

CSCI8850 Advanced Biomedical Imaging Analysis

Paper Presentation

Christian McDaniel

Multi-View Graph Convolutional Network and Its Applications on Neuroimage Analysis for Parkinson's Disease

Zhang, *et. al.*

2018

Presentation Roadmap:

- Parkinson's Disease
- Neuroimage Analysis
 - Modalities
 - Previous efforts
 - PPMI Dataset
- Background (The Emerging Field of Signal Processing on Graphs):

Presentation Roadmap (cont):

- Proposed Framework and Experiment
 - Multi-View Graph Convolutional Network (MVGCN)
 - Graph Convolutional Network (GCN)
 - Multi-View Pooling
 - Pairwise Matching
 - Relationship Prediction
 - Experiment Settings
 - Results
 - Conclusions

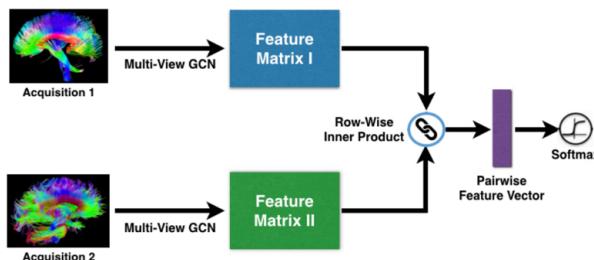
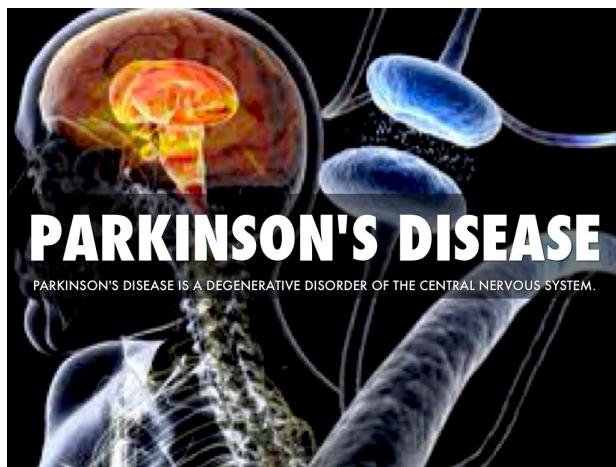


Figure 1: Overall flowchart of our framework.

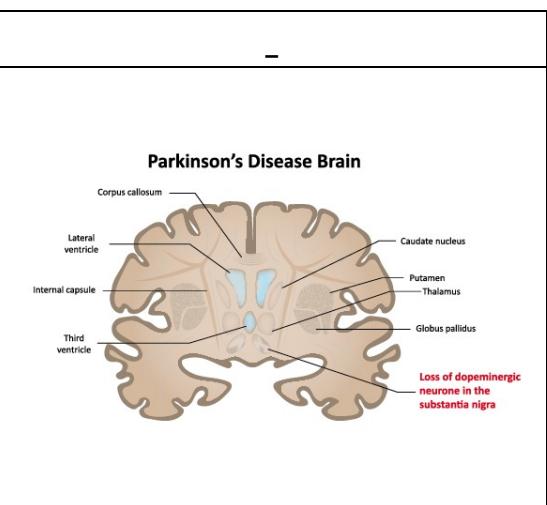
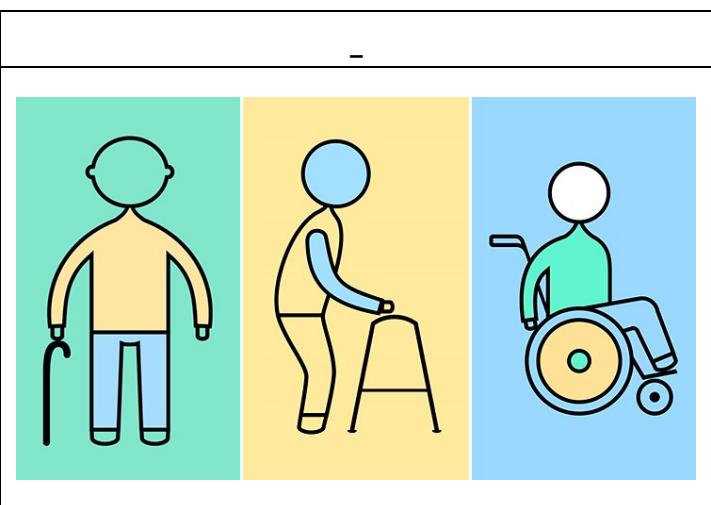
Parkinson's Disease

- One of the most prevalent neurodegenerative diseases
 - 500,000+ Americans
 - 14th highest cause of death in US



Parkinson's Disease - Pathology

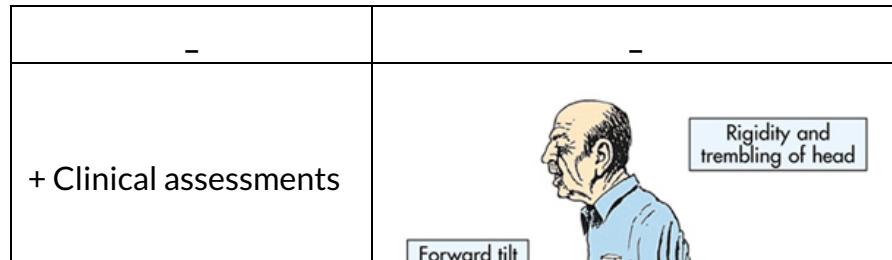
- Predominantly affects the dopaminergic neurons in the substantia nigra
- Highly progressive
- Highly heterogeneous - clinical manifestations vary on an individual basis

Parkinson's Disease - Diagnosis

Parkinson's Disease - Diagnosis

Diagnosis is largely clinically based:



Parkinson's Disease - Diagnosis

+ Diagnosis has traditionally involved a degree of subjectivity, and its distinction between similar diseases can be somewhat arbitrary

Parkinson's Disease - Research

Parkinson's Disease - Research

- + There has been much effort to develop more rigorous, qualitative diagnostic tools

Parkinson's Disease - Research

Much of the research is done using:

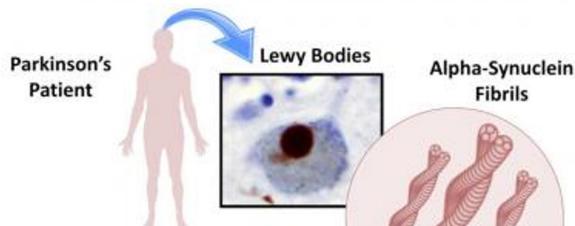
Clinical Data



Parkinson's Disease - Research

Much of the research is done using:

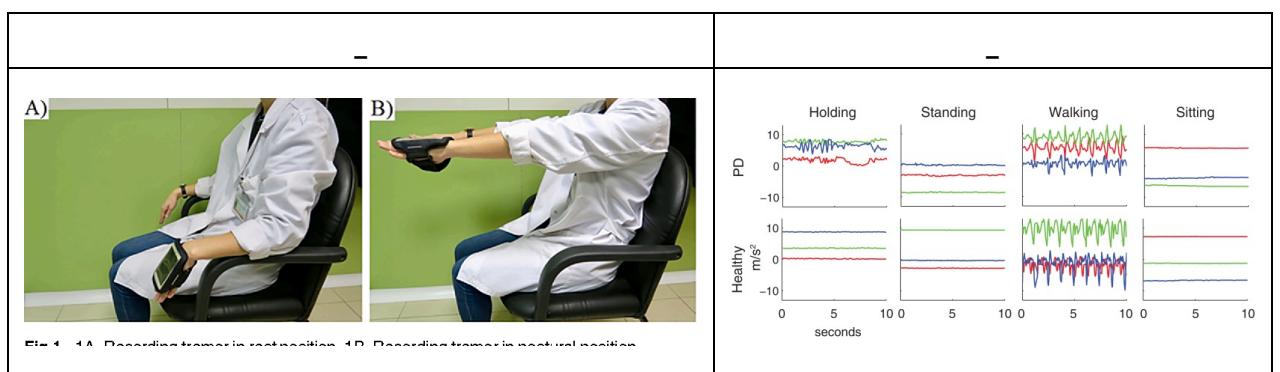
Genetic and Molecular Data



Parkinson's Disease - Research

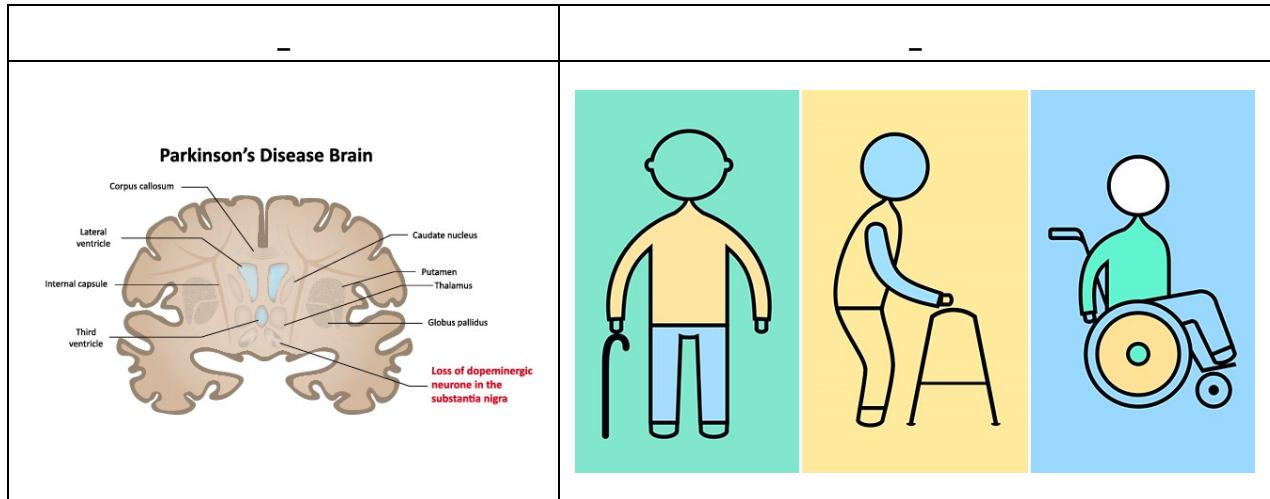
Much of the research is done using:

Accelerometer Data



Parkinson's Disease - Pathology

- Predominantly affects the dopaminergic neurons in the substantia nigra
- Highly progressive
- Highly heterogeneous - clinical manifestations vary on an individual basis

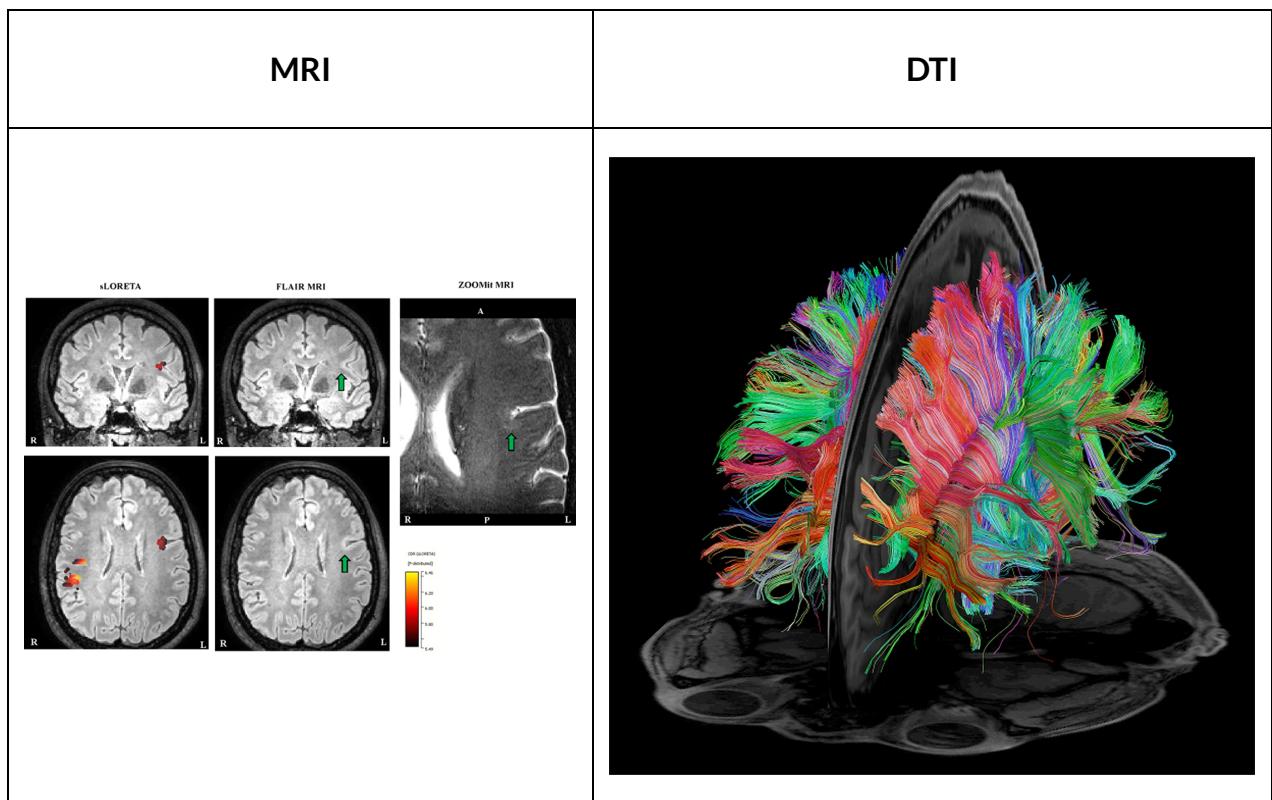


Neuroimaging analysis emerged as a promising tool to:

- Capture the structural/functional differences observed
- Capture the progressive changes in the disease over time
- Account for the individual differences across PD pt's

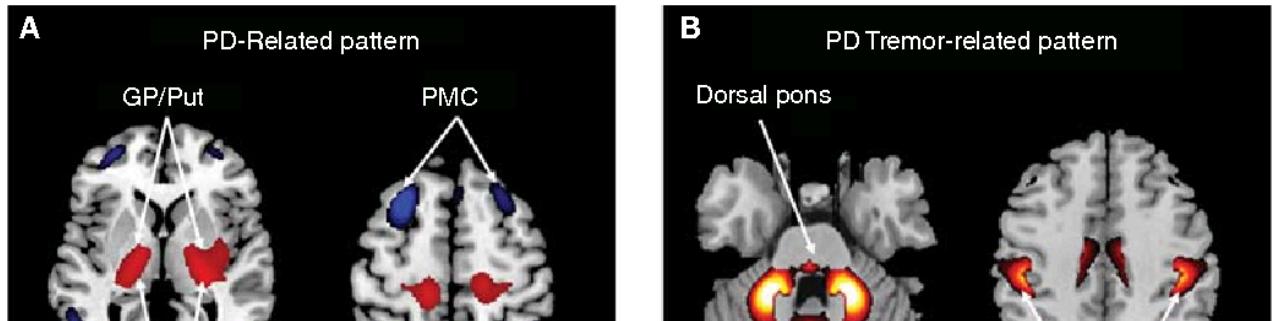
Structural-functional relationship in PD patients

Neuroimaging Review:



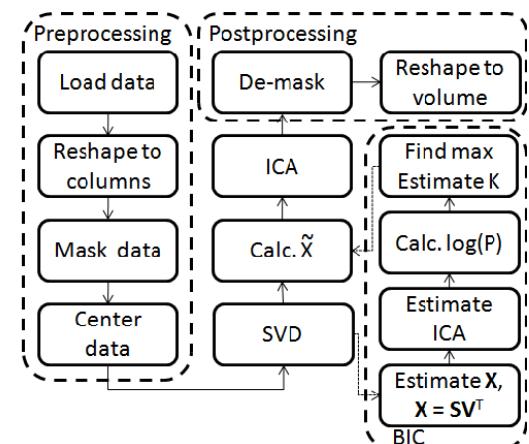
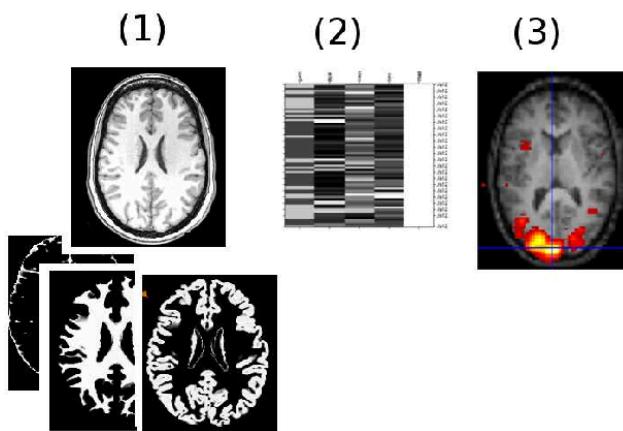
Previous Neuroimaging research successes for PD research:

- Volumetric differences in the **substantia nigra**
- Volumetric loss in **olfactory bulbs/tracts** in PD pt's, and increased loss over PD duration
- Decreased FA in substantia nigra (DTI)
- greater FA reductions in **caudal (vs. middle/rostral) regions** of the SN
 - distinguished PD from controls with 100% sensitivity *and* specificity



Prevailing computational methods for neuroimaging analysis:

- data-driven and hypothesis free
- Mostly linear, multilinear
- Drive network discovery and imaging genomics



Limitations:

- + Linear methods for non-linear data
- + Too dependent on choice of algorithm/ parameter settings

These limitations inspired the development of data driven methods...

These limitations inspired the development of data driven methods...

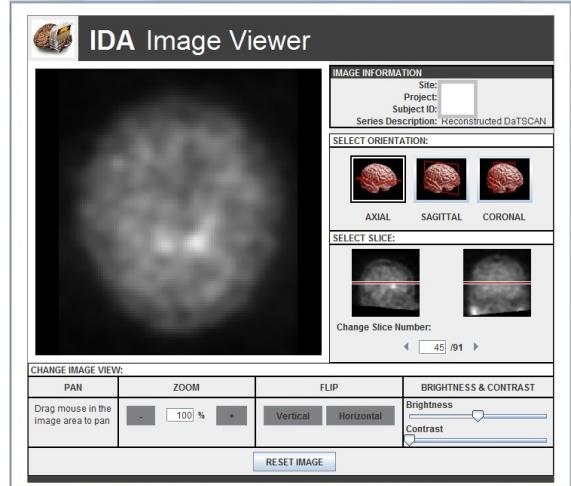
... in need of large amounts of data

The emergence of a large-scale PD-related neuroimaging database:

The PPMI Dataset

The PPMI Dataset

The PPMI Dataset

-	-
<ul style="list-style-type: none">+ Thousands of image sets<ul style="list-style-type: none">+ Multiple modalities+ "Test-retest" feature+ Raw and processed images	

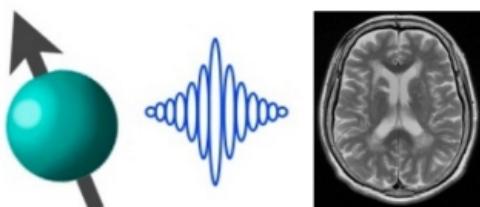
Now that we have our dataset...

... we can plug-n-chug into our favorite state-of-the-art
DL architecture, right?



Feature representation:

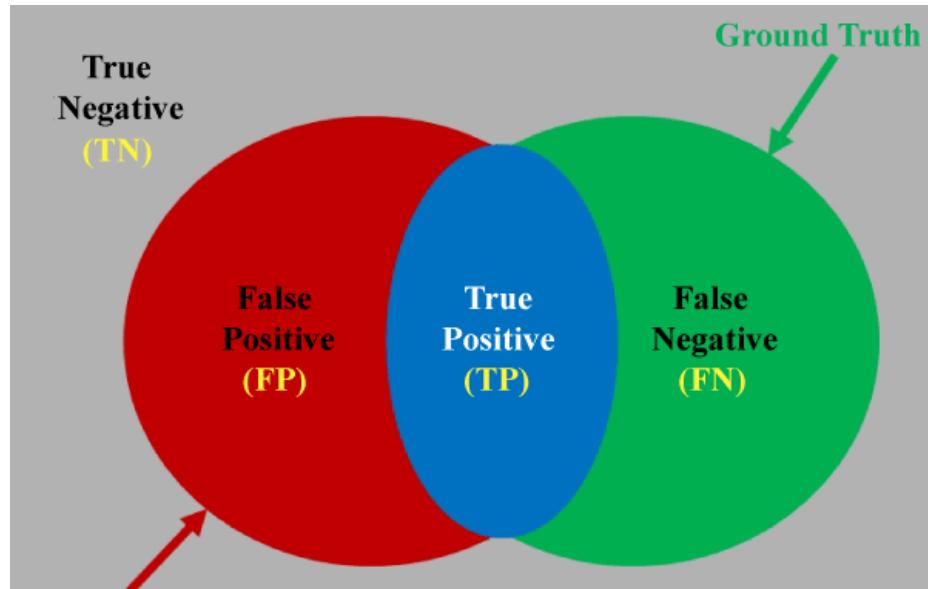
- Biomedical imaging is often **not optical imaging**
- Instead, we are *recreating* an image from signal transmitted through an object (X-ray, CT) or from signals received from the object (MRI, DTI)
 - The signals received are quite **noisy and incomplete**
 - Additionally, each image typically varies on a subject-to-subject basis



Feature representation:

The case for graph convolutional networks:

- dice coefficient / IoU as loss function (used for image segmentation) is **not sufficient** for internal structure/function representation



Feature representation:

The multi-view graph convolutional network:

Hypothesis: it would be nice to borrow from *signal processing* techniques, which are used to dealing with such messy data

Hypothesis: it would be nice to borrow from *signal processing* techniques, which are used to dealing with such messy data

Enter: Shuman, *et. al.* (2012)

The Emerging Field of Signal Processing on Graphs

Extending High-Dimensional Data Analysis to Networks and Other Irregular Domains

Shuman *et. al.*

The Emerging Field of Signal Processing on Graphs

Extending High-Dimensional Data Analysis to Networks and Other Irregular Domains

Shuman *et. al.*

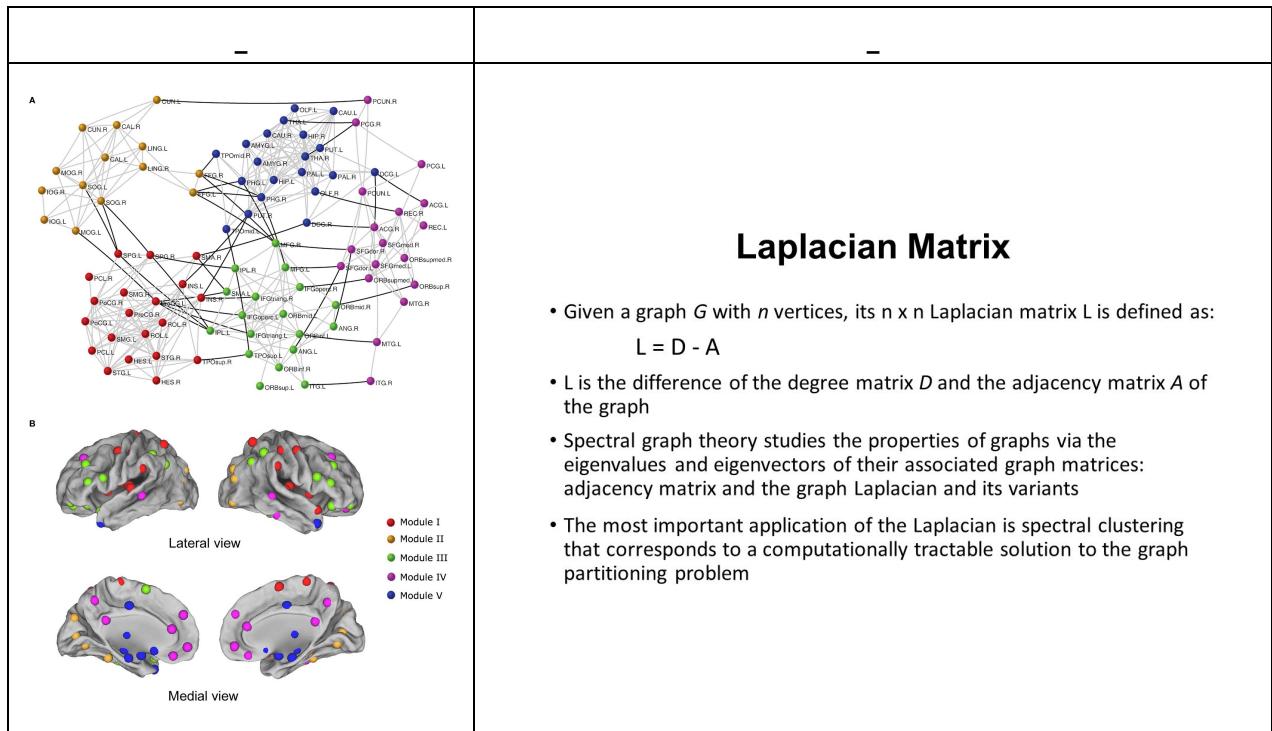
The goal was devised to process a graph signal as if it were a discrete-time signal

- Both a signal on a graph with N vertices and a classical discrete-time signal with N vertices can be viewed as vectors in \mathbb{R}^N
- So, noisy fMRI images can be approximated as signals on weighted graphs

Defining the graph:

- $G = \{V, E, W\}$
 - one graph = one image
 - V = the samples in each image, i.e., ROI's
 - a **signal** is defined on each vertex
 - $E = e(i,j)$ exists if the ROI's are structurally/functionally connected

Defining the graph - The graph Laplacian



Spectral Graph Domain:

The graph Laplacian

<p>Background spectral clustering is a classic technique to partition graphs by looking at eigenvectors [Fiedler73]</p> <p>Graph Laplacian Eigenvector</p> <p>Cluster</p>	<p>Graph G</p> <p>Incidence Matrix $In(G)$</p> $In(G) = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & -1 & & \\ 2 & 1 & -1 & \\ 3 & & 1 & -1 \\ 4 & & & 1 \\ 5 & & & & 1 \end{bmatrix}$ <p>Laplacian Matrix $L(G)$</p> $L(G) = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & -1 & 2 & -1 & \\ 3 & -1 & 2 & -1 & \\ 4 & & -1 & 2 & -1 \\ 5 & & & -1 & 1 \end{bmatrix}$ <p>Incidence and Laplacian Matrices</p> <p>Nodes numbered in black Edges numbered in blue</p>

Spectral Graph Domain:

The graph Laplacian

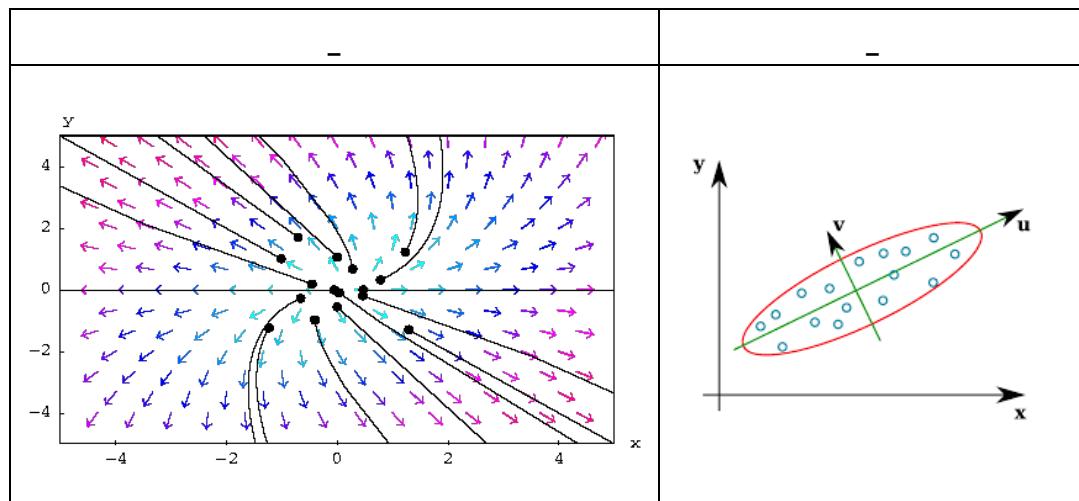
- Matrix whose i,j th entry is:
 - negative iff i and j are adjacent (connected)
 - 0 if $i \neq j$ and not connected
 - arbitrary if $i=j$ (often the degree)
- Symmetric matrix:
 - $A =$ its transpose;
 - A has a complete set of orthonormal eigenvectors with associated real, non-negative eigenvalues
- non-normalized graph Laplacian: $L = D - W$
 - $L = U V U^T$
 - U is the matrix of eigenvectors on a graph
 - V is the diagonal matrix of eigenvalues
 - optimizes an objective relative to the number of nodes in the graph
- normalized graph Laplacian: $L^\Delta = I - D^{-1/2} W D^{-1/2}$
 - optimizes relative to the volume of the graph

Defining the graph signal

The graph signal is defined as the eigenvectors associated with the graph Laplacian at each vertex

Review:

Eigenvectors as descriptors of a data-space:

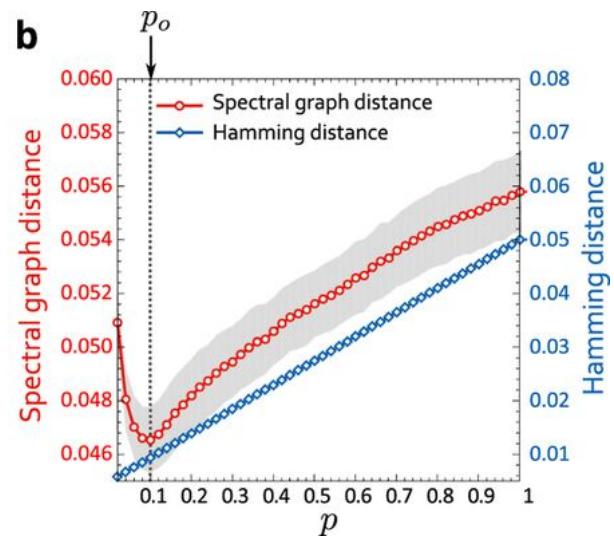
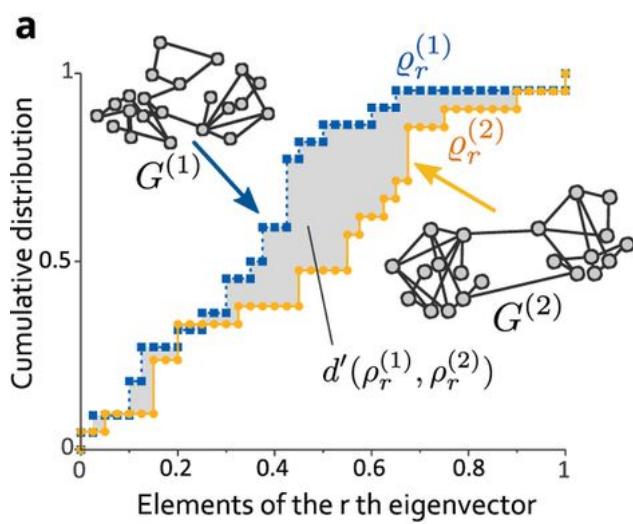


A signal is now a linear combination of parts

The Spectrum is the set of eigenvectors, their eigenvalues and their multiplicity

Review:

Spectral Graph Theory

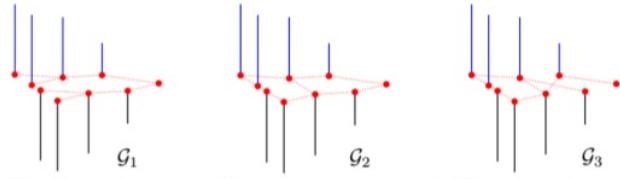


Defining the graph signal

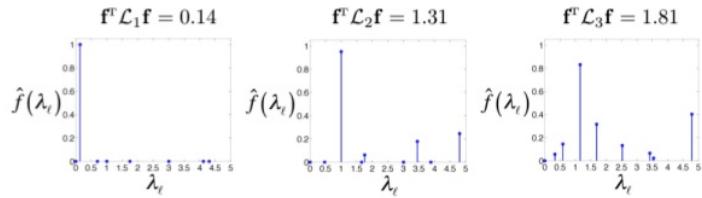
The graph signal is defined as the eigenvectors associated with the graph Laplacian at each vertex

Graph Signal Smoothness

- Quadratic form on L:



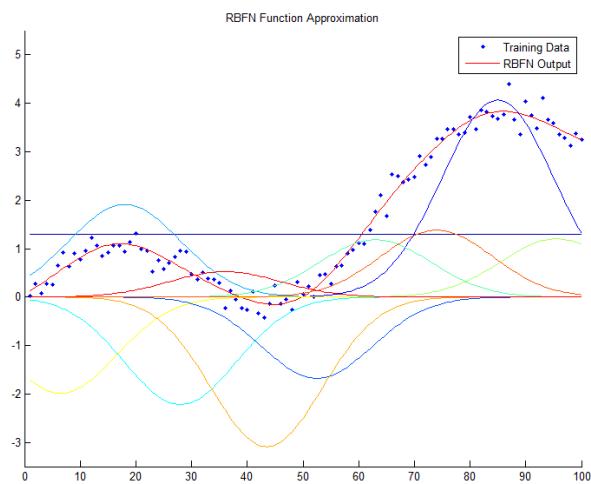
The same signal has different smoothness wrt different graphs



Processing the Signal

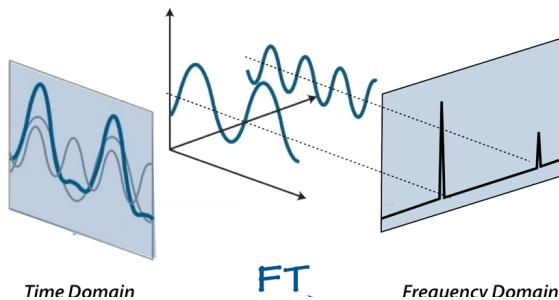
Signal Processing:

Function Approximation

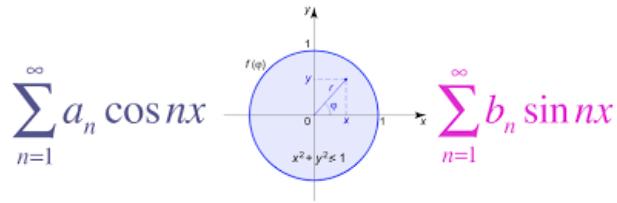


Signal Processing:

The Fourier Transform



Fourier Series --> linearly separable components	Fourier space = Complex space
<p>A unit circle in the complex plane with axes labeled \mathcal{R} (Real) and \mathcal{I} (Imaginary). A vector from the origin to a point on the circle is labeled $e^{i\varphi} = \cos \varphi + i \sin \varphi$. The angle φ is shown between the positive Real axis and the vector.</p>	<p>A 3D plot illustrating the complex exponential signal $e^{j2\pi f_0 t}$ in the Fourier space. The vertical axis is the Imaginary axis, ranging from -2 to 2. The horizontal axes are the Real axis (ranging from -2 to 2) and Time (ranging from 0 to 3). The signal is represented by a series of concentric circles in the complex plane, with arrows indicating the direction of rotation. Labels include $e^{j2\pi f_0 t}$, $\sin(2\pi f_0 t)$, and $\cos(2\pi f_0 t)$.</p>



$$f(x) \approx F_n(x) = a_0 + \sum_{k=1}^n a_k \cos(kx) + \sum_{k=1}^n b_k \sin(kx)$$

where

$$a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) dx$$

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(kx) dx \text{ for } k > 0$$

$$b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(kx) dx \text{ for } k > 0$$

Connecting the Laplacian to the Fourier Transform

- Fourier Transform:
 - Expansion of a function in terms of the complex exponentials (i.e., $\cos(e) + i \sin(e)$)
 - offers a discretized function for the flux of a system
- Laplace Operator (1-D):
 - The divergence of the gradient of a function (scalar notion of the gradient's flux)
 - Those same complex exponentials are the eigenfunctions of the one-dimensional Laplace operator
 - The eigenvalues $2(\pi) f(t)$ carry a notion of *frequency*

Connecting the Laplacian to the Fourier Transform

- The graph FT can be defined on the vertices of G as the expansion of any function (signal) f in terms of the eigenvectors of the graph Laplacian
 - $L = U \Lambda U^T$
 - Let x be the signal defined on the vertices of the graph
 - The Graph Fourier Transform (GFT) is defined as $x^* = U^T * x$
 - This converts the signal x to the spectral domain spanned by the Fourier basis U
- Magnitude of the eigenvalue is proportional to the frequency
 - Eigenvectors associated with larger eigenvalues oscillate more rapidly between connected vertices

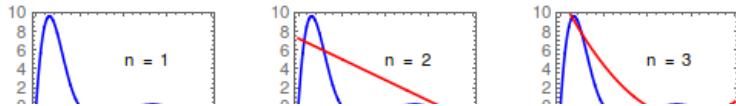
Fourier series best approximate continuous periodic functions

Fourier series best approximate continuous periodic functions

And our functions are discrete, in limited space, and irregularly spaced

Let's simplify further ...

Chebyshev Polynomial



Chebyshev Polynomial

Chebyshev polynomials - interpolation

... we are now faced with the same problem as with the Fourier series. We want to approximate a function $f(x)$, this time not a periodical function but a function which is defined between $[-1, 1]$.

We are looking for $g_n(x)$

$$f(x) \approx g_n(x) = \frac{1}{2} c_0 T_0(x) + \sum_{k=1}^n c_k T_k(x)$$

... and we are faced with the problem, how we can determine the coefficients c_k . Again we obtain this by finding the extremum (minimum)

$$\frac{\partial}{\partial c_k} \left[\int_{-1}^1 (g_n(x) - f(x))^2 \frac{dx}{\sqrt{1-x^2}} \right] = 0$$

Chebyshev Polynomial

The coefficients $c(k)$ *are* the Fourier coefficients $a(k)$

THUS, ...

... if we can learn the Chebyshev coefficients associated with the GFT of each image's signal ...

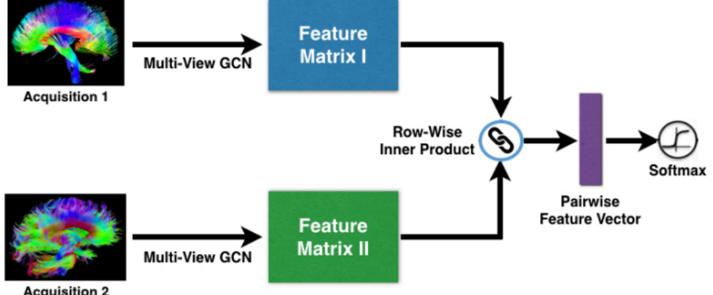
... (comprised of the Laplacian eigenvectors across connected ROI's), ...

... this can govern the convolutional operator of our CNN.

i.e., the Graph Convolutional Network

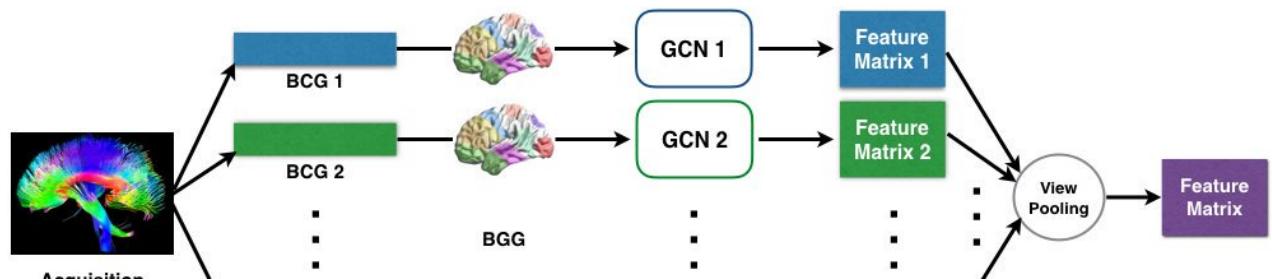
Now let's apply this to neuroimaging analysis for PD:

Current Paper:

Three Main Components	
<p>MV-GCN: Incorporate multiple modalities to learn a feature representation containing global and local information.</p> <p>Similarity Measure: Use this abstract feature representation to learn a similarity vector for a given pair of images</p> <p>Softmax Classifier: so as to cluster similar images</p>	 <p>Figure 1: Overall flowchart of our framework.</p>

Feature representation:

The multi-view graph convolutional network:



Now let's apply this to neuroimaging analysis for PD:

Graph Construction

Graph Construction:

- determine ROI's across all MRI images

Graph Construction:

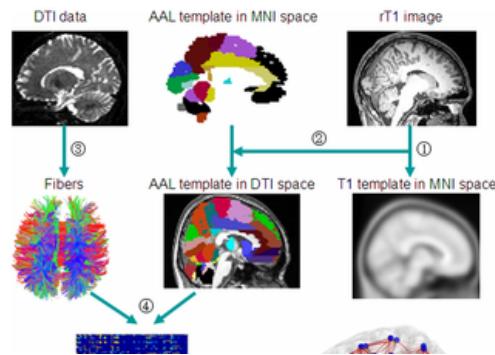
- determine ROI's across all MRI images
- each ROI becomes a vertex on the **Brain Geometry Graph (BGG)**
 - K-Nearest Neighbor graph
 - edges are weighted by Gaussian similarity function of Euclidean distances

$$k(x_i, x_j) = e^{\frac{||x_i - x_j||_2^2}{2\sigma^2}}$$

- This yields an adjacency matrix **A** representing the similarity to nearest similar ROIs
- All subjects share the same BGG
- We can define a shared graph Laplacian **L** on this map

Define the feature space:

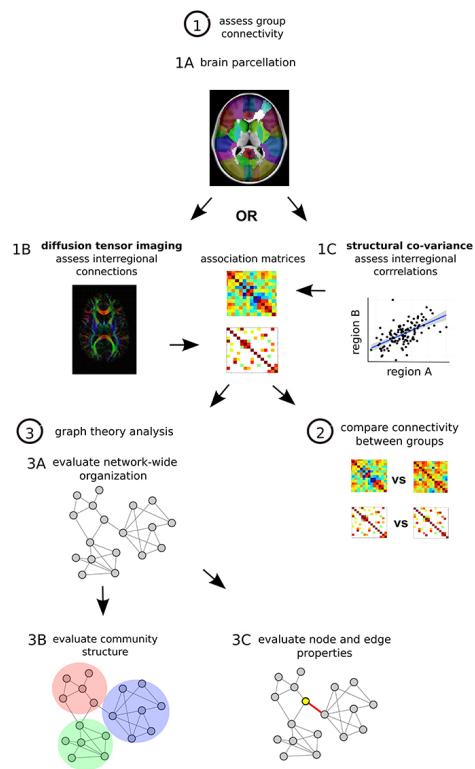
- Use multiple brain tractography algorithms on **DTI** images to yield multiple **Brainn Connectivity Graphs (BCGs)**



Define the feature space:

- Use multiple brain tractography algorithms on **DTI** images to yield multiple **Brainn Connectivity Graphs (BCGs)**
 - V again equals the ROI's of the image
 - E consists of the *connectivity strength* between each ROI
 - again yielding a similarity matrix X
- Each of N image sets is associated with M BCG's
 - V is the same across BCG's, while E varies for each one
 - Each subject is represented by a *group of similarity matrices*

Relationship Prediction - Graph Convolutional Networks



Relationship Prediction - Graph Convolutional Networks

- Learn a feature matrix from each BCG and their shared BGG
 - Each BCG can be represented by a **vector of Chebyshev coefficients**

Graph Convolutional Network

- Conveniently, convolutions in the vertex domain become multiplication operations in the graph spectral domain
- Remembering our equations for the graph Laplacian: $\mathbf{L} = \mathbf{U} \Lambda \mathbf{U}^T$
- and the GFT: $\mathbf{x}^{\wedge} = \mathbf{U}^T \mathbf{T}^* \mathbf{x}$
- we can define the graph convolution as such:

$$\mathbf{y} = g_{\theta}(\mathbf{L})\mathbf{x} = g_{\theta}(\mathbf{U} \Lambda \mathbf{U}^T) \mathbf{x} = \mathbf{U} g_{\theta}(\Lambda) \mathbf{U}^T \mathbf{x},$$

- where θ is a vector of Fourier coefficients
- $g(\theta)$ is the filter, a function of Λ such that, taking from our Chebyshev equation

$$f(x) \approx g_n(x) = \frac{1}{2} c_0 T_0(x) + \sum_{k=1}^n c_k T_k(x)$$

we can define

$$g_{\theta}(\Lambda) = \sum_{p=0}^{s-1} \theta_p T_p(\tilde{\Lambda})$$

- where $\theta(p)$ is a vector of Chebyshev coefficients
- $T(\tilde{\Lambda})$ is the Chebyshev polynomial of order p evaluated at $\tilde{\Lambda} = 2\Lambda/\lambda_{\max} - \mathbf{I}$ (diag. matrix of scaled eigenvalues)

Graph Convolutional Network

This substitution yields:

$$\mathbf{y} = g_{\theta}(\mathbf{L})\mathbf{x} = \sum_{p=0}^{s-1} \theta_p T_p(\tilde{\mathbf{L}})\mathbf{x},$$

where $\tilde{\mathbf{L}} = \frac{2\Lambda}{\lambda_{\max}} \mathbf{L} - \mathbf{I}$

Graph Convolutional Network

Further:

- If we define

$$\tilde{x}_p = T_p(\tilde{L})x$$

- we see that

$$\tilde{x}_i = 2 \tilde{L} \tilde{x}_{p-1} - \tilde{x}_{p-2}$$

with $\tilde{x}_0 = x$ and $\tilde{x}_1 = \tilde{L}x$

- i.e.,

$$\tilde{x}_p = T_p(\tilde{L})x$$

can be defined recursively from the normalized graph Laplacian and only the coefficients need to be learned

Graph Convolutional Network

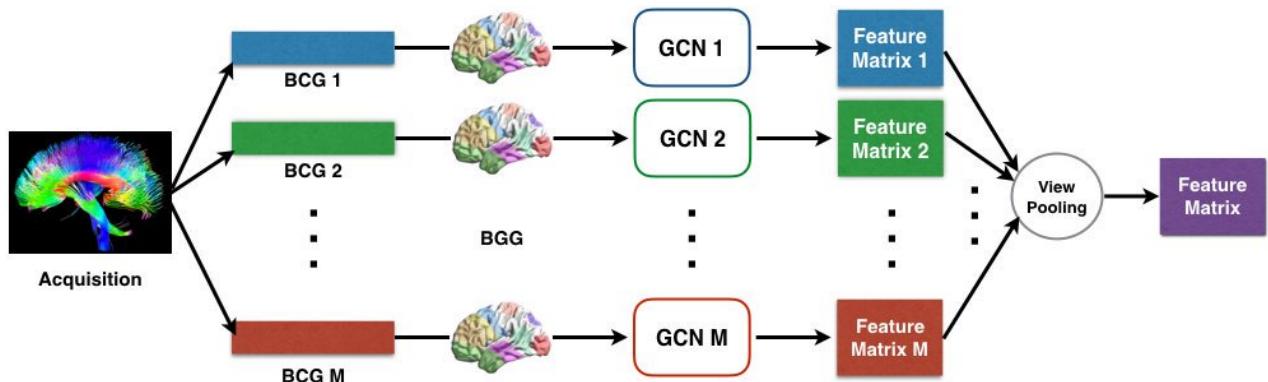
- the jth output feature map from a GCN is given by

$$\mathbf{y}_j = \sum_{i=1}^{F_{in}} g_{\theta_{i,i}}(\mathbf{L}) \mathbf{x}_i$$

- where $\mathbf{x}(i)$ is the i-th row of the input connectivity matrix (BCG) \mathbf{X}

View pooling:

- Multiple tractographies are aggregated together
- An element-wise maximum operation is used across all M feature matrices for a given subject
 - maximum operation combines the views' more informative features
 - instead of averaging (weakening the strongest features)
- This produces a shared feature matrix \mathbf{Z} for each subject
 - giving a combined vector of Chebyshev coefficients for each ROI



Pairwise Matching Strategy

1. Compile the dataset of all pairs of feature matrices (Z)
 - Normalize each matrix so that the sum of squares of each row = 1
 2. Define a pairwise Similarity measure:
 - if two subjects are similar (re: PD v. HC)
 - they should have a high probability of having the same class label
 -
 -
- $\text{sim}(\mathbf{z}_p^i, \mathbf{z}_q^i) = \mathbf{z}_p^{i^T} \mathbf{z}_q^i, \quad i = 1, 2, \dots, n$
- where \mathbf{z}_p^i and \mathbf{z}_q^i are the i-th row vectors of the normalized Z matrices

Relationship Prediction

- The pairwise matching layer yields a feature vector r
 - each element in r is a row-wise similarity
- r is passed to a fully-connected Softmax layer for classification
 -

$$p(y = j | r) = \frac{\exp(\mathbf{w}_j^T \mathbf{r})}{\sum_{c=1}^C \exp(\mathbf{w}_c^T \mathbf{r})}$$

- where $w(c)$ is the weight vector of the c -th class and r is the final abstract representation of the input example

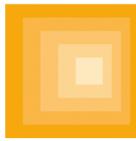
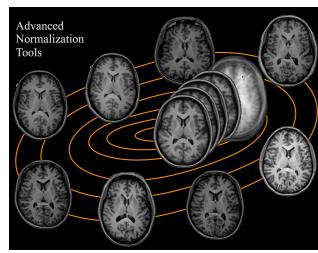
IN SUMMARY:

- This network has three main components:
 - Multi-View Graph Convolutional Network (MVGCN)
 - Learn a feature representation for each subject across multiple BCG's (tractography views)
 - Features obtained via the normalized graph Laplacian matrix and the graph Fourier Transform
 - Pairwise Matching Strategy
 - Softmax Relationship Prediction
- Each component is trained using backpropagation and stochastic optimization

The Experiment

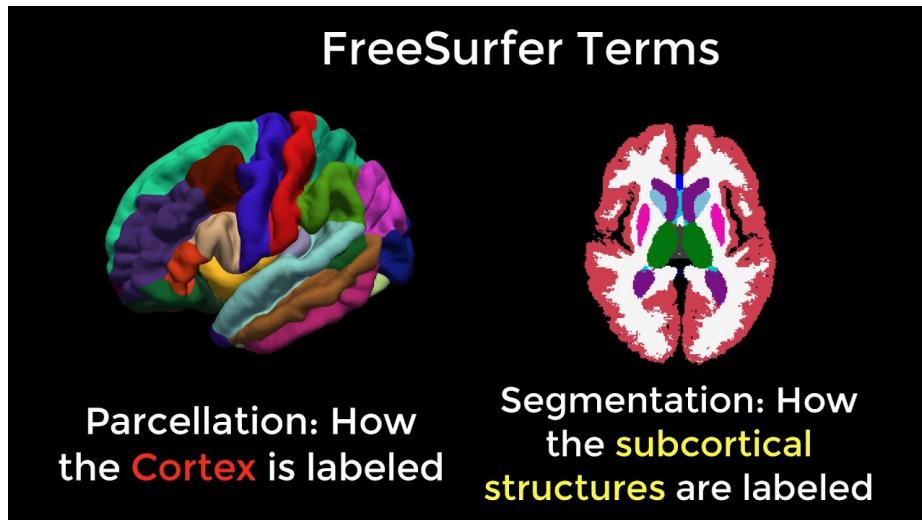
The Data: DTI

- DTI acquisitions of 754 subjects (PPMI)
 - 596 PD
 - 158 HC
- Head Motion and eddy-current distortion correction:
 - FSL eddy-correct tool
- Skull-stripping/ brain segmentation:
 - FSL Brain Extraction Tool (BET)
- echo-planar induced (EPI) susceptibility artifacts (e.g., tissue-fluid interface distortions)
 - co-registered to respective preprocessed structural MRI images
 - Advanced Normalization Tools (ANT) with SyN non-linear registration algorithm

-	-	-
 PARKINSON'S PROGRESSION MARKERS INITIATIVE <small>Play a Part in Parkinson's Research</small>	 FSL stands for FMRI Software Library  <small>Abbreviations.com</small>	 Advanced Normalization Tools

The Data: MRI

- 84 ROIs are parcellated from T1-weighted structural MRI using Freesurfer
 - BGG: Each ROI's coordinate is defined using the mean coordinate for all voxels



BCGs: Tractography algorithms

Based on the 84 ROIs in the RGG

Creating the pairwise dataset:

- 283,881 pairs in total (matching v. non-matching brain networks)
 - PD - PD pair is a match
 - HC - HC pair is a match
 - PD - HC pair is a non-match
 - 189,713 matching pairs
 - 94,168 non-matching pairs

Experiment Settings:

- 5-fold stratified CV
- BGG: 10-NN BGG is constructed
 - 84 vertices, 527 edges
- For fully connected layers:
 - Tested one-layer: 1024 units
 - and 2 layer: 1024 units - 64 units
- Adam optimizer with initial learning rate = 0.005
- Architecture settings were optimized via CV

Results (1): GCN

- Compared GCN to raw-edges weights and PCA performance for each DTI tractography algorithm
- Used same matching component and softmax component for each method

Table 1: Results for classifying matching vs. non-matching brain networks in terms of AUC metric.

Methods	Views					
	FACT	RK2	SL	TL	ODF-RK2	Hough
Raw Edges	58.47±4.05	62.54±6.88	59.39±5.99	61.94±5.00	60.93±5.60	64.49±3.56
PCA	64.10±2.10	63.40±2.72	64.43±2.23	62.46±1.46	60.93±2.63	63.46±3.52
FCN	66.17±2.00	65.11±2.63	65.00±2.29	64.33±3.34	68.80±2.80	61.91±3.42
FCN _{2l}	82.36±1.87	81.02±4.28	81.68±2.49	81.99±3.44	82.53±4.74	81.77±3.74
GCN	92.67±4.94	92.99±4.95	92.68±5.32	93.75±5.39	93.04±5.26	93.90±5.48

Results (2): MVGCN

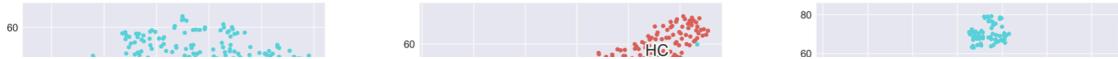
- Compared the clustering abilities of the MVGCN (re: PD vs. HC)
 - Clustering performance measured via Normalized Mutual Information (NMI)

Table 2: Comparison of binary classification (AUC) and acquisition clustering (NMI) results using both single-view and multi-view architectures.

Architectures	AUC	NMI
PCA100-M-S	64.43±2.23	0.39
FCN1024-M-FCN64-S	82.53±4.74	0.87
GCN128-M-S	93.75±5.39	0.98
MVGCN128-M-S _{mean}	94.74±5.62	1.00
MVGCN128-M-S _{max}	95.37±5.87	1.00

Results (3): Visualization - Binary Similarity

- used the relationship prediction generated from the various models to map all 754 DTI acquisitions distanced by their predicted similarity



Results (4): Visualisation - ROI Similarity

- The MVGCN output consists of ROI-wise pairwise similarity
- Able to visualize the 10 most similar or dissimilar ROIs for PD vs. HC

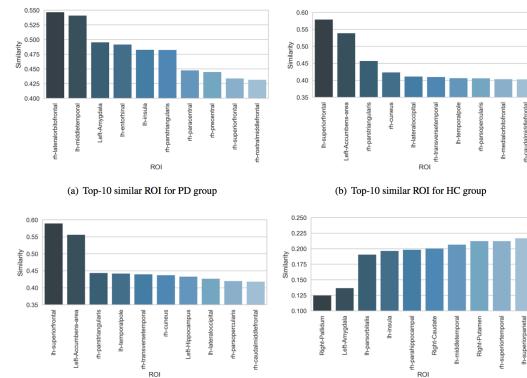


Figure 4: Visualization of the learned ROI-ROI similarities from averaged pairwise feature vectors in the certain groups. Top-10 similar or dissimilar ROIs for PD and HC groups and the corresponding values are shown in (a)-(d) respectively.

- Key Findings:**
 - Lateral orbitofrontal area, middle temporal and amygdala are three most similar ROIs for PD
 - Caudate and putamen are discriminative b/t Pd and HC

Discussion/ Conclusions:

- MVGCN allows modeling multiple brain connectivity networks (BCGs) and brain geometry graphs (BGG) based on common ROIs and tractography algorithms
 - BCGs are non-Euclidean, so standard convolution is not straightforward to use and must be specifically defined
- Multi-view approach allows the exploration of various aspects of the data
- Pairwise method increases the size of the dataset
- Straightforward interpretations of the networks were learned