A computational framework for analysis of functional MRI data for developing imaging-derived biomarkers for HIV Associated Neurocognitive Disorders

# PhD Thesis Proposal

<u>Anas Zainul Abidin</u> <u>Supervised by : Dr. Axel Wismüller</u>



# Proposed research

- Computational approach for analysis of functional MRI data
- Focus on improved characterization of information flow within the brain
- Network analysis approaches for identifying changes in the brain during neurologic diseases
- Study of HIV Associated Neurocognitive disorders

### Functional Magnetic Resonance Imaging (fMRI)

- A technique for measuring brain activity over time using MRI
- Detects the changes in oxygenated blood flow, hence referred to as BOLD fMRI
- Resting state fMRI without any stimulus has been used to capture explicit patterns in brain activity
- Reproducible resting state networks have been observed





Specific Aims



Directed Connectivity

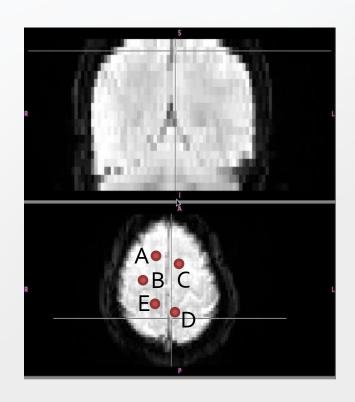


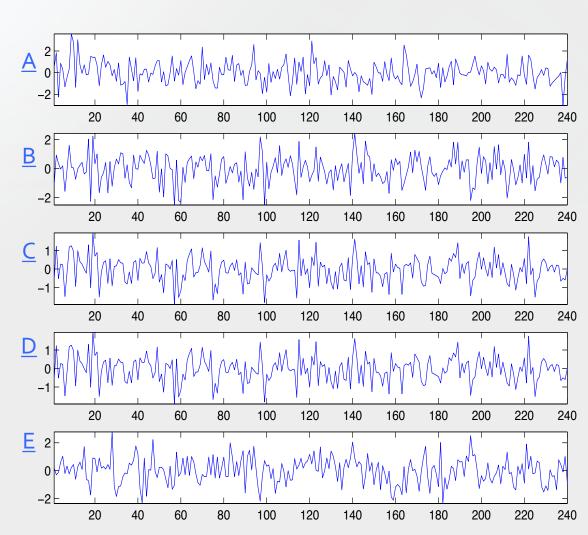
Network Analysis



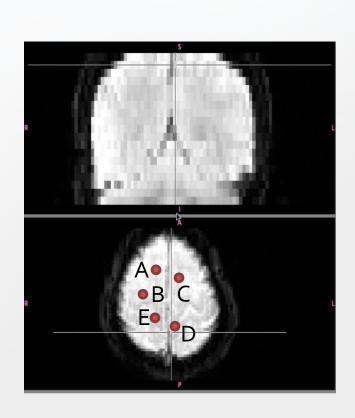
Brain Network in HAND

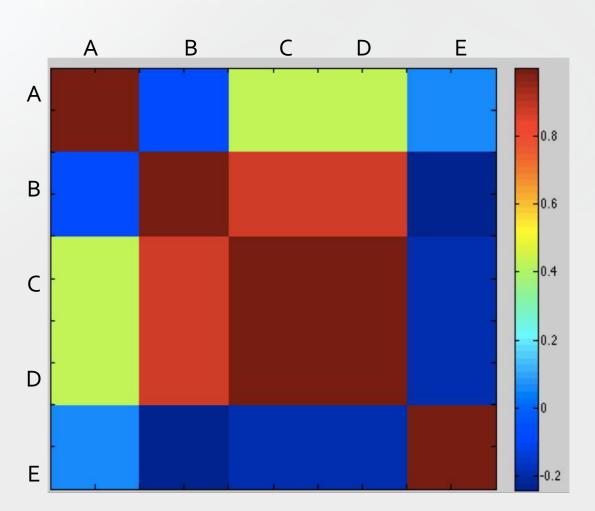
### fMRI time series





### fMRI time series



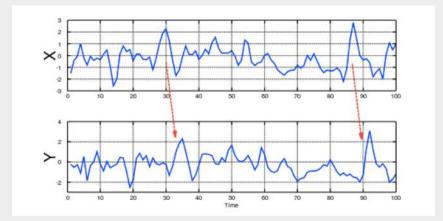


# Information Theory and Causality

- Methods in information science enable us to quantify the information flow between such time series
- Can one time series be helpful in forecasting another?

Measuring such a prediction ability indicates the presence of

a directional influence



### HIV Associated Neurocognitive Disorder (HAND)

- Reports estimate that about 1.2 million Americans and 40 million people worldwide live with HIV/AIDS
- HIV is capable of invading the brain soon after seroconversion and causing continuous neuronal injury
- Can eventually lead to deficits in multiple cognitive domains commonly referred to as HIV-associated neurocognitive disorders (HAND)
- Difficult to detect, particularly when mild impairment persists





Specific Aims



Directed Connectivity

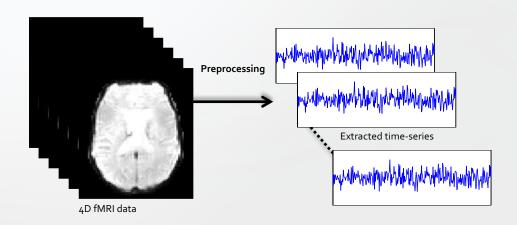


Network Analysis

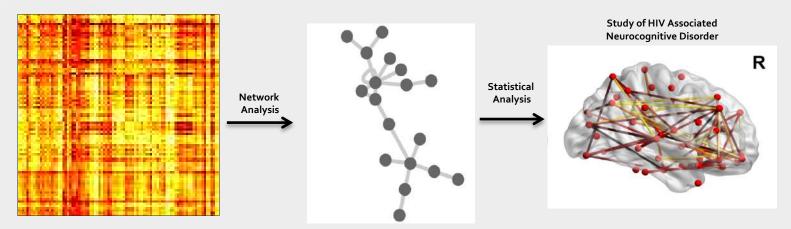


Brain Network in HAND

## Outline



### **Directed Connectivity**



# Specific Aims

- To develop and evaluate novel time-series analysis methods for exploring connectivity in high-dimensional time-series ensemble.
  - Develop the framework utilizing state-space reconstruction and radial basis function networks, for capturing non-linear pair-wise interactions in time-series.
  - Explore the applicability of multivariate time-series analysis methods for recovery of underlying network interactions using large-scale Granger causality.
  - Systematically test and evaluate the performance of the methods for adequate recovery of underlying network structure via the use of benchmark datasets and functional MRI simulations.

- Existing methods have certain disadvantages owing to *α priori* assumptions, linear models
- Can the use of non-linear techniques enhance the characterization of connectivity?
- Mutual Connectivity Analysis using Generalized Radial Basis Function network (MCA-GRBF)

- According to dynamical systems theory, the state of a system at every instant is controlled by its state variables
  - Two time-series are causally linked if they share a common state space
- fMRI time-series are not the actual state variables, but only a projection on a lower dimensional space
- Interactions can be better understood using state-space reconstruction
  - A time-delayed sequence of observations can be used to represent/understand the original state-space of systems

• We can construct a *d*-dimensional state space

For a given time-series, say x(t)

$$\mathbf{x}_{t} = (x(t), x(t+1), ..., x(t+d)), t = 1, 2, ..., l-d,$$

where, d is the embedding dimension

& l is the length of the time-series

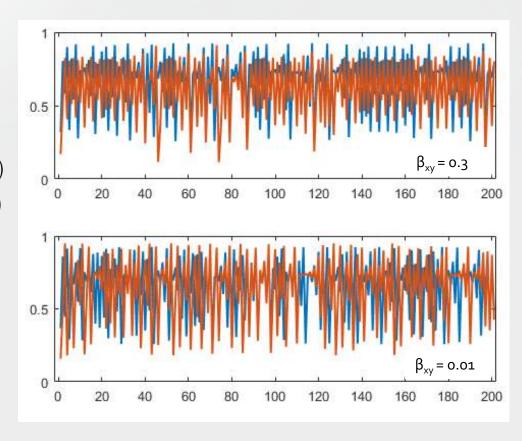


 Exact embedding dimensions of a system are not known but can be estimated quantitatively (Cao's criterion)

Lotka-Volterra 2 species models\*

$$x(t+1) = x(t)(r_x - r_x x(t) - \beta_{xy} y(t))$$
  
$$y(t+1) = y(t)(r_y - r_y y(t) - \beta_{yx} x(t))$$

where,  $r_x = 3.7, r_y = 3.8, \beta_{yx} = 0.01$ 



Lotka-Volterra 2 species models

$$x(t+1) = x(t)(r_x - r_x x(t) - \beta_{xy} y(t))$$

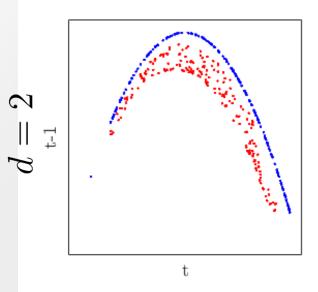
$$y(t+1) = y(t)(r_y - r_y y(t) - \beta_{yx} x(t))$$

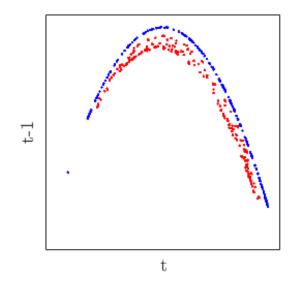
where,  $r_x = 3.7, r_y = 3.8, \beta_{yx} = 0.01$ 

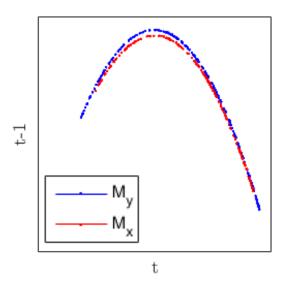
$$\beta_{xy} = 0.3$$

$$\beta_{xy} = 0.15$$

$$\beta_{xy} = 0.01$$







Lotka-Volterra 2 species models

$$x(t+1) = x(t)(r_x - r_x x(t) - \beta_{xy} y(t))$$

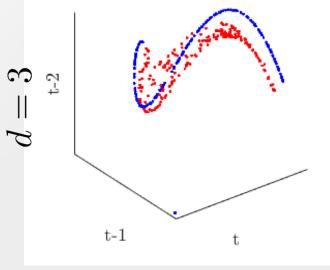
$$y(t+1) = y(t)(r_y - r_y y(t) - \beta_{yx} x(t))$$

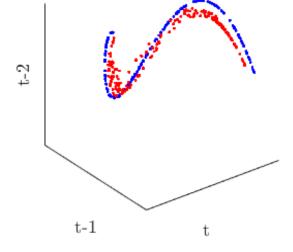
where,  $r_x = 3.7, r_y = 3.8, \beta_{yx} = 0.01$ 

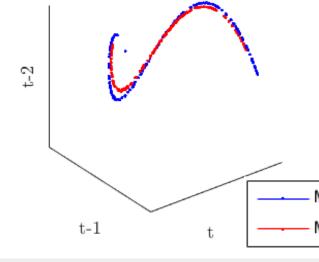
$$\beta_{xy} = 0.3$$

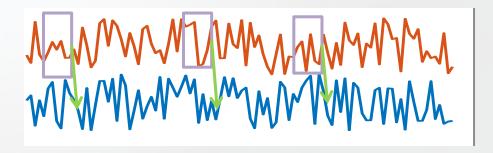
$$\beta_{xy} = 0.15$$

$$\beta_{xy} = 0.01$$





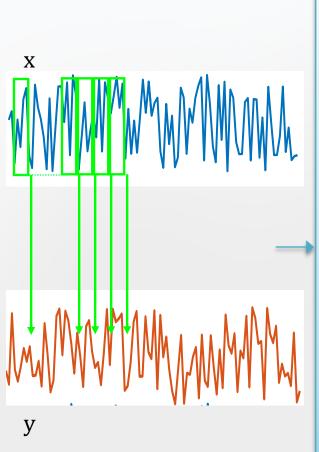


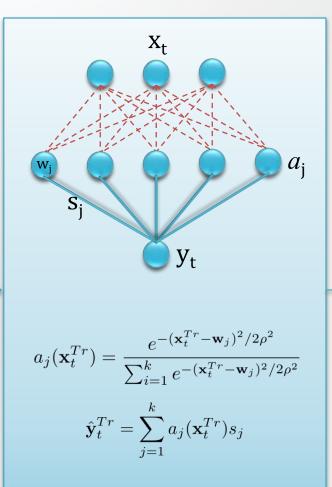


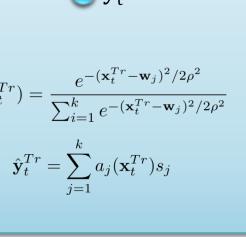
$$\mathbf{y}_t = \mathbf{f}(\mathbf{x}_t)$$

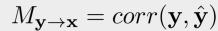
- Using a radial basis function neural network (Mutual Connectivity Analysis \*)
- An input, hidden and output layer
- Can the past of x be used to predict/cross-map a future y

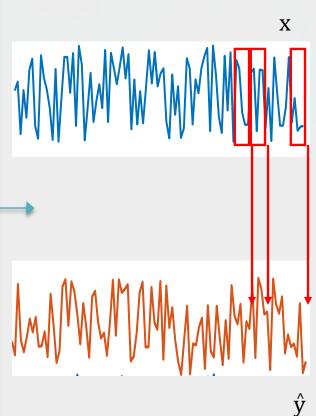
# Approach – MCA GRBF











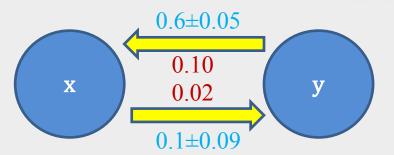
# Validation(s)

Lotka-Volterra 2 species models

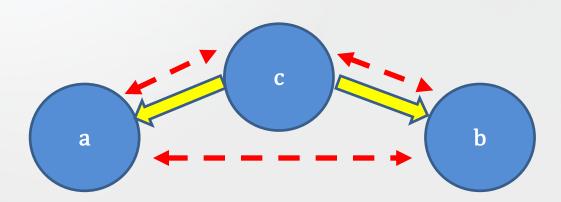
$$x(t+1) = x(t)(r_x - r_x x(t) - \beta_{xy} y(t))$$

$$y(t+1) = y(t)(r_y - r_y y(t) - \beta_{yx} x(t))$$

where, 
$$r_x = 3.7, r_y = 3.8$$



### Indirect interactions

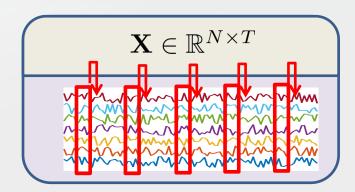


# **Granger Causality**

### **Vector Auto-Regressive (VAR) Modelling**

Consider a system, **X**, with *N* time series.

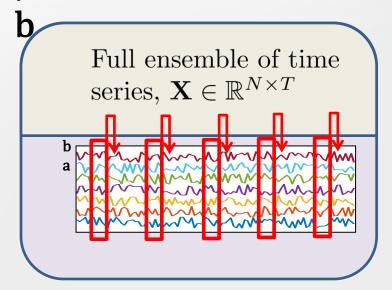
$$x(t) = \sum_{j=1}^{m} \mathbf{AR}_{j} x(t-j) + e(t)$$

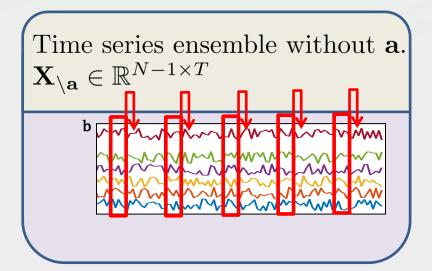


We get an estimate  $\widehat{\mathbf{X}}$  using VAR modelling.  $\mathbf{AR}_j \in \mathbb{R}^{N \times N}$  are the model parameters. m is the order of the process.

# **Granger Causality\***

If the prediction quality of **b** improves when the past of **a** is used in its prediction as compared to its quality of prediction in the absence of **a**, then **a** *Granger Causes* 





$$F_{\mathbf{a} \to \mathbf{b}} = \log \frac{\sigma(e_{\mathbf{b}_{\mathbf{X}\setminus \mathbf{a}}})}{\sigma(e_{\mathbf{b}_{\mathbf{X}}})} \text{ where, } e_{\mathbf{b}_{\mathbf{X}\setminus \mathbf{a}}} = \mathbf{b} - \hat{\mathbf{b}}_{\mathbf{X}\setminus \mathbf{a}}$$





Directed Connectivity



Network **Analysis** 

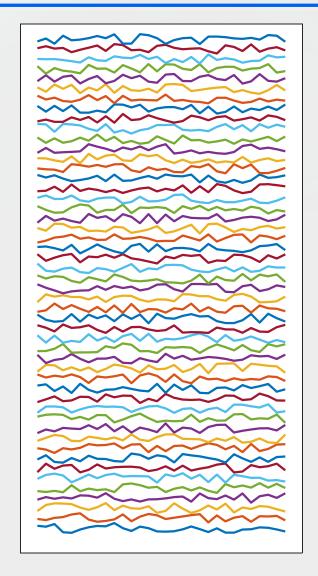


Brain Network in **HAND** 

# Limitation(s)

Background

- Cannot be extended to a full brain analysis as  $N\gg T$ 
  - As Nincreases number of parameter increase by  $N^2$
  - This results in an underdetermined problem.
- Granger causality is generally applied on fMRI datasets through
  - a priori ROI selection
  - Longer time series



# Large scale Granger Causality (IsGC)\*

- We try to circumvent the  $N \gg T$  problem
- Incorporate an additional step of invertible dimension reduction
- We can obtain the directional information scores at voxel level, in a multivariate sense



Specific Aims



Directed Connectivity

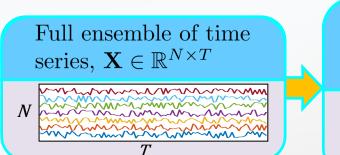


Network Analysis

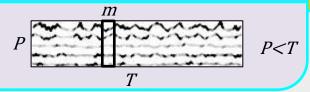


Brain Network in HAND

# Large scale Granger Causality (IsGC)

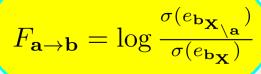


Dimension reduction using PCA on  $\mathbf{X}$ . We take  $1^{st}$  P components.  $\mathbf{Z}=\mathbf{W}\mathbf{X}$ 



VAR results in estimating  $\hat{\mathbf{Z}}$ . Obtain  $\hat{\mathbf{X}} = \mathbf{W}^{-1}\hat{\mathbf{Z}}$ 

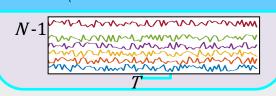




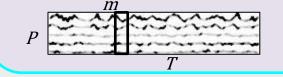


Time series ensemble without **a**.

$$\mathbf{X}_{\setminus \mathbf{a}} \in \mathbb{R}^{N-1 \times T}$$



Dimension reduction using PCA on  $\mathbf{X}_{\backslash \mathbf{a}}$ . We take  $1^{st}$  P components.  $\mathbf{Z}_{\backslash \mathbf{a}} = \mathbf{W}_{\mathbf{a}} \mathbf{X}_{\backslash \mathbf{a}}$ 



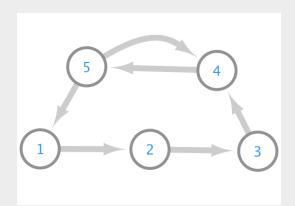
VAR results in estimating  $\hat{\mathbf{Z}}_{\mathbf{a}}$ .

Obtain 
$$\hat{\mathbf{X}_{\mathbf{a}}} = \mathbf{W}_{\mathbf{a}}^{-1} \hat{\mathbf{Z}_{\mathbf{a}}}$$

Background

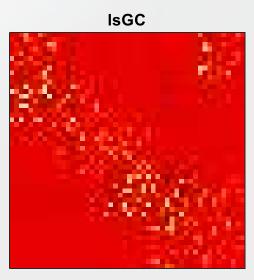
# Large scale Granger Causality (IsGC)

- To validate this IsGC we generated a system of timeseries with known structure
- 50-dimensional stationary multivariate autoregressive (MVAR) processes of order two were generated
- Corresponding AR parameters were chosen according to Baccala et. al.
- Network consists of 5 clusters



# Large scale Granger Causality (IsGC)

# Network Structure





**Thresholded** 

AUC = 0.91 csen = 0.87

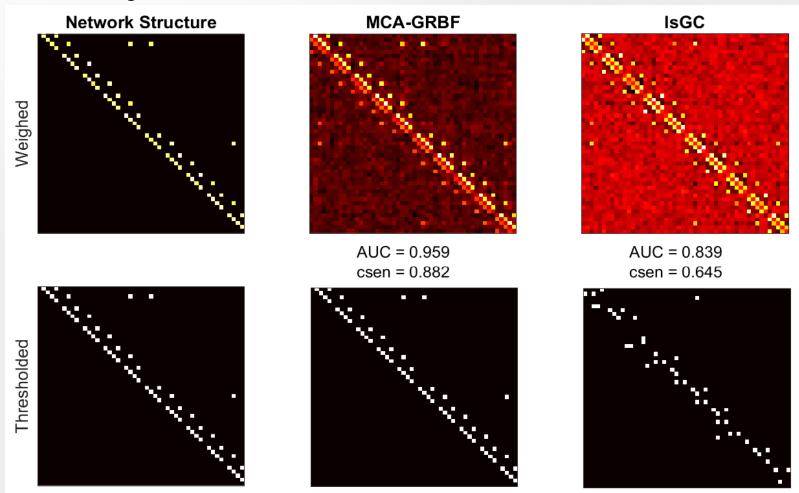
# Validation(s) – fMRI simulations

- Realistic fMRI simulations
  - As proposed by Smith et. al.\* specifically for validation of network modelling methods
  - Based on the dynamic causal modelling with non-linear hemodynamic response on top of a neural network
- Network of varied complexity can be used for simulation of BOLD time-series



# Validation(s) – fMRI simulations

Results over 50 iterations for each simulation



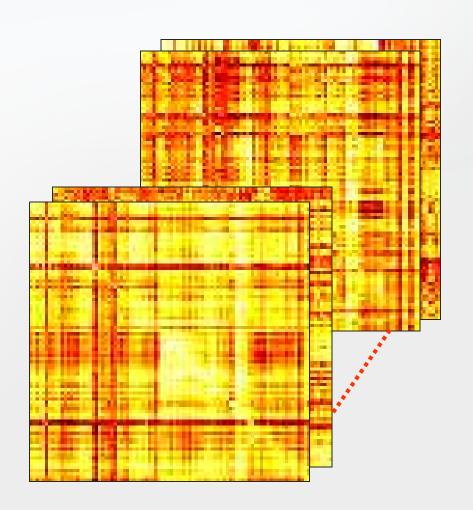
# Comparative Analysis

- Conventional Methods
  - Correlation based analysis (linear, bivariate)
  - Mutual information (non-linear, bivariate)
  - Partial correlation (linear, multi-variate)

# Specific Aims

- To model a suitable network analysis framework for identifying changes in the brain connectivity profiles.
  - Quantitatively characterize the network profiles (at a global as well as regional level) based on graph theoretic approaches with the aim of developing imaging-derived biomarkers for disease.
  - Apply and adapt the network based statistic framework for analyzing connectivity profiles to detect edge level effects of disease.
  - Develop a statistical framework for performing seedbased connectivity in a non-linear and multivariate sense for studying whole-brain connectivity maps.

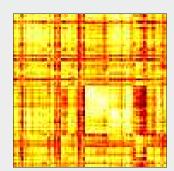
# Approach

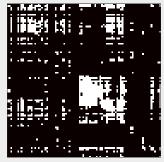


- The connectivity matrices obtained in Aim 1 can be considered as representative network graphs
- A graph consists of nodes and edges (pair-wise interactions)
- A system with N nodes results in a graph with N(N-1) edges.

# Approach – Graph Theoretic Measures

- Can characterize local or global properties
- Can be specific to nodes or the actual links
- If these are based on nodes then they are characterized by a distribution (based on the number of nodes).
- Functional Segregation
- 2. Functional Integration
- 3. Small World Properties
- 4. Network Motifs & Centrality
- 5. Resilience





- **Degree**: Number of nodes connected to an individual node
- <u>Distribution</u>: Degrees of all nodes comprise this. Mean is measure of density or the total wiring cost of the network.

# Global Graph Measures

Measure	Undirected definition	Explanation
Characteristic Path Length	$L = \frac{1}{n} \sum_{i \in N} L_i = \frac{1}{n} \sum_{i \in N} \frac{\sum_{j \in N, j \neq 1} d_{ij}}{n - 1}$ where, L <sub>i</sub> is the distance between node <i>i</i> , and all other nodes	Characteristic Path Length is a measure of functional integration and is primarily influenced by long paths. The inverse (global efficiency) measure is considered a superior measure and is primarily influenced by
Global Efficiency	$E = \frac{1}{n} \sum_{i \in N} E_i = \frac{1}{n} \sum_{i \in N} \frac{\sum_{j \in N, j \neq 1} d_{ij}^{-1}}{n - 1}$	short paths.
Clustering Coefficient	$C = \frac{1}{n} \sum_{i \in N} C_i = \frac{1}{n} \frac{2t_i}{k_i(k_i - 1)}$ where $C_i$ is the clustering coefficient of node $i$ ( $C_i = 0$ for $k_i < 2$ ).	The mean clustering coefficient represents the presence of clustered connectivity around individual nodes
Modularity	$Q = \sum_{u \in M} e_{uu} - (\sum_{v \in M} e_{uu})^2$ where, the network is fully subdivided into $M$ non overlapping modules and $e_{uv}$ is the proportion of all links that connect nodes in module $u$ to those in $v$	Measures size and composition of individual smaller networks.
Small Worldness	$S = \frac{C/C_{rand}}{L/L_{rand}}$ where, $C$ and $C_{rand}$ are the clustering coefficients, and $L$ and $L_{rand}$ are the characteristic path lengths of the respective tested network and a random network.	Describes how well a network is clustered when compared to a random graph of the same size. Smallworld networks often have $S\gg 1$

# Regional Characteristics

Measure	Undirected definition	Explanation
Degree	$k_i = \sum_{j \in N} a_{ij}$	Defined for each node <i>i</i> ; Represents the total number of links connected to a node. It is a marker of network development and resilience.
Local Efficiency	$\begin{split} E_{loc} &= \frac{1}{n} \sum_{i \in N} E_{loc,i} = \frac{1}{n} \frac{\sum_{j,h \in N, j \neq 1} a_{ij} a_{ih} [d_{jh} N_i]^{-1}}{k_i (k_i - 1)} \\ &\text{where, } E_{loc,i} \text{ is the local efficiency of node } i, \\ &\text{and } d_{jh} \left( Ni \right) \text{ is the length of the shortest path between } j \text{ and } h, \text{ that contains only neighbors of } i. \end{split}$	Characterizes the efficiency of node connections at a smaller scale.
Clustering Coefficient	$C = \frac{1}{n} \sum_{i \in N} C_i = \frac{1}{n} \frac{2t_i}{k_i(k_i - 1)}$ where $C_i$ is the clustering coefficient of node $i$ ( $C_i = 0$ for $k_i < 2$ ).	The mean clustering coefficient represents the presence of clustered connectivity around individual nodes
Betweenness Centrality	$b(i) = \frac{1}{(n-1)(n-2)} \sum_{\substack{h,j \in \mathbb{N} \\ h \neq j, h \neq i, i \neq j}} \frac{\rho_{hj}(i)}{\rho_{hj}}$ where, $\rho_{hj}$ is the number of shortest paths between $h$ and $j$ , and $\rho_{hj}(i)$ is the number of shortest paths between $h$ and $j$ passing through node $i$	Summarizes the importance of a particular node in the full network

Directed Connectivity

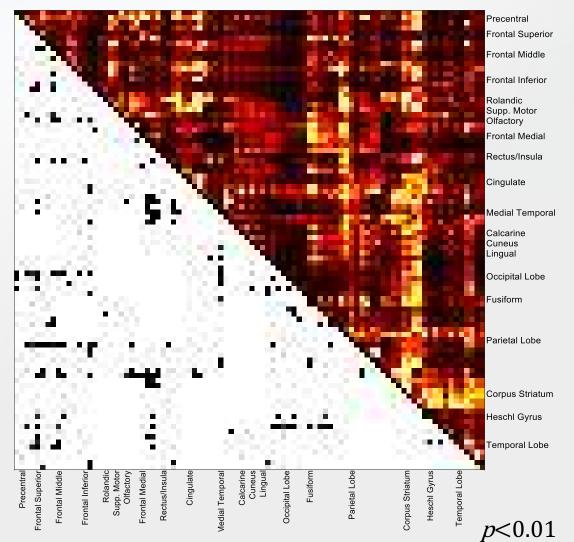


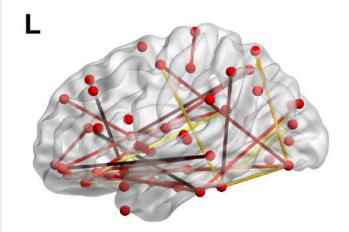
Network Analysis

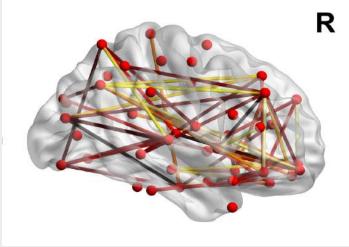


Brain Network in HAND

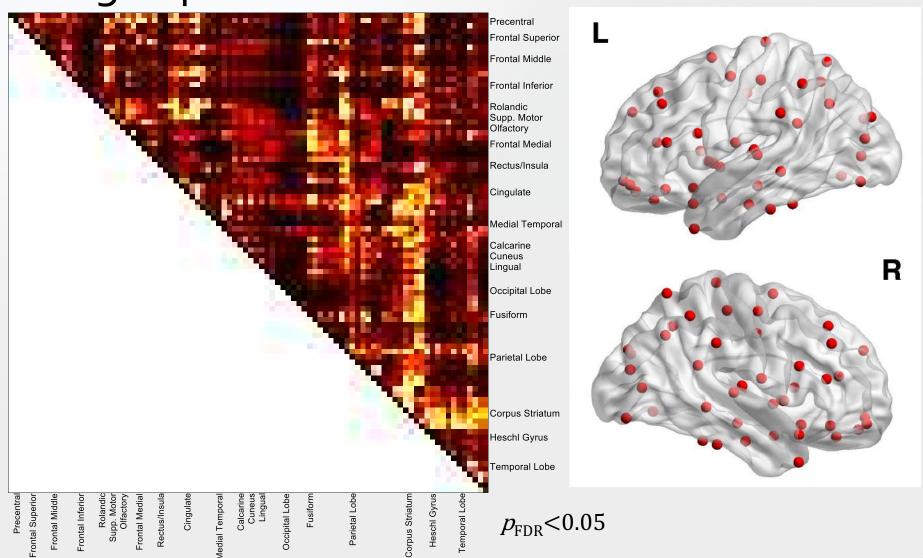
# Edge-specific differences







# Edge-specific differences



#### Network-Based Statistic\*

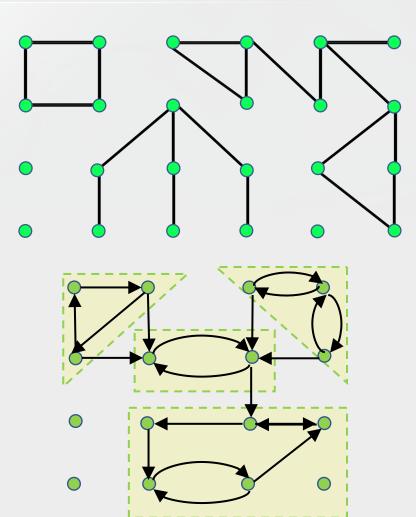
- The NBS approach offers better control over Family wise error when multiple comparisons are required.
- It seeks to identify connected nodes or clusters instead of focusing only on individual links.

### Network-Based Statistic\*

- A t-statistic is computed for each link to define a set of links above a threshold of (p=0.05).
- A non-parametric permutation approach (5000 permutations) is used to estimate the significance of the individual components (each of size  $M_i$ ).
- The null distribution of the maximal component size for each iteration is obtained
- The corrected p-value is determined by calculating the proportion of the 5000 permutations for which the maximal connected component was larger than  $M_{\rm i}$ .

#### Network-Based Statistic

- The critical step here is the detection of connected components
- In the case of undirected networks, it is done via a simple breadth first search
- For directed networks, 'strongly' connected components can be obtained using Kosaraju's algorithm



# Seed Based Connectivity

- Define a seed or a region of interest in the brain
- Compute connectivity with all other voxels
- Gives an option of quantifying the extent of connections
- A more interpretable notion of connectivity
- With methods developed in Aim 1 we can also detect the notion of directed connectivity

# Seed Based Connectivity – MCA-GRBF

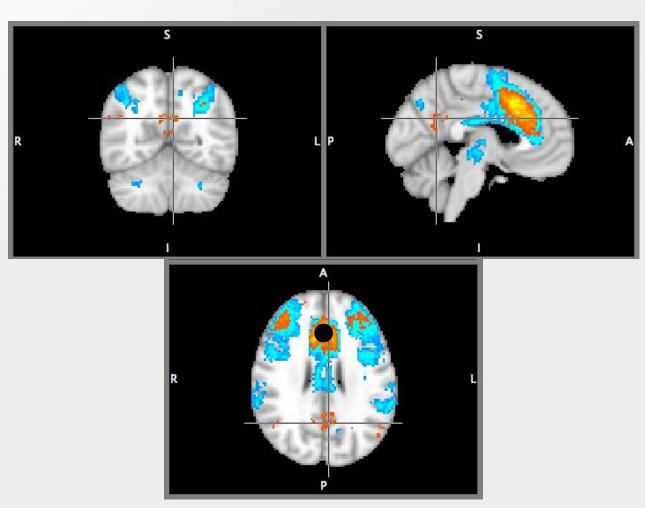
#### **Preliminary Result**

Seed – Anterior Cingulate Gyrus

Blue/Light Blue – Correlation Map

Red/Yellow – MCA-GRBF

Simple Mean Effect p<0.05, FDR corrected



# Seed Based Connectivity – MCA-GRBF

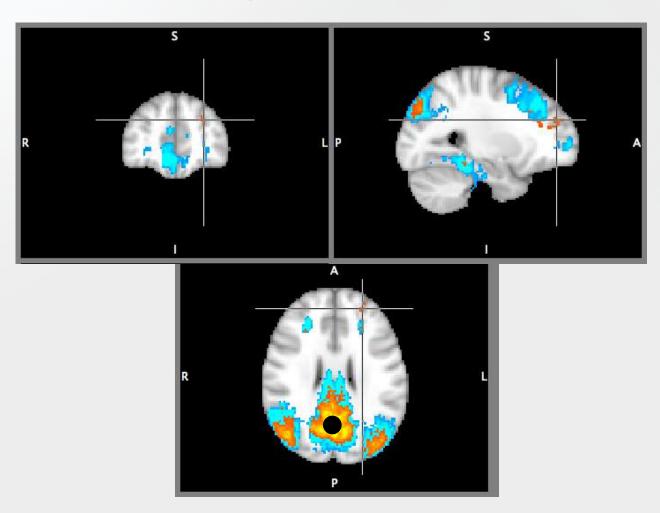
#### Preliminary Result

Seed – Posterior Cingulate Gyrus

Blue/Light Blue – Correlation Map

Red/Yellow – MCA-GRBF

Simple Mean Effect p<0.05, FDR corrected



# Seed Based Connectivity

- Conventionally, most studies are based on a correlation analysis
  - Bivariate in nature
- Directional information (seed-to-voxel as well as voxel to seed) effects can be studied
- Multivariate techniques for such an analysis are less explored
- The IsGC framework can enable a purely multivariate seed based analysis, not performed previously

# Specific Aims

- 3. To quantitatively evaluate our system in a clinical pilot study by performing an in-depth assessment of changes in resting-state brain activity in subjects with HIV infection.
  - Preprocess and analyze clinical, neuropsychological testing and neuroimaging data acquired in a pilot study for identifying subjects with HAND based on current clinical standards.
  - Evaluate relationships between neuropsychological testing scores and connectivity derived summary measures, for assessing their applicability as clinical biomarkers.
  - Assess regional changes in brain networks occurring as a result of neural injury caused by HIV infection.

## HIV Associated Neurocognitive Disorder (HAND)

- Synpato-dendritic complex has been shown to loose its densely branched structure during HIV infection
- Can occur long before symptoms manifest in an individual
- Such damage can trigger changes in global connectivity
- The efficacy of treatment paradigm (particularly cART combined anti-retroviral therapy) has helped dramatically control viral loads HIV+ individuals leading to longer life expectancy

### HIV Associated Neurocognitive Disorder (HAND)

- Before cART severe dementia could affect ~20% of the individuals
- Following cART, the incidence of severe disabling dementia has reduced however milder forms of the disorder are prevalent
- Neuropsychologic performance testing is the clinical standard for both diagnosis and monitoring
  - It can miss subtle effects of disease

## Approach – Data

- 40 subjects were recruited at the University of Rochester Medical Center, as part of a NIH funded study (Ro1-DA-034977).
- Imaging Protocol:
  - MPRAGE scan (TI = 950 ms, TE/TR = 3.87 ms/1,620 ms, 1-mm isotropic resolution)
  - 4 gradient echo (GE) EPI sequences, with the following parameters: 25 contiguous axial slices, 4mm slice thickness, TR=1650 ms, TE= 23ms, 96x96 matrix – 240 time points
- Detailed neuropsychological testing
  - executive function, information processing speed, attention and working memory, learning, memory, motor function etc.

# Demographics & Clinical characteristics

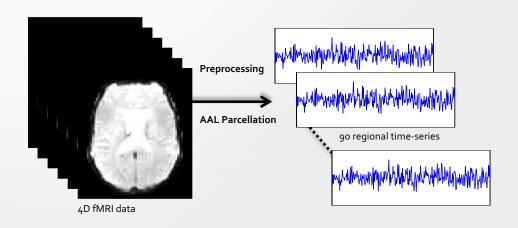
	HIV-	HIV+
Number of Patients	20	20
Age - in years	41.45 (9.98)	41.60 (15.42)
Gender (Num. female/male)	9/11	5/15
Nadir CD4 (cells/mm³)	-NA-	314.63 (211.94)
CD4 (cells/mm³)	-NA-	702.8 (464.5)
VL (log <sub>10</sub> scale)	-NA-	1.76 (1.79)
HIV - in years	-NA-	11.44 (9.11)
NP Z-scores		
Attention	0.432 (0.8)	-0.541 (1.042)
Executive	0.284 (0.982)	-0.36 (0.953)
Learning	0.363 (0.915)	-0.323 (0.878)
Memory	0.345 (1.04)	-0.254 (0.76)
Motor	0.536 (0.664)	-0.54 (0.957)
Speed of Information Processing	0.366 (0.769)	-0.42 (1.066)
Overall	2.327 (2.842)	-2.41 (3.711)
HAND Classification (%)		
WNL	-NA-	6 (30%)
ANI	-NA-	12 (60%)
MND	-NA-	2 (10%)

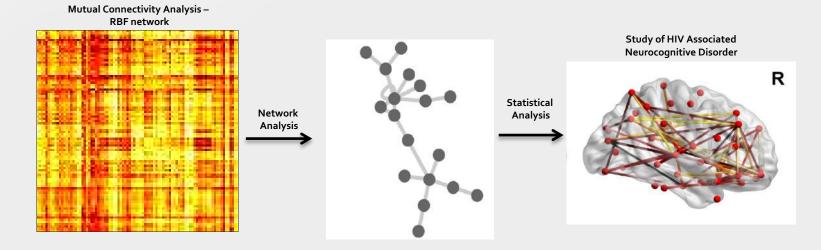
## Approach – Data

- Standard preprocessing
  - Motion correction
  - Linear detrending
  - Correction for interleaved acquisition
  - Normalization the MNI template
- Anatomic Parcellation
  - AAL template (90 regions)
  - Dosenbach template (160 regions)



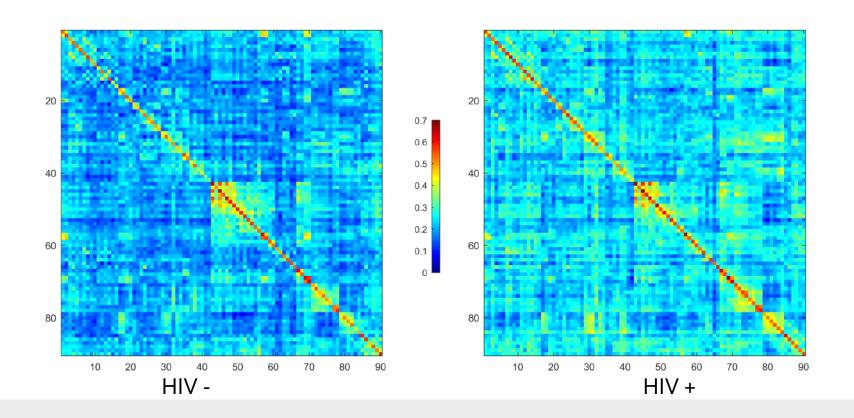
# Approach





# Non-Linear Connectivity

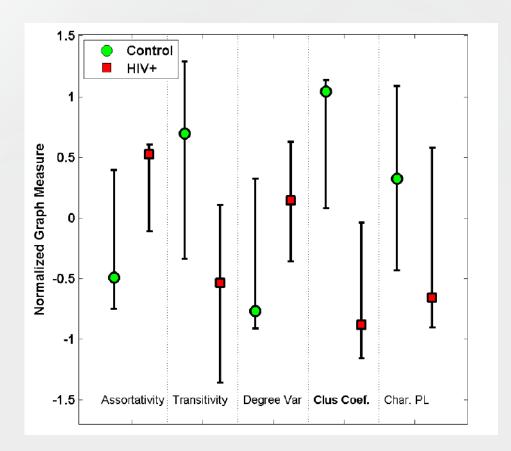
Mean Group Matrices – MCA GRBF



# Global Graph Statistics

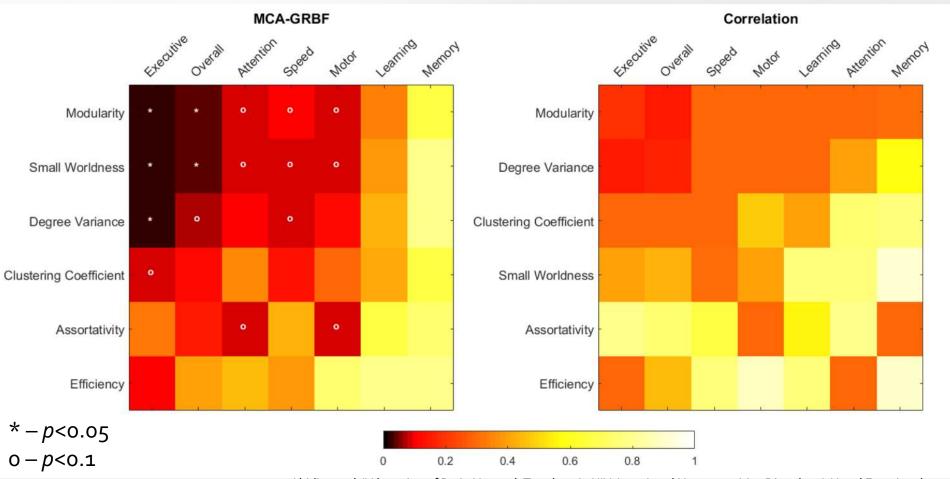
 Comparison between subjects with and without HIV infection

 Regression analysis when controlled for age and gender



# Global Graph Statistics

Associations with NP assessment scores









Directed Connectivity



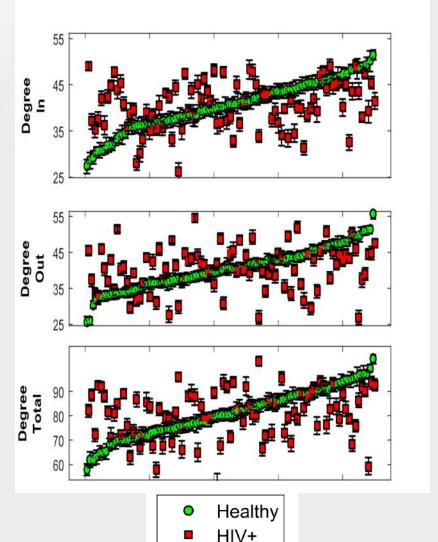
Network Analysis



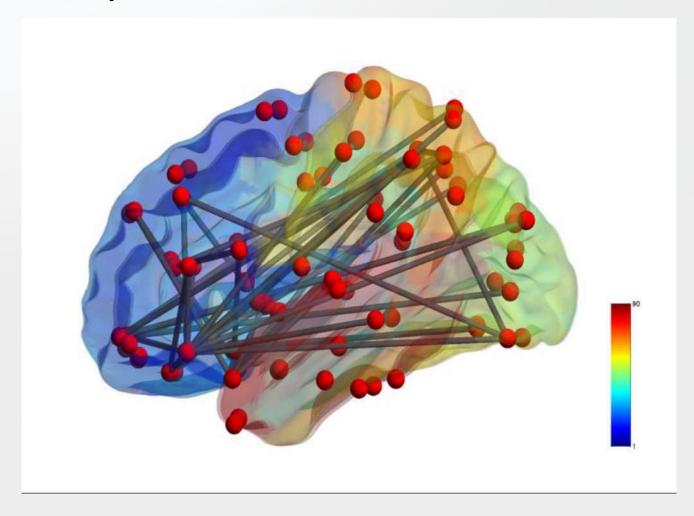
Brain Network in HAND

# Regional Properties (Preliminary Data)

- The properties of each node or region being analyzed can be quantified.
- Can the directional properties show differences?



# NBS – Expected outcome



# Seed Based Analysis

- Certain regions of the brain are more susceptible to damage by HIV
- Clinical presentation of HAND has changed due to the efficacy of cART
- Primarily affects the FSTC (fronto-striato-thalamocortical) circuits
- Candidate seed regions
  - Basal Ganglia
  - Inferior Frontal Cortex
  - Posterior Cingulate Cortex

# **Expected Goals**

- Novel time-series analysis methods for improved characterization of brain connectivity
- Statistical framework for analysis of network data
- Insights for improved diagnosis of HAND

2.

# **Current Progress**

 To develop and evaluate novel time-series analysis methods for exploring connectivity in highdimensional time-series ensemble.

Background

- Develop the framework
  utilizing state-space
  reconstruction and radial basis
  function networks, for
  capturing non-linear pair-wise
  interactions in time-series.
- Explore the applicability of multivariate time-series analysis methods for recovery of underlying network interactions using large-scale Granger causality.
- Systematically test and evaluate the performance of the methods for adequate recovery of underlying network structure via the use of benchmark datasets and functional MRI simulations.

- To model a suitable network analysis framework for identifying changes in the brain connectivity profiles.
- Quantitatively characterize the network profiles (at a global as well as regional level) based on graph theoretic approaches with the aim of developing imaging-derived biomarkers for disease.
- Apply and adapt the network based statistic framework for analyzing connectivity profiles to detect edge level effects of disease.
- Develop a statistical framework for performing seed-based connectivity in a non-linear and multivariate sense for studying wholebrain connectivity maps.

- 3. To quantitatively evaluate our system in a clinical pilot study by performing an in-depth assessment of changes in resting-state brain activity in subjects with HIV infection.
  - Preprocess and analyze clinical, neuropsychological testing and neuroimaging data acquired in a pilot study for identifying subjects with HAND based on current clinical standards.
  - Evaluate relationships
    between neuropsychological
    testing scores and
    connectivity derived
    summary measures, for
    assessing their applicability
    as clinical biomarkers.
  - Assess regional changes in brain networks occurring as a result of neural injury caused by HIV infection.

# Acknowledgements

- National Institute of Health (NIH) Award Ro1-DA-034977
- CTSI Upstate New York Translational Research Network
- Dr. Axel Wismueller
- Adora D'Souza
- Uday Chockanathan
- Dr. Mahesh Nagarajan

- Bernstein Group for Computational Neuroscience Jena, Institute of Medical Statistics, Computer Science, and Documentation Jena, Germany
  - Lutz Leistritz
- University of Rochester
  - Dr. Xing Qiu
  - Wang Lu

- Dr Kevin Parker
- Dr Jianhui Zhong
- Dr. Giovanni Schifitto



