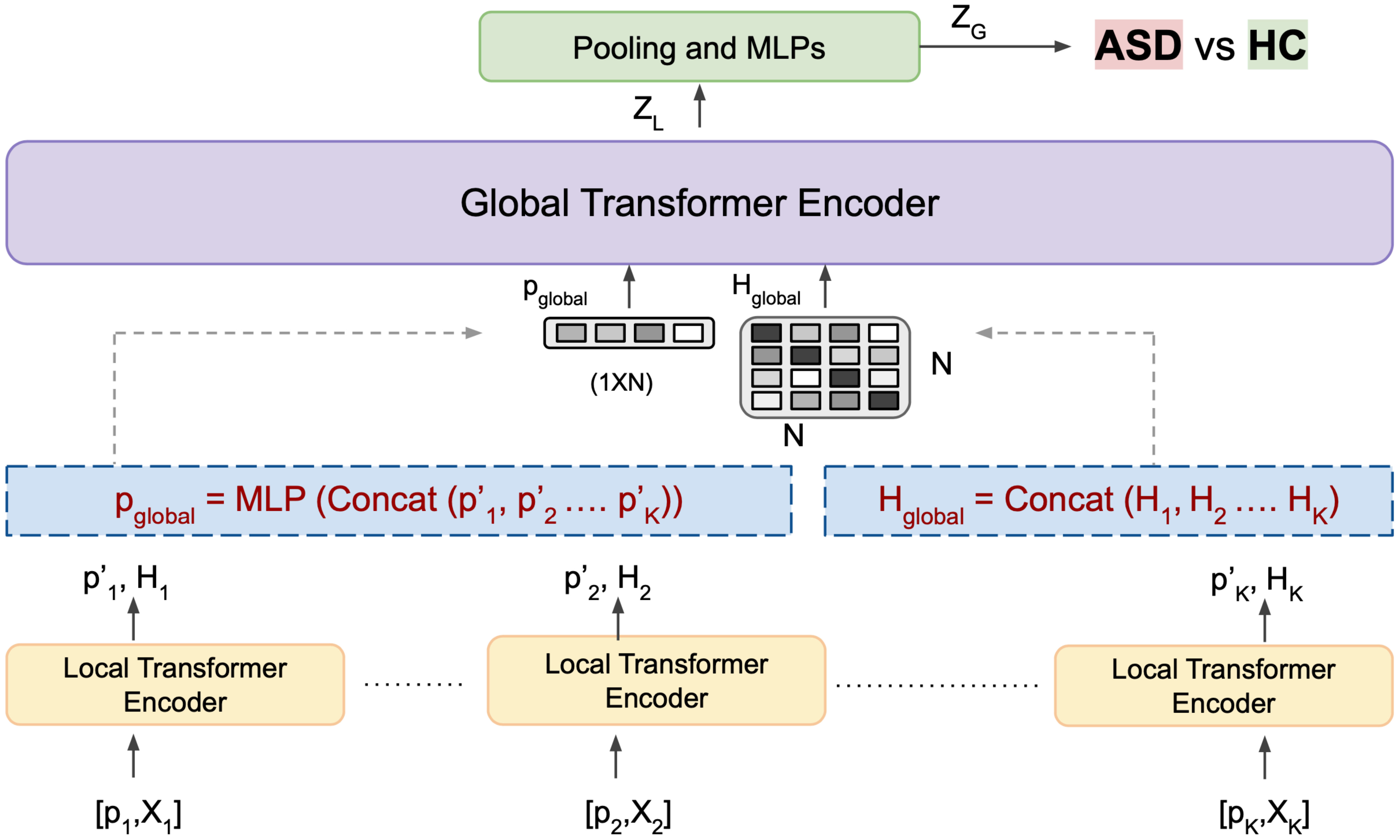
Summary

✵ They proposed a novel local-global hier. transformer Comm-BrainTF

- Utilizes ROI level info
- Utilizes community level info

✵ Avoided over-parameterization

- Community prompt tokens
- Local parameters shared



✵ Experiments show efficacy

- Captures functional community patterns
- Specific to ASD vs HC classification