Supervised Graph Representation Learning for Modeling the Relationship between Structural and Functional Brain Connectivity

Yang Li

Dept. of Electrical and Computer Engineering
University of Rochester
yli131@ur.rochester.edu
http://www2.ece.rochester.edu/~yli131/

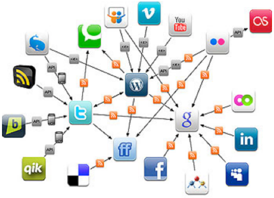
Co-authors: Rasoul Shafipour, Gonzalo Mateos and Zhengwu Zhang

Acknowledgment: NSF Awards CCF-1750428 and ECCS-1809356

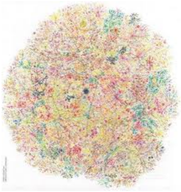
IEEE ICASSP, May 8, 2020

Network Science analytics

Online social media



Internet



Clean energy and grid analytics

- Network as graph $G(\mathcal{V}, \mathcal{E})$: encode pairwise relationships
- Desiderata: Process, analyze and learn from network data [Kolaczyk'09]

  $\Rightarrow$ Use G to study graph signals, data associated with nodes in $\mathcal{V}$
- Ex: Opinion profile, buffer congestion levels, **brain signal analyses**

Networks of the brain

- Challenge: understand human brain function and structure
  - Neuroimaging advances $\Rightarrow$ Data increase in volume and complexity
  - Graph-centric analysis and methods of network science [Sporns'10]
- Brain networks can reflect two connectivity patterns
  - Structural connectivity (SC). How is the brain wired?
    $\Rightarrow$ Anatomical tracts connecting brain regions (DTI)
  - Functional connectivity (FC). How the brain functions?
    $\Rightarrow$ Correlation between neural signals in different regions (fMRI)

Networks of the brain

- Challenge: understand human brain function and structure
  - Neuroimaging advances $\Rightarrow$ Data increase in volume and complexity
  - Graph-centric analysis and methods of network science [Sporns'10]
- Brain networks can reflect two connectivity patterns
  - Structural connectivity (SC). How is the brain wired?
    $\Rightarrow$ Anatomical tracts connecting brain regions (DTI)
  - Functional connectivity (FC). How the brain functions?
    $\Rightarrow$ Correlation between neural signals in different regions (fMRI)
- Key problem: deciphering the relationship between SC and FC
  - Simulations of nonlinear cortical activity models [Honey et al'09]
  - Diffusion-based parametric inverse problem [Abdelnour et al'14]
  - Network deconvolution [Li-Mateos'19]
- **Goal:** investigate SC-FC relationship

  $\Rightarrow$ Reconstruct FC from SC as a regression problem

  $\Rightarrow$ Learn lower-dimensional graph representations for classification

Graph signal processing (GSP)

- Graph $G(\mathcal{V}, \mathcal{E})$ with $N = |\mathcal{V}|$ nodes $\Rightarrow$ $\mathbf{A}$ and $\mathbf{L} = \mathbf{D} - \mathbf{A}$

Graph signal processing (GSP)

- Graph $G(\mathcal{V}, \mathcal{E})$ with $N = |\mathcal{V}|$ nodes $\Rightarrow$ $\mathbf{A}$ and $\mathbf{L} = \mathbf{D} - \mathbf{A}$
- Graph signals mappings $x : \mathcal{V} \to \mathbb{R}$, represented as vectors $\mathbf{x} \in \mathbb{R}^N$
- To understand GS $\Rightarrow$ Graph-shift operator $\mathbf{S} \in \mathbb{R}^{N \times N}$

  $\Rightarrow$ Local $S_{ij} = 0$ for $i \neq j$ and $(i, j) \notin \mathcal{E}$ $\Rightarrow$ Ex: $\mathbf{A}$ or $\mathbf{L}$

  $\Rightarrow$ Spectrum of $\mathbf{S} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T$

Graph signal processing (GSP)

- Graph $G(\mathcal{V}, \mathcal{E})$ with $N = |\mathcal{V}|$ nodes $\Rightarrow$ $\mathbf{A}$ and $\mathbf{L} = \mathbf{D} - \mathbf{A}$
- Graph signals mappings $x : \mathcal{V} \to \mathbb{R}$, represented as vectors $\mathbf{x} \in \mathbb{R}^N$
- To understand GS $\Rightarrow$ Graph-shift operator $\mathbf{S} \in \mathbb{R}^{N \times N}$

  $\Rightarrow$ Local $S_{ij} = 0$ for $i \neq j$ and $(i, j) \notin \mathcal{E}$ $\Rightarrow$ Ex: $\mathbf{A}$ or $\mathbf{L}$

  $\Rightarrow$ Spectrum of $\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$
- Graph Fourier Transform (GFT): $\tilde{\mathbf{x}} = \mathbf{U}^T\mathbf{x}$

Graph signal processing (GSP)

- Graph $G(\mathcal{V},\mathcal{E})$ with $N = |\mathcal{V}|$ nodes $\Rightarrow$ $\mathbf{A}$ and $\mathbf{L} = \mathbf{D} - \mathbf{A}$
- Graph signals mappings $x : \mathcal{V} \to \mathbb{R}$, represented as vectors $\mathbf{x} \in \mathbb{R}^N$
- To understand GS $\Rightarrow$ Graph-shift operator $\mathbf{S} \in \mathbb{R}^{N \times N}$
  - $\Rightarrow$ Local $S_{ij} = 0$ for $i \neq j$ and $(i,j) \notin \mathcal{E}$ $\Rightarrow$ Ex: $\mathbf{A}$ or $\mathbf{L}$
  - $\Rightarrow$ Spectrum of $\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$
- Graph Fourier Transform (GFT): $\tilde{\mathbf{x}} = \mathbf{U}^T\mathbf{x}$
- Generalize machine learning models for network data
  - $\Rightarrow$ GSP to define convolutions on graphs [Ortega et al'19]
  - $\Rightarrow$ Tool to integrate brain structure and function [Wang et al'19]

Graph convolutions

- Given a graph filter $\mathbf{H} = \sum_{i=0}^{K} h_i \mathbf{L}^i$ with $\mathbf{h} := [h_0, \ldots, h_K]^T$

$$\mathbf{H}\mathbf{x} = \left(\sum_{i=0}^{K} h_i \mathbf{L}^i\right)\mathbf{x} = \mathbf{U}\left(\sum_{i=0}^{K} h_i \boldsymbol{\Lambda}^i\right)\mathbf{U}^T\mathbf{x} = \mathbf{U}\left(\mathrm{diag}(\tilde{\mathbf{h}})\right)\tilde{\mathbf{x}}$$

Graph convolutions

- Given a graph filter $\mathbf{H} = \sum_{i=0}^{K} h_i \mathbf{L}^i$ with $\mathbf{h} := [h_0, \ldots, h_K]^T$

$$\mathbf{Hx} = \left(\sum_{i=0}^{K} h_i \mathbf{L}^i\right) \mathbf{x} = \mathbf{U}\left(\sum_{i=0}^{K} h_i \mathbf{\Lambda}^i\right) \mathbf{U}^T \mathbf{x} = \mathbf{U}\left(\mathrm{diag}(\tilde{\mathbf{h}})\right) \tilde{\mathbf{x}}$$

- ChebNet: Chebyshev polynomials of $\mathbf{\Lambda}$ [Defferrard et al'16]

$$\mathrm{diag}(\tilde{\mathbf{h}}) = \sum_{i=0}^{K} h_i \mathbf{\Lambda}^i \approx \sum_{i=0}^{K} \theta_i T_i(\bar{\mathbf{\Lambda}})$$

  - θ_i: polynomial filter coefficients to be learned
  - $\bar{\mathbf{\Lambda}} = 2\mathbf{\Lambda}/\lambda_{\max} - \mathbf{I}_N$, $\lambda_{\max}$: largest eigenvalue of $\mathbf{L}$
  - $T_k(x) = 2x T_{k-1}(x) - T_{k-2}(x)$ with $T_0(x) = 1$, $T_1(x) = x$
- Graph convolution operator becomes $(\bar{\mathbf{L}} = (2/\lambda_{\max})\mathbf{L} - \mathbf{I}_N)$

$$\mathbf{Hx} = \sum_{i=0}^{K} \theta_i T_i(\bar{\mathbf{L}})\mathbf{x}$$

  - Combines signal values at nodes K-hops away in G

Graph convolutional network (GCN)

- First-order approximation of ChebNet [Kipf-Welling'17]
  - Set $K = 1$, $\lambda_{\text{max}} = 2$, $\theta = \theta_0 = -\theta_1$

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

Graph convolutional network (GCN)

- First-order approximation of ChebNet [Kipf-Welling'17]
  - Set $K = 1$, $\lambda_{\text{max}} = 2$, $\theta = \theta_0 = -\theta_1$

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

- Compact rule for per-layer filtering update

$$\bar{\mathbf{X}} \leftarrow \bar{\mathbf{A}}\mathbf{X}\boldsymbol{\Theta}$$

  - $\bar{\mathbf{A}} = \mathbf{I}_N + \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$
  - $\mathbf{X}$: set of multiple observations of the graph signal $\mathbf{x}$
  - $\boldsymbol{\Theta}$ collects the learnable filter parameters

- $\bar{\mathbf{X}}$ integrates nodal attributes in $\mathbf{X}$ and topology in $\bar{\mathbf{A}}$

Graph convolutional network (GCN)

- First-order approximation of ChebNet [Kipf-Welling'17]
  - Set $K = 1$, $\lambda_{\max} = 2$, $\theta = \theta_0 = -\theta_1$

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

- Compact rule for per-layer filtering update

$$\bar{\mathbf{X}} \leftarrow \bar{\mathbf{A}}\mathbf{X}\boldsymbol{\Theta}$$

  - $\bar{\mathbf{A}} = \mathbf{I}_N + \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$
  - $\mathbf{X}$: set of multiple observations of the graph signal $\mathbf{x}$
  - $\boldsymbol{\Theta}$ collects the learnable filter parameters
- $\bar{\mathbf{X}}$ integrates nodal attributes in $\mathbf{X}$ and topology in $\bar{\mathbf{A}}$
- Q: Aggregate information from multiple hops within G?

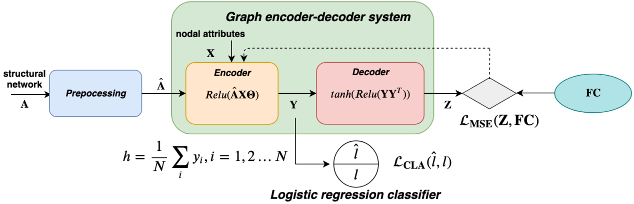
  $\Rightarrow$ Stack convolutional layers with pointwise non-linear activation

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

Problem statement

Model the relationship between brain structural and functional network

- Goal: summarize SC-FC relationship by simultaneously learning
  - Node embeddings to reconstruct FC from the given SC networks
  - Graph embeddings for graph classification
- Model: supervised graph encoder-decoder system



- Analysis: investigate group-wise difference within reconstructed FCs

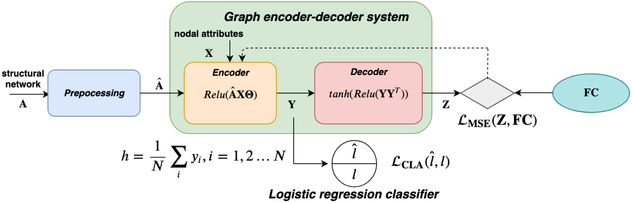
Model architecture: Encoder

- Input SC network $\mathbf{A} \in \mathbb{R}^{N \times N}$, N regions from brain atlas
  - $\Rightarrow$ Edge weights represent SC between brain regions
  - $\Rightarrow$ Preprocessing: $\hat{\mathbf{A}} := \bar{\mathbf{D}}^{-1/2}\bar{\mathbf{A}}\bar{\mathbf{D}}^{-1/2}$, $\bar{\mathbf{A}} = \mathbf{I}_N + \mathbf{A}$
- 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
- A single-layer GCN used for encoder to learn node embeddings

$$\mathbf{Y} = \mathrm{Relu}(\hat{\mathbf{A}}\mathbf{X}\boldsymbol{\Theta}) \in \mathbb{R}^{N \times F}$$

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

Model architecture: Decoder

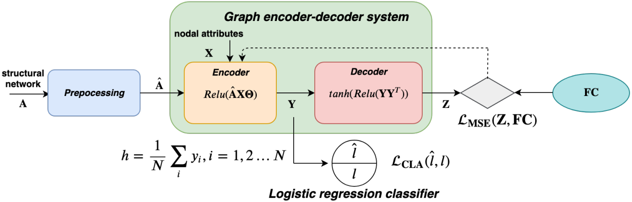


- Node embeddings $\mathbf{Y} \in \mathbb{R}^{N \times F}$ go through the outer-product decoder

$$\mathbf{Z} = \tanh(\mathrm{Relu}(\mathbf{Y}\mathbf{Y}^T)) \in \mathbb{R}^{N \times N}$$

- Weights in empirical FC networks restricted to [0,1]
  - Ensure the output of the decoder in the same range
  - Choose tanh and Relu over Sigmoid
- Loss function: MSE between $\mathbf{Z}$ and empirical FC

Model architecture: Classifier



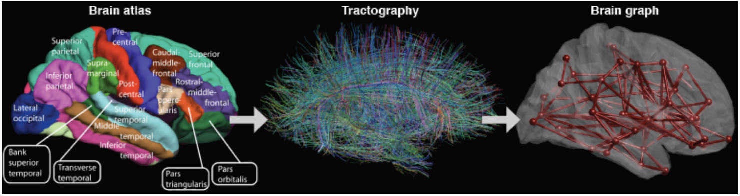
- Apply row-wise average-pooling on the encoder output $\mathbf{Y}$

  $\Rightarrow$ Vector summarizing SC-FC relationship, i.e., graph embedding
- Construct logistic regression classifier to predict subject labels
  - Sigmoid cross-entropy loss between predicted and empirical labels
- Loss function: $\mathcal{L} = \mathcal{L}_{\mathbf{MSE}}(\mathbf{Z}, \mathbf{FC}) + \lambda \times \mathcal{L}_{\mathbf{CLA}}(\hat{l}, l)$

Numerical tests: Data

- 466 subjects from Human Connectome Project (HCP)
  - Two classes: 245 non-drinkers, 221 heavy drinkers
- Preprocessed SC network $\mathbf{A}$ from diffusion MRI

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



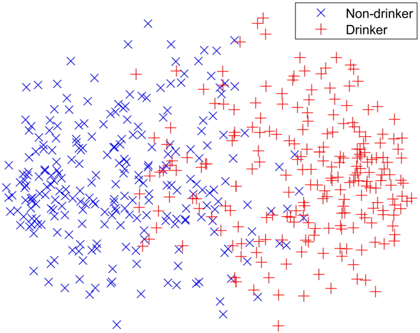
- Preprocessed FC network from functional MRI

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

  $\Rightarrow$ Estimated FC $\Leftrightarrow$ Pearson correlation between BOLD signals
- One-hot encoding as the signal on each graph node ($\mathbf{X} = \mathbf{I}_N$)

Numerical tests: Results

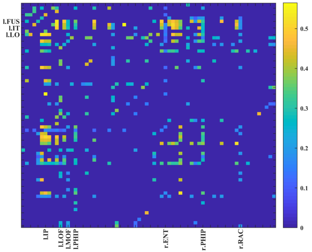
- MSE between reconstructed and empirical FC networks
  - Average test reconstruction error = 0.034174 with std = 0.00208
  - Captured population patterns of SC-FC relationship
- Classification accuracy: $67.4 \pm 2\%$
  - Captured discriminative patterns within each group

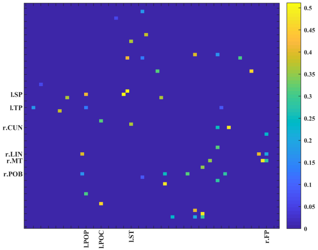


  - Reduced dimensional graph embeddings exhibit cluster structure

Numerical tests: Reconstructed FC

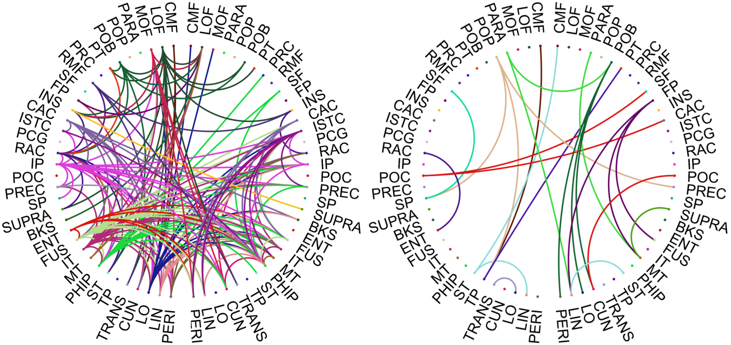
- Investigate group-wise difference within reconstructed FCs
  - Captured difference between subjects in latent representations
- Test for significant group-wise difference in functional connections
  - Edge-wise T-tests ($p < 0.05$) with FDR correction





  - Connections weaker (left) & stronger (right) in drinkers

Numerical tests: Class differences



- Left: subnetwork of connections weaker in drinkers
  - *Entorhinal, Parahippocampus*, limbic system impaired in drinkers
  - Overall decrease in connection strengths in drinkers
- Right: subnetwork of connections stronger in drinkers
  - Involve regions in multiple cortices $\Rightarrow$ neural compensation
  - Additional connections compensate for alcohol damages

Summary and the road ahead

- Preliminary study of human brain SC-FC relationship
- Supervised graph encoder-decoder system to simultaneously learn
  - Node embeddings for FC reconstruction with given SC
  - Graph embeddings as representations for graph classification
- Latent variables contain vital information integrating
  - Universal SC-FC relationship patterns in the population
  - Discriminative power preserved in key connections
  - Information from both types of brain networks (FC and SC)

Summary and the road ahead

- Preliminary study of human brain SC-FC relationship
- Supervised graph encoder-decoder system to simultaneously learn
  - Node embeddings for FC reconstruction with given SC
  - Graph embeddings as representations for graph classification
- Latent variables contain vital information integrating
  - Universal SC-FC relationship patterns in the population
  - Discriminative power preserved in key connections
  - Information from both types of brain networks (FC and SC)
- Envisioned application domains
  (a) Integrate temporal graph signals as node attributes
  (b) Investigate different sets of parameters to train the model
  (c) Explore parameters within GCN models, e.g. graph saliency map