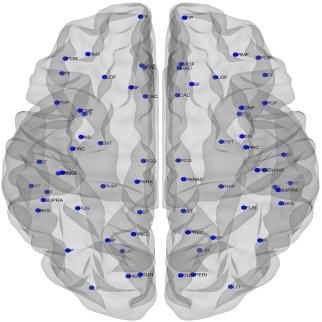
Learning Similarity-Preserving Representations of Brain Structure-Function Coupling

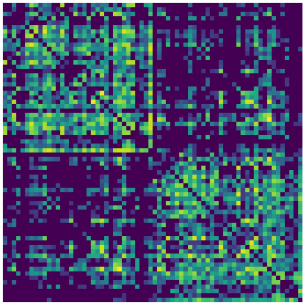
Gonzalo Mateos

Dept. of ECE and Goergen Institute for Data Science
University of Rochester
gmateosb@ece.rochester.edu
http://www.hajim.rochester.edu/ece/sites/gmateos

Co-author: Yang Li
Acknowledgment: NSF Awards CCF-1750428, CCF-1934962 and ECCS-1809356

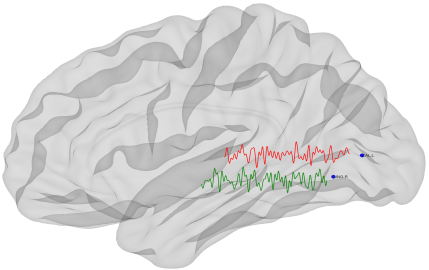
EUSIPCO 2022, Belgrade, Serbia
September 1, 2022

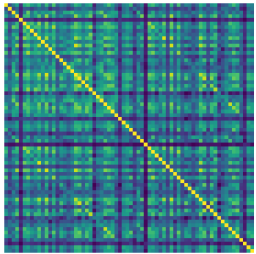




- Structural connectivity (SC) How is the brain wired?
  - $\Rightarrow$ Anatomical tracts connecting brain regions
  - $\Rightarrow$ Sparse with fewer connections between hemispheres

- Functional connectivity (FC) How the brain functions?
  - Statistical correlation between neural signals in different regions
  - Blood oxygen-level dependent (BOLD) signals from fMRI





- FC depends on anatomical connections
- Strong FC exists between regions with weak or none SC connections
  - FC correlates SC at an aggregate level
  - FC between brain regions related with indirect SC connections

Key problem: deciphering the relationship between SC and FC

- SC-FC correlation, graph feature comparisons [Hagmann et al'14]
- Simulations of nonlinear cortical activity models [Honey et al'09]
- Diffusion-based parametric inverse problem [Abdelnour et al'14]

Contributions: novel investigation of the SC-FC relationship

- Machine learning for network data
  - $\Rightarrow$ Reconstruct FC from SC as a regression problem
  - $\Rightarrow$ Learn lower-dimensional graph representations for classification
- Siamese network framework design
  - $\Rightarrow$ Each input is a pair of graphs $\rightarrow$ enlarge data size
  - $\Rightarrow$ Increase robustness to integrate various prior information
  - $\Rightarrow$ Similarity-preserving graph embeddings

- Learn a mapping from a discrete graph to a continuous domain
- Given $G(\mathcal{V}, \mathcal{E})$ with weighted adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$
  $\Rightarrow$ N brain regions, $A_{ij} :=$ anatomical connection strengths
- **Goal**: learn low dimensional vector representation, node embeddings
  $\Rightarrow$ Capture information of the node and its neighbors
- Approach: Graph signal processing (GSP) and graph convolution
- Extensions to *embed the whole graph*, i.e. graph embeddings
  $\Rightarrow$ Graph clustering/classification
- Potential of GRL for neuroimaging data analysis rather unexplored
  $\Rightarrow$ Inductive GRL for brain network analysis among population

- Graph $G(\mathcal{V}, \mathcal{E})$ with $N = |\mathcal{V}|$ nodes $\Rightarrow$ $\mathbf{A}$ and $\mathbf{L} = \mathbf{D} - \mathbf{A}$
  - $\Rightarrow$ $\mathbf{D}$: diagonal degree matrix
  - $\Rightarrow$ Symmetric $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$
- Graph signals represented as vectors $\mathbf{x} \in \mathbb{R}^N$
- Graph Fourier Transform (GFT): $\hat{\mathbf{x}} = \mathbf{U}^T\mathbf{x}$
- Generalize machine learning models for network data
  - $\Rightarrow$ GSP to define convolutions on graphs [Ortega et al'19]

- Graph convolution (filter) with frequency response $\hat{\mathbf{H}} = \mathrm{diag}(\hat{\mathbf{h}})$

$$\mathbf{y} = \mathbf{H}\mathbf{x} = \mathbf{U}\hat{\mathbf{H}}\mathbf{U}^T\mathbf{x}$$

- ChebNet: Chebyshev polynomials of $\mathbf{\Lambda}$ [Defferrard et al'16]
- First-order approximation of ChebNet [Kipf-Welling'17]

$$\mathbf{y} = \mathbf{H}\mathbf{x} = \theta(\mathbf{I}_N + \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2})\mathbf{x}$$

- Compact rule for per-layer filtering update

$$\tilde{\mathbf{X}} \leftarrow \tilde{\mathbf{A}}\mathbf{X}\mathbf{\Theta}$$

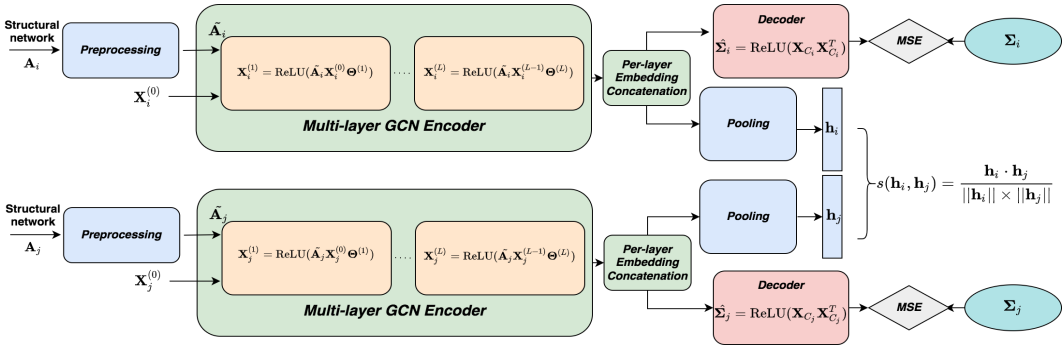
  - $\tilde{\mathbf{A}} = \mathbf{I}_N + \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$
  - $\mathbf{X}$: set of multiple observations of the graph signal $\mathbf{x}$
  - $\tilde{\mathbf{X}}$ integrates nodal attributes in $\mathbf{X}$ and topology in $\tilde{\mathbf{A}}$

- GCN: aggregate information from multiple hops within G

  $\Rightarrow$ Stack convolutional layers with pointwise non-linear activation

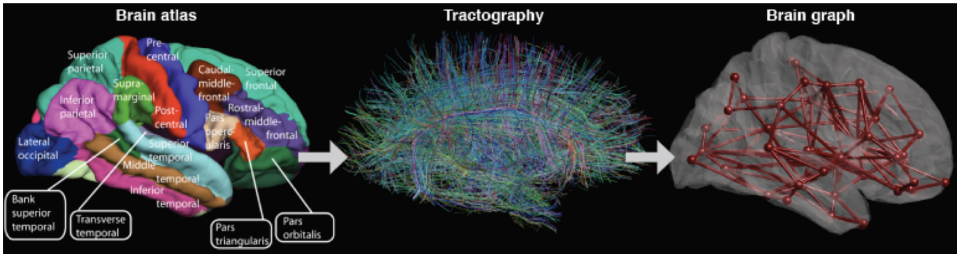
  $\Rightarrow$ Capture *indirect interactions* across the network

- Goal: summarize SC-FC relationship by simultaneously learning
  - Node embeddings $\mathbf{X}_C$ to reconstruct FC Σ from input SC networks
    $\Rightarrow \mathbf{X}_C$ captures SC-FC relationship
  - Graph embeddings for graph classification
    $\Rightarrow$ Approximately preserve similarity input graphs



- 412 subjects from Human Connectome Project (HCP)
  - Two classes: 191 non-drinkers, 221 heavy drinkers
- Preprocessed SC network **A** from diffusion MRI

  $\Rightarrow$ Fiber counts between $N = 68$ cortical surface regions



- Preprocessed FC network from functional MRI

  $\Rightarrow$ Blood oxygen-level dependent (BOLD) signals

  $\Rightarrow$ Estimated FC $\Leftrightarrow$ Pearson correlation between BOLD signals

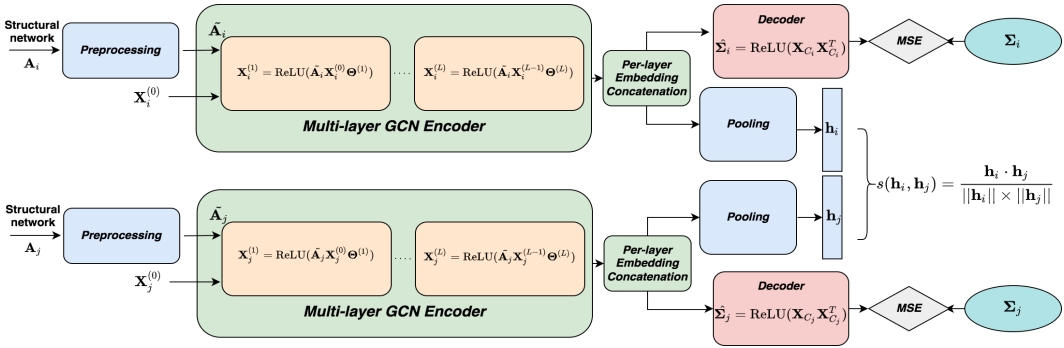
  $\Rightarrow$ Discard negative correlation and keep only positive connections
- One-hot encoding as the signal on each graph node $(\mathbf{X} = \mathbf{I}_N)$

- Input SC network $\mathbf{A} \in \mathbb{R}^{N \times N}$, $N = 68$ regions from brain atlas
  - $\Rightarrow$ Edge weights represent SC between brain regions
  - $\Rightarrow$ Preprocessing: $\tilde{\mathbf{A}} := \hat{\mathbf{D}}^{-1/2}\hat{\mathbf{A}}\hat{\mathbf{D}}^{-1/2}$, $\hat{\mathbf{A}} = \mathbf{I}_N + \mathbf{A}$
- Learn vertex representations (i.e., embeddings) that capture
  - (i) Nodal attributes, e.g. intrinsic properties of brain regions
  - (ii) Graph topology information, e.g. regional connection strengths
- First GCN layer of the encoder to learn node embeddings

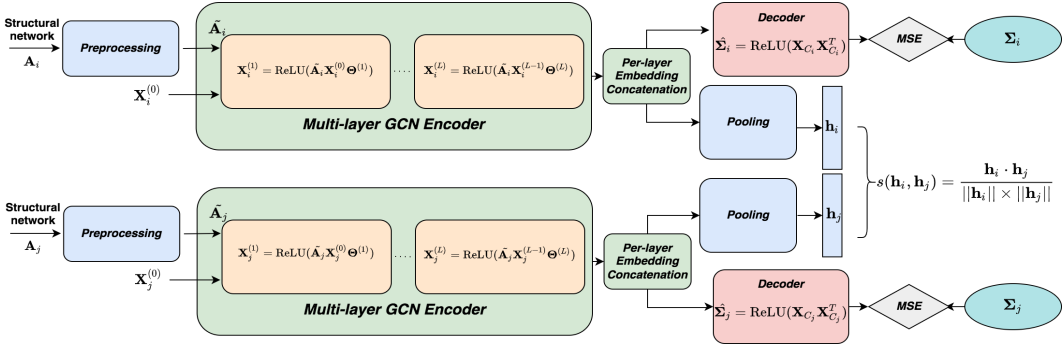
$$\mathbf{X}^{(1)} = \mathrm{ReLU}(\tilde{\mathbf{A}}\mathbf{X}^{(0)}\mathbf{\Theta}^{(1)}) \in \mathbb{R}^{N \times F_1}$$

  - $\mathbf{X}^{(0)} \in \mathbb{R}^{N \times d_0}$: input signal matrix
  - $\mathbf{\Theta}^{(1)} \in \mathbb{R}^{T \times F_1}$: learnable GCN filter coefficients
  - $\mathrm{ReLU}(x) = \max(0, x)$ activation for training the network

Model architecture: Encoder



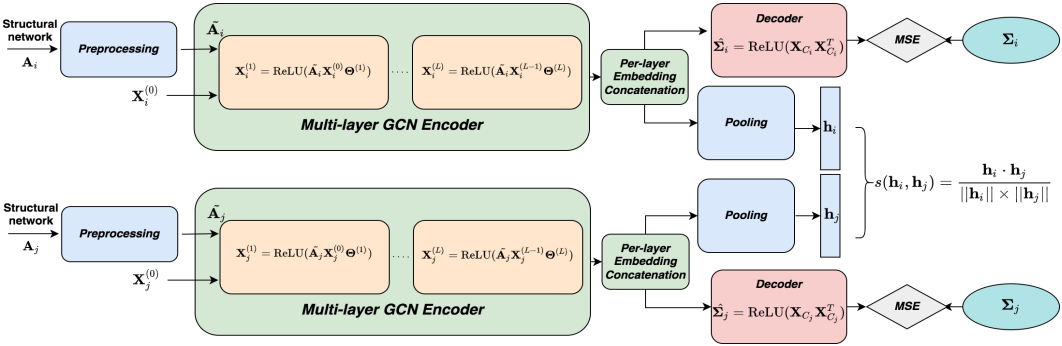
- Multi-layer GCN with dimension $32 \times 16 \times 8$
- Node embeddings concatenation
- Global row-wise mean pooling



- Node embeddings $\mathbf{X}_C$ go through the outer-product decoder

$$\hat{\mathbf{\Sigma}} = \mathrm{ReLU}(\mathbf{X}_C\mathbf{X}_C{}^T) \in \mathbb{R}^{N\times N}$$

- Generate estimate of empirical FC networks $\mathbf{\Sigma}$
- Reconstruction loss $\mathcal{L}_{\mathsf{MSE}}(\hat{\mathbf{\Sigma}}, \mathbf{\Sigma})$



- Obtain whole graph embedding $\mathbf{h}_i$ via row-wise average of $\mathbf{X}_C$
- For input graph pairs $\{\mathbf{A}_i, \mathbf{A}_j\}$ with label $\{l_i, l_j\}$, $\mathbf{h}_i, \mathbf{h}_j$ shall be
  - $\Rightarrow$ Highly (less) similar if $l_i = l_j$ ($l_i \neq l_j$)
  - $\Rightarrow$ Measured by cosine similarity $s(\mathbf{h}_i, \mathbf{h}_j)$

- N_s (N_d) graph pairs with same (different) labels
  - Pairwise similarity global loss function [Ktena et al'18]

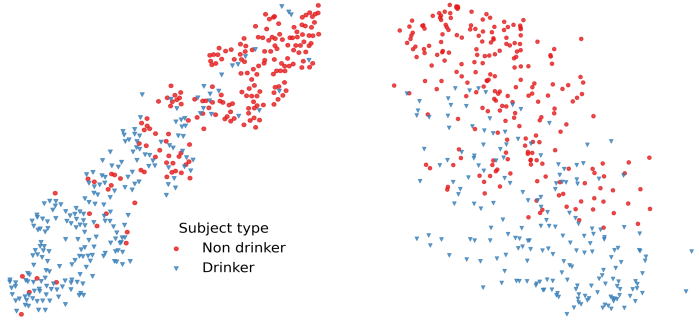
$$\mathcal{L}_{\mathbf{SIM}} = (\sigma^{2+} + \sigma^{2-}) + w \times \max(0, m - (\mu^{+} - \mu^{-}))$$

  - Minimize $\mu^{-} = \sum s(\mathbf{h}_i, \mathbf{h}_j)/N_d, l_i \neq l_j$
    $\Rightarrow$ Mean similarity between embeddings of different classes
  - Maximize $\mu^{+} = \sum s(\mathbf{h}_i, \mathbf{h}_j)/N_s, l_i = l_j$
    $\Rightarrow$ Mean similarity between embeddings of the same classes
  - Minimize the variance of similarities within and between classes
- Overall loss function is

$$\mathcal{L} = \mathcal{L}_{\mathbf{SIM}} + \lambda \times (\mathcal{L}_{\mathbf{MSE}}(\hat{\mathbf{\Sigma}}_i, \mathbf{\Sigma}_i) + \mathcal{L}_{\mathbf{MSE}}(\hat{\mathbf{\Sigma}}_j, \mathbf{\Sigma}_j))$$

- Classify test sample (KNN= 5) by cosine similarity in vector space
- Baseline: supervised GRL model [Li et al'21]
  - ⇒ Reconstruct FC from SC
  - ⇒ Subject classification via a logistic regression classifier

Model	Accuracy	F score
Siamese model	$\mathbf{0.6843 \pm 0.016}$	$\mathbf{0.7391 \pm 0.016}$
Baseline	0.6610 ± 0.043	0.6962 ± 0.030



Summary

- Building the Siamese network and training it with graph pairs
  - Increase the amount of training data
  - Bring in the flexibility to incorporate additional prior information
- Graph reconstruction of FC from SC
  - Parsimonious representation of population-level SC-FC relationship
- Similarity estimation between input graphs
  - Use subject labels as additional inputs for supervised classification
- Prospect of using graph-level, similarity-preserving embeddings
  - Measure SC-FC coupling for brain network analysis

Future work

- Reflect more intrinsic differences between drinkers and non-drinkers
- Spatio-temporal analysis of dynamic brain networks