Predictive Inference from Replicated Networks

David Dunson

Department of Statistical Science, Duke University

Snedecor Lecture, May 2018

Outline

Background & Motivation

Unsupervised approaches
- Nonparametric Bayes models
- Fast algorithms

Supervised methods
- SBR for subgraph extraction
- MrTensor for spatial networks

Outline

Datasets

- Soccer passing networks data

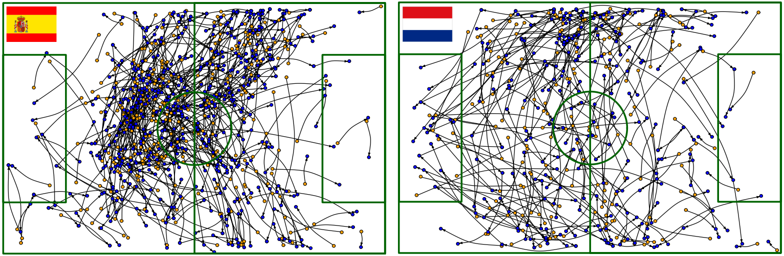


Figure: Spatial passing networks in a 2014 FIFA world cup match (Spain 1-5 Netherlands). Orange & blue nodes indicates origin-destination of pass. Team attack from left → right.

Datasets

- Soccer passing networks data

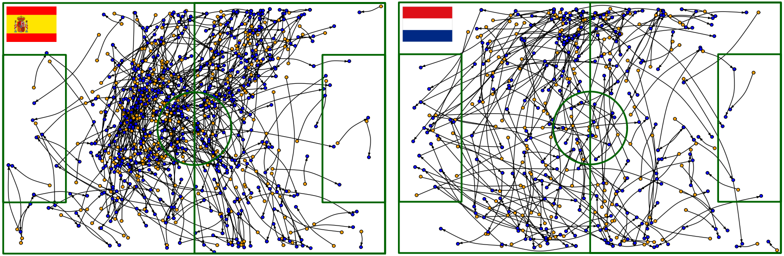


Figure: Spatial passing networks in a 2014 FIFA world cup match (Spain 1-5 Netherlands). Orange & blue nodes indicates origin-destination of pass. Team attack from left $\rightarrow$ right.

Datasets

- Soccer passing networks data

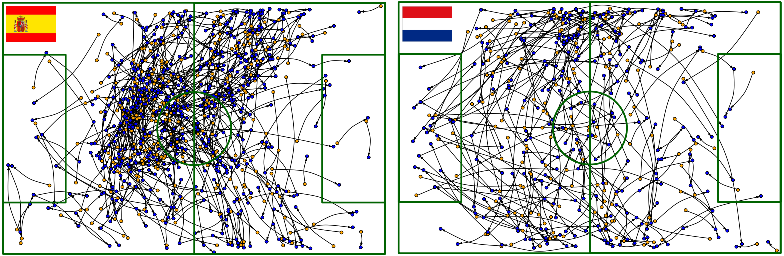


Figure: Spatial passing networks in a 2014 FIFA world cup match (Spain 1-5 Netherlands). Orange & blue nodes indicates origin-destination of pass. Team attack from left → right.

- Human Connectome Project (HCP) dataset

Datasets

- Soccer passing networks data

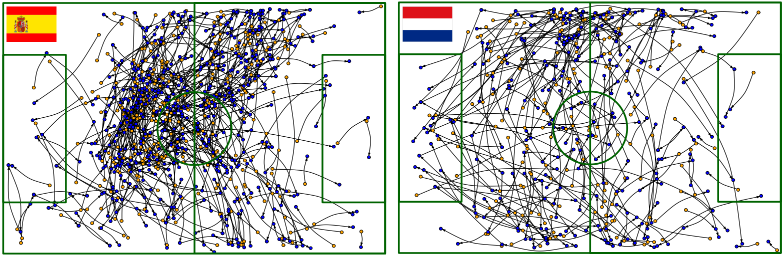


Figure: Spatial passing networks in a 2014 FIFA world cup match (Spain 1-5 Netherlands). Orange & blue nodes indicates origin-destination of pass. Team attack from left → right.

- Human Connectome Project (HCP) dataset
  - Brain imaging data for 1065 healthy adults between 22 $\sim$ 37

Datasets

- Soccer passing networks data

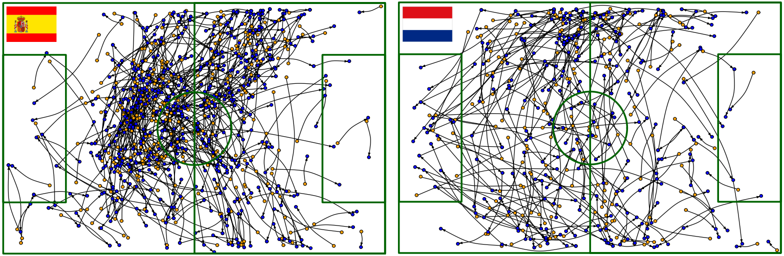


Figure: Spatial passing networks in a 2014 FIFA world cup match (Spain 1-5 Netherlands). Orange & blue nodes indicates origin-destination of pass. Team attack from left $\rightarrow$ right.

- Human Connectome Project (HCP) dataset
  - Brain imaging data for 1065 healthy adults between 22 $\sim$ 37
  - Rich information on traits for each subject (cognitive, motor, sensory, emotional, etc.)

Datasets

- Soccer passing networks data

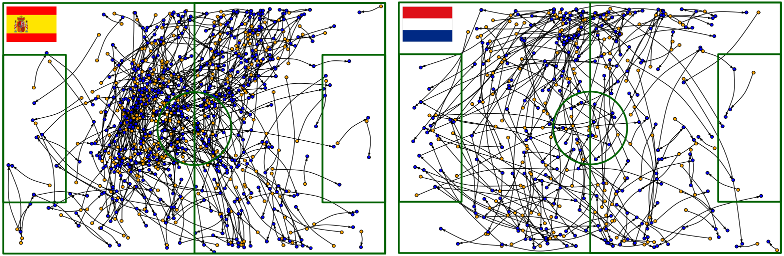
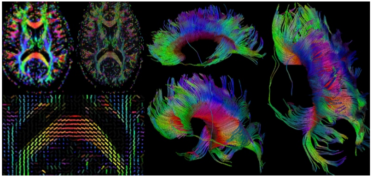


Figure: Spatial passing networks in a 2014 FIFA world cup match (Spain 1-5 Netherlands). Orange & blue nodes indicates origin-destination of pass. Team attack from left $\rightarrow$ right.

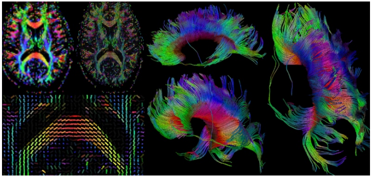
- Human Connectome Project (HCP) dataset
  - Brain imaging data for 1065 healthy adults between 22 $\sim$ 37
  - Rich information on traits for each subject (cognitive, motor, sensory, emotional, etc.)
  - Processed by Dr. Zhengwu Zhang, University of Rochester

Modeling variation in brain connectomes



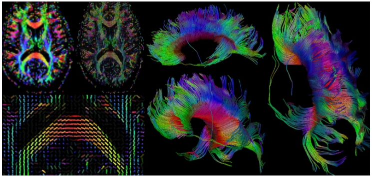
- For each individual i, we extract a structural connectome X_i from MRI data

Modeling variation in brain connectomes



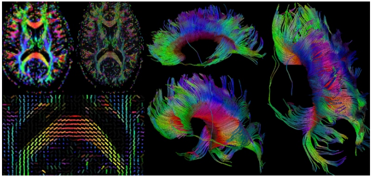
- For each individual i, we extract a structural connectome X_i from MRI data
- A single person's connectome is illustrated above & can be represented mathematically in different ways

Modeling variation in brain connectomes



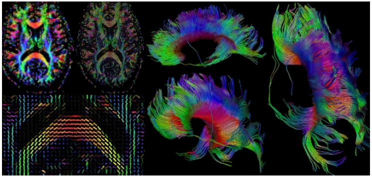
- For each individual i, we extract a structural connectome X_i from MRI data
- A single person's connectome is illustrated above & can be represented mathematically in different ways
- One simple representation is as an $R \times R$ adjacency matrix, with $R = \#$ regions of interest (ROIs)

Modeling variation in brain connectomes



- For each individual i, we extract a structural connectome X_i from MRI data
- A single person's connectome is illustrated above & can be represented mathematically in different ways
- One simple representation is as an $R \times R$ adjacency matrix, with $R = \#$ regions of interest (ROIs)
- Then, $X_{i[u,v]} = 1$ if there is any connection between regions u & v for individual i, and $X_{i[u,v]} = 0$ otherwise

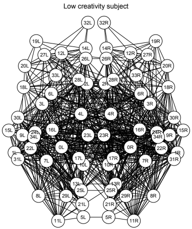
Modeling variation in brain connectomes



- For each individual i, we extract a structural connectome X_i from MRI data
- A single person's connectome is illustrated above & can be represented mathematically in different ways
- One simple representation is as an $R \times R$ adjacency matrix, with $R = \#$ regions of interest (ROIs)
- Then, $X_{i[u,v]} = 1$ if there is any connection between regions u & v for individual i, and $X_{i[u,v]} = 0$ otherwise
- Goal: study variation in X_i across individuals & interpretable predictive model for phenotypes y_i

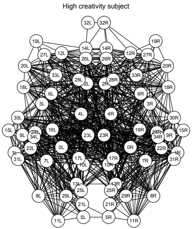
Outline

A nonparametric model of variation in brain networks



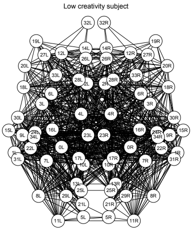
- Variation in brain networks across individuals: $X_i \sim P$, $P = ?$.

A nonparametric model of variation in brain networks



- Variation in brain networks across individuals: $X_i \sim P$, $P = ?$.
- For each brain region (r) & component (h), assign individual-specific score $\eta_{ih[r]}$

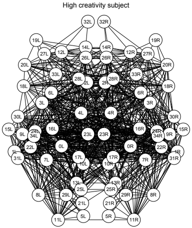
A nonparametric model of variation in brain networks



- Variation in brain networks across individuals: $X_i \sim P$, $P = ?$.
- For each brain region (r) & component (h), assign individual-specific score $\eta_{ih[r]}$
- Characterize variation among individuals with:

$$\begin{aligned}\text{logit}\{\text{pr}(X_{i[u,v]} = 1)\} &= \mu_{[u,v]} + \sum_{h=1}^{K} \lambda_{ih}\eta_{ih[u]}\eta_{ih[v]}, \\ \theta_i &= \{\lambda_{ih}, \eta_{ih}\} \sim Q, \quad Q \sim \text{DP}\end{aligned}$$

A nonparametric model of variation in brain networks

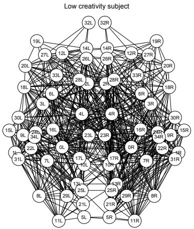


- Variation in brain networks across individuals: $X_i \sim P$, $P = ?$.
- For each brain region (r) & component (h), assign individual-specific score $\eta_{ih[r]}$
- Characterize variation among individuals with:

$$\begin{aligned}\text{logit}\{\text{pr}(X_{i[u,v]} = 1)\} &= \mu_{[u,v]} + \sum_{h=1}^{K} \lambda_{ih}\eta_{ih[u]}\eta_{ih[v]}, \\ \theta_i &= \{\lambda_{ih}, \eta_{ih}\} \sim Q, \quad Q \sim \text{DP}\end{aligned}$$

- Using Bayesian nonparametrics, allow Q (& P) to be unknown

A nonparametric model of variation in brain networks

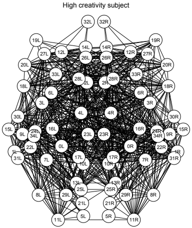


- Variation in brain networks across individuals: $X_i \sim P$, $P = ?$.
- For each brain region (r) & component (h), assign individual-specific score $\eta_{ih[r]}$
- Characterize variation among individuals with:

$$\begin{aligned}
\text{logit}\{\text{pr}(X_{i[u,v]} = 1)\} &= \mu_{[u,v]} + \sum_{h=1}^{K} \lambda_{ih}\eta_{ih[u]}\eta_{ih[v]}, \\
\theta_i &= \{\lambda_{ih}, \eta_{ih}\} \sim Q, \quad Q \sim \text{DP}
\end{aligned}$$

- Using Bayesian nonparametrics, allow Q (& P) to be unknown

A nonparametric model of variation in brain networks

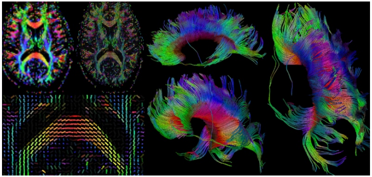


- Variation in brain networks across individuals: $X_i \sim P$, $P = ?$.
- For each brain region (r) & component (h), assign individual-specific score $\eta_{ih[r]}$
- Characterize variation among individuals with:

$$
\begin{aligned}
\text{logit}\{\text{pr}(X_{i[u,v]} = 1)\} &= \mu_{[u,v]} + \sum_{h=1}^{K} \lambda_{ih} \eta_{ih[u]} \eta_{ih[v]}, \\
\theta_i &= \{\lambda_{ih}, \eta_{ih}\} \sim Q, \quad Q \sim \text{DP}
\end{aligned}
$$

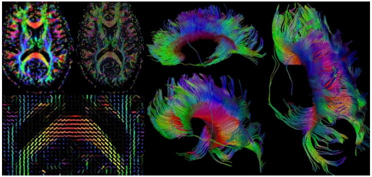
- Using Bayesian nonparametrics, allow Q (& P) to be unknown

Bayesian inferences



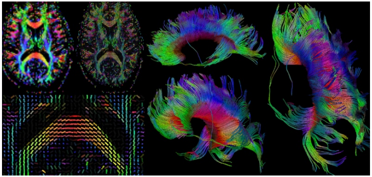
- Common *dictionary* representing the brain structure

Bayesian inferences



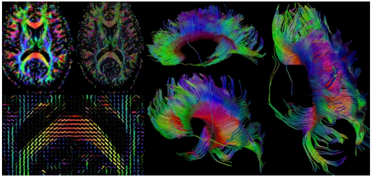
- Common *dictionary* representing the brain structure
- Pop dist of weights on dictionary elements varies with traits

Bayesian inferences



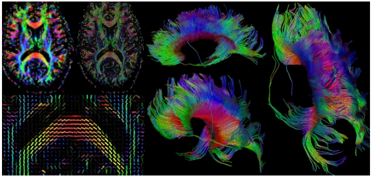
- Common *dictionary* representing the brain structure
- Pop dist of weights on dictionary elements varies with traits
- Induces a nonparametric model of variation in brain structure with phenotypes $(X_i|Y_i = y) \sim P_y$

Bayesian inferences



- Common *dictionary* representing the brain structure
- Pop dist of weights on dictionary elements varies with traits
- Induces a nonparametric model of variation in brain structure with phenotypes $(X_i|Y_i = y) \sim P_y$
- Allows global & local testing for relationships with traits (*Alzheimer's disease, creative reasoning, IQ*)

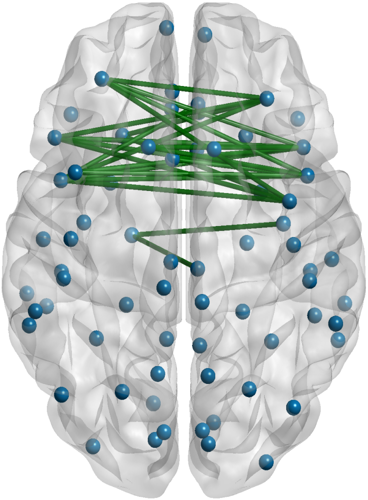
Bayesian inferences



- Common *dictionary* representing the brain structure
- Pop dist of weights on dictionary elements varies with traits
- Induces a nonparametric model of variation in brain structure with phenotypes $(X_i|Y_i = y) \sim P_y$
- Allows global & local testing for relationships with traits (*Alzheimer's disease, creative reasoning, IQ*)
- Induces predictive model for traits given brain structure:
$$f(y|X_i = x) = \frac{f_0(y)P_y(x)}{\int_{\mathcal{Y}} f_0(y)P_y(x)dy}.$$

Application to creativity

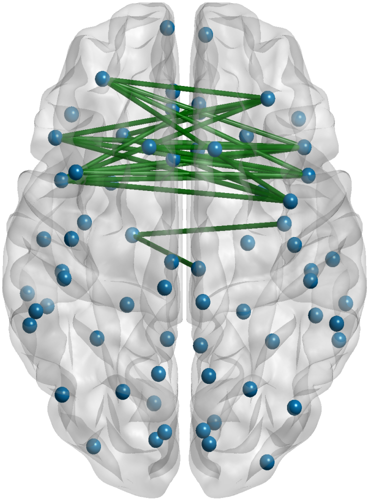
Results from local testing



- Apply model to brain networks of 36 subjects (19 with high creativity, 17 with low creativity—measured via CCI).

Application to creativity

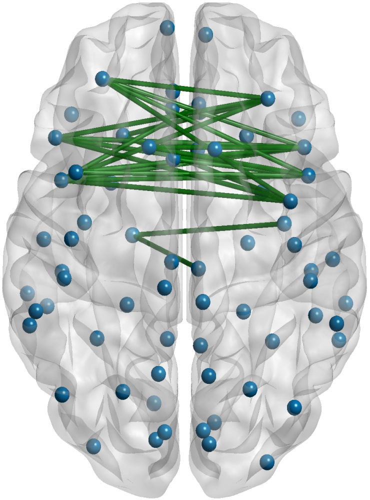
Results from local testing



- Apply model to brain networks of 36 subjects (19 with high creativity, 17 with low creativity—measured via CCI).
- $\hat{\text{pr}}(H_1 \mid \text{data}) = 0.995.$

Application to creativity

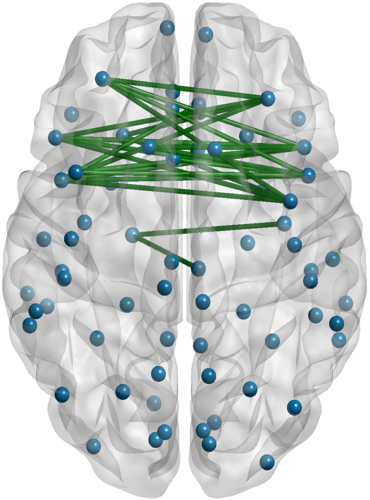
Results from local testing



- Apply model to brain networks of 36 subjects (19 with high creativity, 17 with low creativity—measured via CCI).
- $\hat{\text{pr}}(H_1 \mid \text{data}) = 0.995$.
- High creative individuals display a significantly higher propensity to form inter-hemispheric connections.

Application to creativity

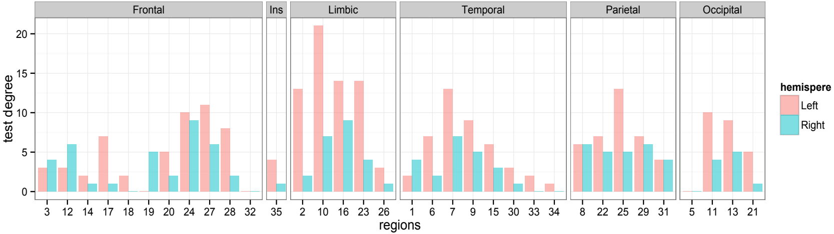
Results from local testing



- Apply model to brain networks of 36 subjects (19 with high creativity, 17 with low creativity—measured via CCI).
- $\hat{\text{pr}}(H_1 \mid \text{data}) = 0.995$.
- High creative individuals display a significantly higher propensity to form inter-hemispheric connections.
- Differences in frontal lobe are consistent with recent findings from fMRI studies analyzing regional activity in isolation.

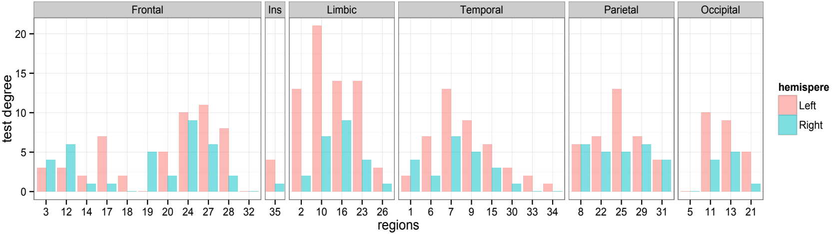
Application to Alzheimer's

- Apply model to brain networks of 92 subjects (42 with AD and 50 age–matched individuals having normal aging)



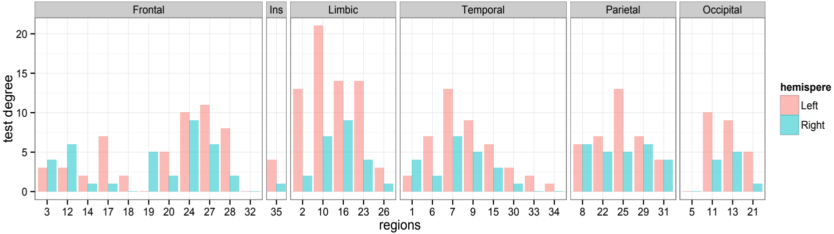
Application to Alzheimer's

- Apply model to brain networks of 92 subjects (42 with AD and 50 age–matched individuals having normal aging)
- $\hat{\text{pr}}(H_1 \mid \text{data}) > 0.99$



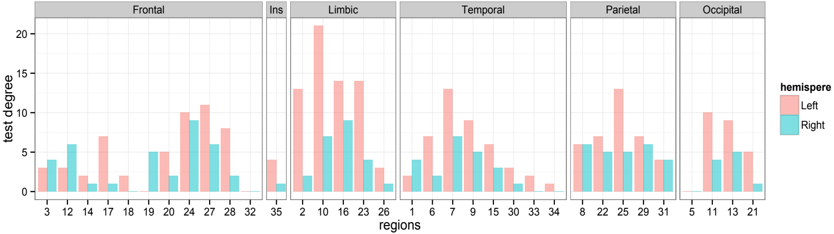
Application to Alzheimer's

- Apply model to brain networks of 92 subjects (42 with AD and 50 age–matched individuals having normal aging)
- $\hat{\text{pr}}(H_1 \mid \text{data}) > 0.99$
- AD people have less intra-hemispheric links in left hemisphere, but there is also a reduction in inter-hemispheric links

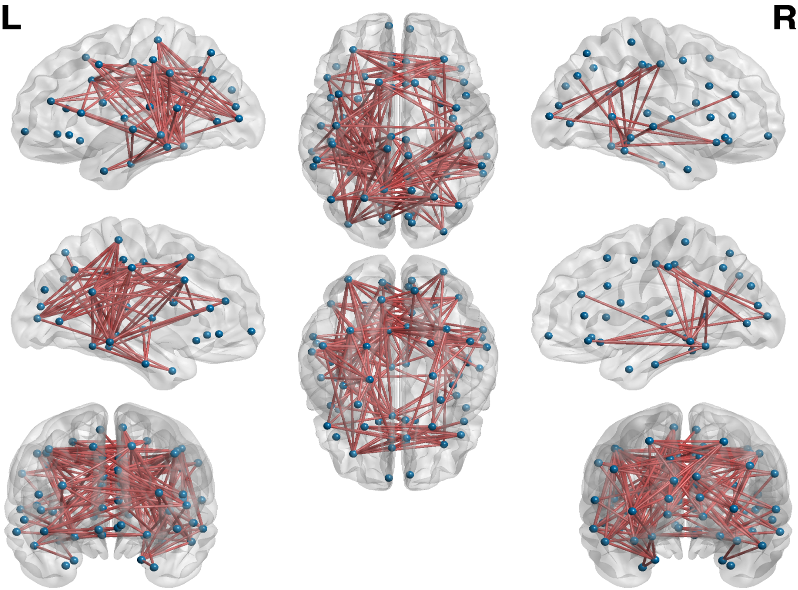


Application to Alzheimer's

- Apply model to brain networks of 92 subjects (42 with AD and 50 age–matched individuals having normal aging)
- $\hat{\text{pr}}(H_1 \mid \text{data}) > 0.99$
- AD people have less intra-hemispheric links in left hemisphere, but there is also a reduction in inter-hemispheric links
- Main differences in the connectivity of the regions in the left limbic lobe



Results for Alzheimer's

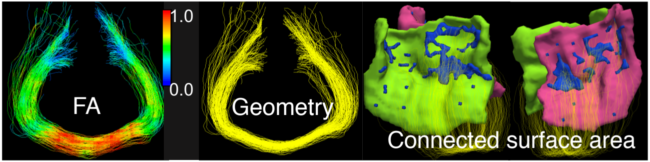


Tensor PCA & Results

- Predicting traits based on brain structural connectomes is extremely interesting.

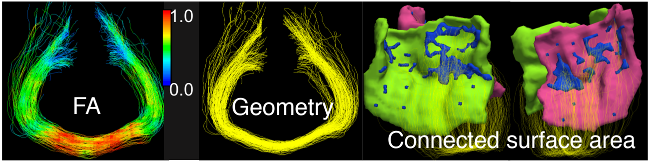
Tensor PCA & Results

- Predicting traits based on brain structural connectomes is extremely interesting.
- To better characterize brain connectomes, we extract different features from streamlines connecting two ROIs: geometry-, diffusion-, and endpoint-related features.



Tensor PCA & Results

- Predicting traits based on brain structural connectomes is extremely interesting.
- To better characterize brain connectomes, we extract different features from streamlines connecting two ROIs: geometry-, diffusion-, and endpoint-related features.



- Connectomes from multiple subjects can form semi-symmetric 3-way or 4-way tensors. Tensor PCA maps connectomes to low-dimensional vectors:

$$\mathcal{X} \approx \sum_{k=1}^{K} d_k \boldsymbol{v}_k \circ \boldsymbol{v}_k \circ \boldsymbol{u}_k. \quad (1)$$

Tensor PCA & Results

Visualization: connectome vectors of subjects with high & low trait scores.

Tensor PCA & Results

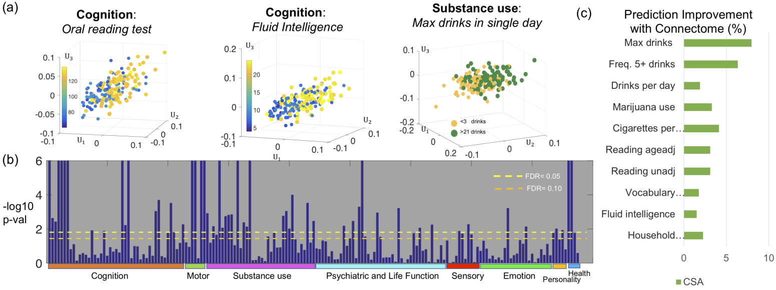
Visualization: connectome vectors of subjects with high & low trait scores.

Hypothesis testing: test distribution difference between subjects with high & low traits.

Tensor PCA & Results

Visualization: connectome vectors of subjects with high & low trait scores.
Hypothesis testing: test distribution difference between subjects with high & low traits.
Prediction: trait prediction improvement with connectomes in addition to age & gender.

Visualization: connectome vectors of subjects with high & low trait scores.
Hypothesis testing: test distribution difference between subjects with high & low traits.
Prediction: trait prediction improvement with connectomes in addition to age & gender.



Tensor PCA & Results

Connectome change: how the connectome varies with trait?

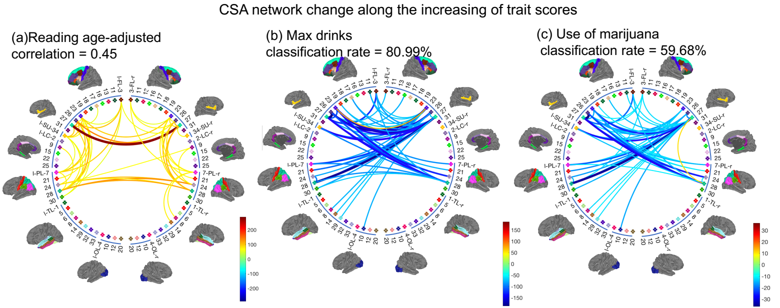
Tensor PCA & Results

Connectome change: how the connectome varies with trait?

Addressed by canonical correlation analysis (for continuous traits) and linear discriminant analysis (for categorical traits).

Connectome change: how the connectome varies with trait?

Addressed by canonical correlation analysis (for continuous traits) and linear discriminant analysis (for categorical traits).



Outline

Background & Motivation

Unsupervised approaches

- Nonparametric Bayes models
- Fast algorithms

Supervised methods

- SBR for subgraph extraction
- MrTensor for spatial networks

Identifying brain subnetworks predictive of traits

- Neuroscientists tend to be very interested in identifying subnetworks

Identifying brain subnetworks predictive of traits

- Neuroscientists tend to be very interested in identifying subnetworks
- Identify networks among a small subset of the brain ROIs

Identifying brain subnetworks predictive of traits

- Neuroscientists tend to be very interested in identifying subnetworks
- Identify networks among a small subset of the brain ROIs
- Individuals over- or under-expressing a subnet have higher or lower values of trait y_i on average

Identifying brain subnetworks predictive of traits

- Neuroscientists tend to be very interested in identifying subnetworks
- Identify networks among a small subset of the brain ROIs
- Individuals over- or under-expressing a subnet have higher or lower values of trait y_i on average
- To identify such subnetworks, start with *Symmetric Bilinear Regression (SBR)*:

$$E(y_i \mid X_i) = \alpha + \langle \theta, X_i \rangle \,,$$

where $\langle \theta, X \rangle = \text{trace}(\theta^\top X) = \text{vec}(\theta)^\top \text{vec}(X)$

Identifying brain subnetworks predictive of traits

- Neuroscientists tend to be very interested in identifying subnetworks
- Identify networks among a small subset of the brain ROIs
- Individuals over- or under-expressing a subnet have higher or lower values of trait y_i on average
- To identify such subnetworks, start with *Symmetric Bilinear Regression (SBR)*:

$$E(y_i \mid X_i) = \alpha + \langle \theta, X_i \rangle \, ,$$

  where $\langle \theta, X \rangle = \text{trace}(\theta^\top X) = \text{vec}(\theta)^\top \text{vec}(X)$
- X_i is symmetric $\rightarrow$ θ is symmetric $\rightarrow$ large p small n (# parameters $= 1 + R(R-1)/2$; e.g. $R = 68 \rightarrow 2279 > n \approx 1000$)

Rank-K Symmetric Bilinear Regression

Suppose θ admits a rank-K CP decomposition

$$\theta = \sum_{h=1}^{K} \lambda_h \boldsymbol{\beta}_h \boldsymbol{\beta}_h^\top \tag{2}$$

with sparsity penalty on $\{\lambda_h \boldsymbol{\beta}_h \boldsymbol{\beta}_h^\top\}_{h=1}^{K}$.

Rank-K Symmetric Bilinear Regression

Suppose θ admits a rank-K CP decomposition

$$\theta = \sum_{h=1}^{K} \lambda_h \boldsymbol{\beta}_h \boldsymbol{\beta}_h^\top \tag{2}$$

with sparsity penalty on $\{\lambda_h \boldsymbol{\beta}_h \boldsymbol{\beta}_h^\top\}_{h=1}^{K}$. The model becomes

$$E(y_i \mid W_i) = \alpha + \left\langle \sum_{h=1}^{K} \lambda_h \boldsymbol{\beta}_h \boldsymbol{\beta}_h^\top, X_i \right\rangle = \alpha + \sum_{h=1}^{K} \lambda_h \boldsymbol{\beta}_h^\top X_i \boldsymbol{\beta}_h \tag{3}$$

Rank-K Symmetric Bilinear Regression

Suppose θ admits a rank-K CP decomposition

$$\theta = \sum_{h=1}^{K} \lambda_h \boldsymbol{\beta}_h \boldsymbol{\beta}_h^\top \tag{2}$$

with sparsity penalty on $\{\lambda_h \boldsymbol{\beta}_h \boldsymbol{\beta}_h^\top\}_{h=1}^{K}$. The model becomes

$$E(y_i \mid W_i) = \alpha + \left\langle \sum_{h=1}^{K} \lambda_h \boldsymbol{\beta}_h \boldsymbol{\beta}_h^\top, X_i \right\rangle = \alpha + \sum_{h=1}^{K} \lambda_h \boldsymbol{\beta}_h^\top X_i \boldsymbol{\beta}_h \tag{3}$$

- Reduce parameters from $(1 + R(R-1)/2)$ to $(1 + R + KR)$, $K \ll V$

Rank-K Symmetric Bilinear Regression

Suppose θ admits a rank-K CP decomposition

$$\theta = \sum_{h=1}^{K} \lambda_h \boldsymbol{\beta}_h \boldsymbol{\beta}_h^\top \tag{2}$$

with sparsity penalty on $\{\lambda_h \boldsymbol{\beta}_h \boldsymbol{\beta}_h^\top\}_{h=1}^K$. The model becomes

$$E(y_i \mid W_i) = \alpha + \left\langle \sum_{h=1}^{K} \lambda_h \boldsymbol{\beta}_h \boldsymbol{\beta}_h^\top, X_i \right\rangle = \alpha + \sum_{h=1}^{K} \lambda_h \boldsymbol{\beta}_h^\top X_i \boldsymbol{\beta}_h \tag{3}$$

- Reduce parameters from $(1 + R(R-1)/2)$ to $(1 + R + KR)$, $K \ll V$
- Maintain flexibility: if set $K = R(R-1)/2$ and $\{\boldsymbol{\beta}_h\}_{h=1}^K = \{\boldsymbol{e}_u + \boldsymbol{e}_v\}_{u<v}$, (3) $\Leftrightarrow$ unstructured linear model.

Rank-K Symmetric Bilinear Regression

Suppose θ admits a rank-K CP decomposition

$$\theta = \sum_{h=1}^{K} \lambda_h \boldsymbol{\beta}_h \boldsymbol{\beta}_h^\top \tag{2}$$

with sparsity penalty on $\{\lambda_h \boldsymbol{\beta}_h \boldsymbol{\beta}_h^\top\}_{h=1}^K$. The model becomes

$$E(y_i \mid W_i) = \alpha + \left\langle \sum_{h=1}^{K} \lambda_h \boldsymbol{\beta}_h \boldsymbol{\beta}_h^\top, X_i \right\rangle = \alpha + \sum_{h=1}^{K} \lambda_h \boldsymbol{\beta}_h^\top X_i \boldsymbol{\beta}_h \tag{3}$$

- Reduce parameters from $(1 + R(R-1)/2)$ to $(1 + R + KR)$, $K \ll V$
- Maintain flexibility: if set $K = R(R-1)/2$ and $\{\boldsymbol{\beta}_h\}_{h=1}^K = \{\boldsymbol{e}_u + \boldsymbol{e}_v\}_{u<v}$, (3) $\Leftrightarrow$ unstructured linear model.
- Interpretation: nonzero entries in each $\lambda_h \boldsymbol{\beta}_h \boldsymbol{\beta}_h^\top$ identify a clique subgraph.

Estimation

Block-relaxation algorithm for tensor regression (Zhou et al., 2013) is not applicable due to symmetry constraint.

Estimation

Block-relaxation algorithm for tensor regression (Zhou et al., 2013) is not applicable due to symmetry constraint.

Elementwise L_1 Regularization

$$\frac{1}{2n}\sum_{i=1}^{n}\left(y_i - \alpha - \sum_{h=1}^{K}\lambda_h \boldsymbol{\beta}_h^\top X_i \boldsymbol{\beta}_h\right)^2 + \gamma \sum_{h=1}^{K}|\lambda_h| \sum_{u=1}^{R}\sum_{v<u}|\beta_{hu}\beta_{hv}| \tag{4}$$

- Avoid scaling problems between λ_h and $\boldsymbol{\beta}_h$ compared to simply penalizing $\sum_{h=1}^{K}\|\boldsymbol{\beta}_h\|_1$ → sufficient to identify each matrix $\lambda_h\boldsymbol{\beta}_h\boldsymbol{\beta}_h^\top$

Estimation

Block-relaxation algorithm for tensor regression (Zhou et al., 2013) is not applicable due to symmetry constraint.

Elementwise L_1 Regularization

$$\frac{1}{2n}\sum_{i=1}^{n}\left(y_i - \alpha - \sum_{h=1}^{K}\lambda_h \boldsymbol{\beta}_h^\top X_i \boldsymbol{\beta}_h\right)^2 + \gamma \sum_{h=1}^{K}|\lambda_h| \sum_{u=1}^{R}\sum_{v<u}|\beta_{hu}\beta_{hv}| \tag{4}$$

- Avoid scaling problems between λ_h and $\boldsymbol{\beta}_h$ compared to simply penalizing $\sum_{h=1}^{K}\|\boldsymbol{\beta}_h\|_1$ → sufficient to identify each matrix $\lambda_h \boldsymbol{\beta}_h \boldsymbol{\beta}_h^\top$
- A simple & efficient coordinate descent algorithm can be derived having analytic updates

Estimation

Block-relaxation algorithm for tensor regression (Zhou et al., 2013) is not applicable due to symmetry constraint.

Elementwise L_1 Regularization

$$\frac{1}{2n}\sum_{i=1}^{n}\left(y_i - \alpha - \sum_{h=1}^{K}\lambda_h \boldsymbol{\beta}_h^\top X_i \boldsymbol{\beta}_h\right)^2 + \gamma \sum_{h=1}^{K}|\lambda_h| \sum_{u=1}^{R}\sum_{v<u}|\beta_{hu}\beta_{hv}| \qquad (4)$$

- Avoid scaling problems between λ_h and $\boldsymbol{\beta}_h$ compared to simply penalizing $\sum_{h=1}^{K}\|\boldsymbol{\beta}_h\|_1$ $\rightarrow$ sufficient to identify each matrix $\lambda_h\boldsymbol{\beta}_h\boldsymbol{\beta}_h^\top$
- A simple & efficient coordinate descent algorithm can be derived having analytic updates
- Can choose K as an upper bound & zero out unnecessary components

Estimation

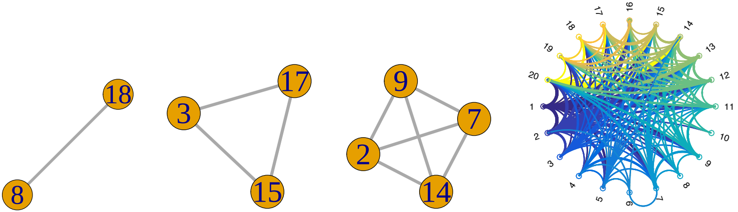
Block-relaxation algorithm for tensor regression (Zhou et al., 2013) is not applicable due to symmetry constraint.

Elementwise L_1 Regularization

$$\frac{1}{2n}\sum_{i=1}^{n}\left(y_i - \alpha - \sum_{h=1}^{K}\lambda_h \boldsymbol{\beta}_h^\top X_i \boldsymbol{\beta}_h\right)^2 + \gamma \sum_{h=1}^{K}|\lambda_h| \sum_{u=1}^{R}\sum_{v<u}|\beta_{hu}\beta_{hv}| \qquad (4)$$

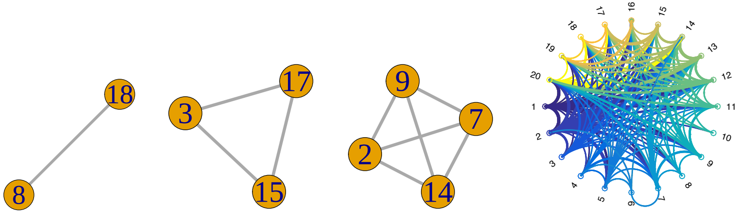
- Avoid scaling problems between λ_h and $\boldsymbol{\beta}_h$ compared to simply penalizing $\sum_{h=1}^{K} \|\boldsymbol{\beta}_h\|_1$ $\rightarrow$ sufficient to identify each matrix $\lambda_h \boldsymbol{\beta}_h \boldsymbol{\beta}_h^\top$
- A simple & efficient coordinate descent algorithm can be derived having analytic updates
- Can choose K as an upper bound & zero out unnecessary components
- Speedup: organize iterations around the nonzero parameters after a few complete cycles (Friedman et al., 2010).

Simulation



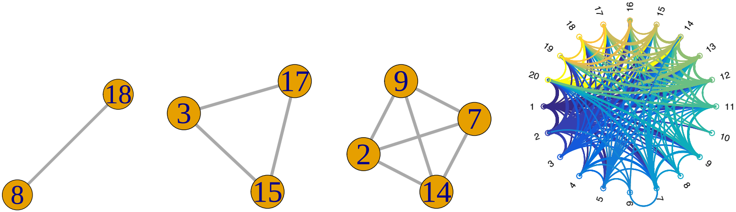
- Considered a variety of data generating processes for $(X_i, y_i), i = 1, \ldots, n$.

Simulation



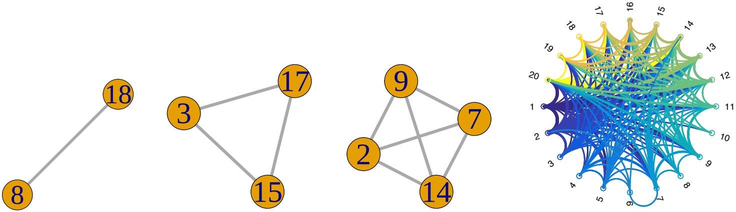
- Considered a variety of data generating processes for $(X_i, y_i), i = 1, \ldots, n$.
- X_i is generated via individual-specific weights λ_{ih} on common subnetworks + Gaussian noise

Simulation



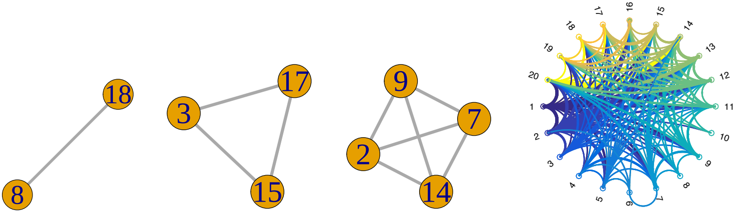
- Considered a variety of data generating processes for $(X_i, y_i), i = 1, \ldots, n$.
- X_i is generated via individual-specific weights λ_{ih} on common subnetworks + Gaussian noise
- A subset of these subnetworks are related to the response y_i

Simulation



- Considered a variety of data generating processes for $(X_i, y_i), i = 1, \ldots, n$.
- X_i is generated via individual-specific weights λ_{ih} on common subnetworks + Gaussian noise
- A subset of these subnetworks are related to the response y_i
- Considered two different signal-to-noise ratios

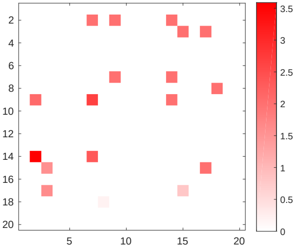
Simulation



- Considered a variety of data generating processes for $(X_i, y_i), i = 1, \ldots, n$.
- X_i is generated via individual-specific weights λ_{ih} on common subnetworks + Gaussian noise
- A subset of these subnetworks are related to the response y_i
- Considered two different signal-to-noise ratios
- Compared performance in different cases w/ Lasso & tensor PCA

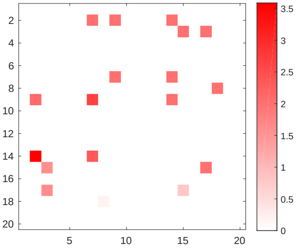
Low Noise

Coefficients of lasso

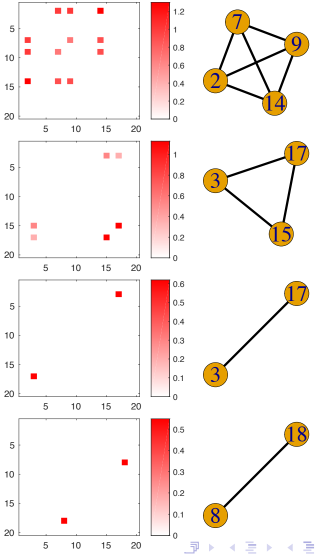


Low Noise

Coefficients of lasso



Coefficients and selected subgraphs of SBL



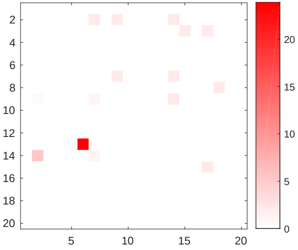
Low Noise

Repeat the procedure above 100 times.

	MSE	TPR	FPR
lasso	10.98±4.40	0.837±0.138	**0.002±0.005**
TN-PCA	**10.04±4.66**	0.449±0.499	0.449±0.499
SBL	10.08±4.51	**0.848±0.169**	0.005±0.007

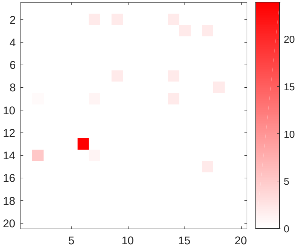
High Noise

Coefficients of lasso

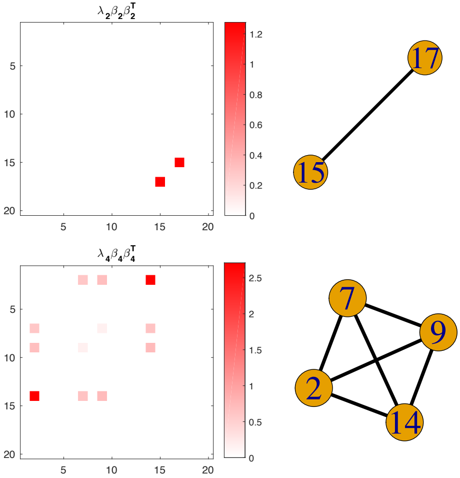


High Noise

Coefficients of lasso



Coefficients and selected subgraphs of SBL



High Noise

Repeat the procedure above 100 times.

	MSE	TPR	FPR
lasso	448.3±195.3	0.445±0.141	**0.025±0.037**
TN-PCA	624.0±287.8	0.060±0.239	0.060±0.238
SBL	**393.7±159.2**	**0.539±0.210**	0.029±0.038

HCP - Picture Vocabulary Data

- Age-adjusted picture vocabulary (PV) scores of 1065 subjects

HCP - Picture Vocabulary Data

- Age-adjusted picture vocabulary (PV) scores of 1065 subjects
  - presented with an audio recording of a word and 4 images

HCP - Picture Vocabulary Data

- Age-adjusted picture vocabulary (PV) scores of 1065 subjects
  - presented with an audio recording of a word and 4 images
  - select the picture that most closely matches the word

HCP - Picture Vocabulary Data

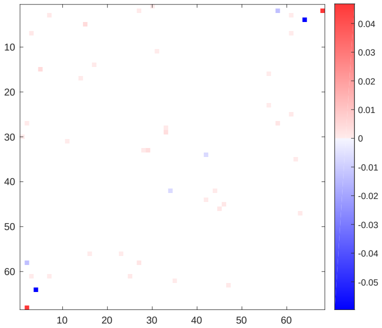
- Age-adjusted picture vocabulary (PV) scores of 1065 subjects
  - presented with an audio recording of a word and 4 images
  - select the picture that most closely matches the word
- Weighted brain network of fiber counts among 68 regions constructed for each subject (Zhang et al., 2018).

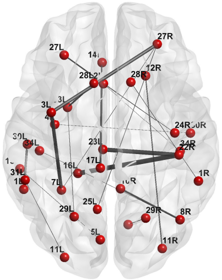
HCP - Picture Vocabulary Data

- Age-adjusted picture vocabulary (PV) scores of 1065 subjects
  - presented with an audio recording of a word and 4 images
  - select the picture that most closely matches the word
- Weighted brain network of fiber counts among 68 regions constructed for each subject (Zhang et al., 2018).
- Training set of 565 subjects & test set of 500 subjects.

HCP - Picture Vocabulary Data

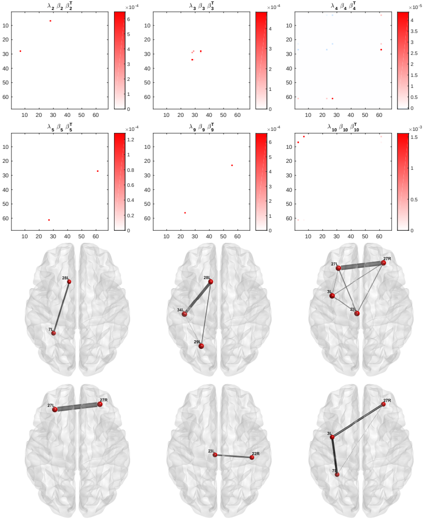
- Age-adjusted picture vocabulary (PV) scores of 1065 subjects
  - presented with an audio recording of a word and 4 images
  - select the picture that most closely matches the word
- Weighted brain network of fiber counts among 68 regions constructed for each subject (Zhang et al., 2018).
- Training set of 565 subjects & test set of 500 subjects.
- Estimated coefficients from lasso





Results from SBL

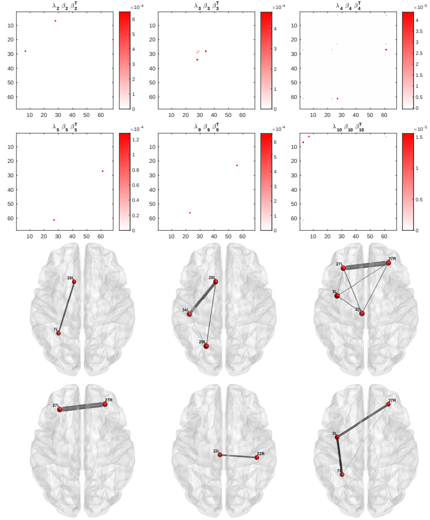
6 nonempty coefficient components out of $\{\lambda_h \boldsymbol{\beta}_h \boldsymbol{\beta}_h^\top\}_{h=1}^{10}$



HCP - Picture Vocabulary Data

Results from SBL

6 nonempty coefficient components out of $\{\lambda_h \boldsymbol{\beta}_h \boldsymbol{\beta}_h^\top\}_{h=1}^{10}$



27L, 27R (left and right superior frontal gyrus), 7L (left inferior parietal gyrus) and 29L (left superior temporal gyrus) are among activated regions when shifting from listening to meaningless pseudo sentences to listening to meaningful sentences (Saur et al., 2008; Dronkers, 2011).

Multiresolution tensor (MrTensor) networks

- Soccer passing networks data

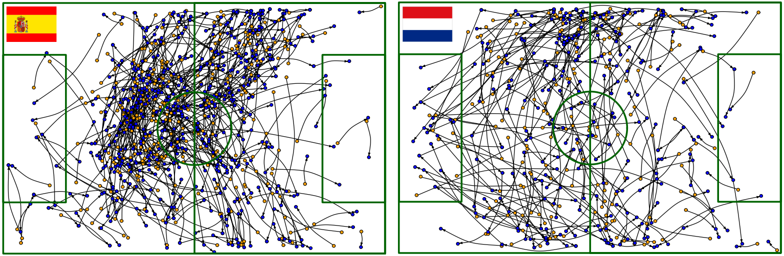


Figure: Spatial passing networks in a 2014 FIFA world cup match (Spain 1-5 Netherlands). Orange & blue nodes indicates origin-destination of pass. Team attack from left → right.

Multiresolution tensor (MrTensor) networks

- Soccer passing networks data

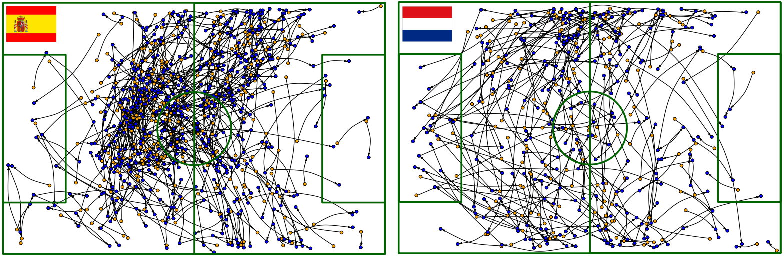


Figure: Spatial passing networks in a 2014 FIFA world cup match (Spain 1-5 Netherlands). Orange & blue nodes indicates origin-destination of pass. Team attack from left → right.

Multiresolution tensor (MrTensor) networks

- Soccer passing networks data

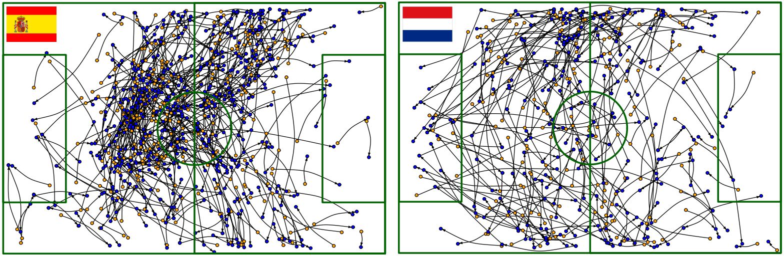


Figure: Spatial passing networks in a 2014 FIFA world cup match (Spain 1-5 Netherlands). Orange & blue nodes indicates origin-destination of pass. Team attack from left → right.

- Spatial replicated networks

Multiresolution tensor (MrTensor) networks

- Soccer passing networks data

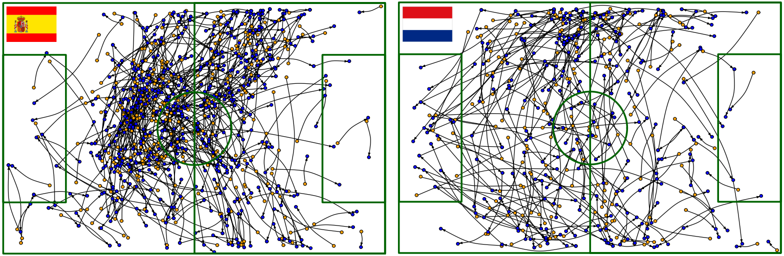


Figure: Spatial passing networks in a 2014 FIFA world cup match (Spain 1-5 Netherlands). Orange & blue nodes indicates origin-destination of pass. Team attack from left → right.

- Spatial replicated networks
  - Important to take spatial location into account

Multiresolution tensor (MrTensor) networks

- Soccer passing networks data

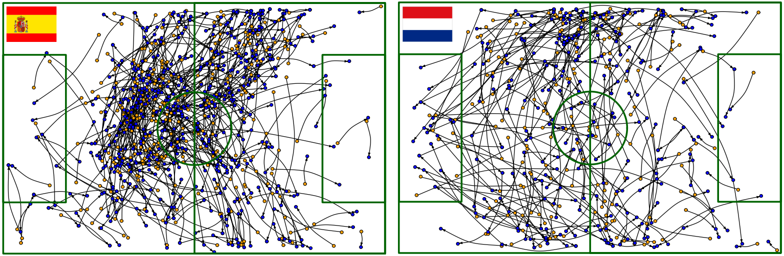


Figure: Spatial passing networks in a 2014 FIFA world cup match (Spain 1-5 Netherlands). Orange & blue nodes indicates origin-destination of pass. Team attack from left → right.

- Spatial replicated networks
  - Important to take spatial location into account
  - For brain nets, we used a pre-specified set of ROIs

Multiresolution tensor (MrTensor) networks

- Soccer passing networks data

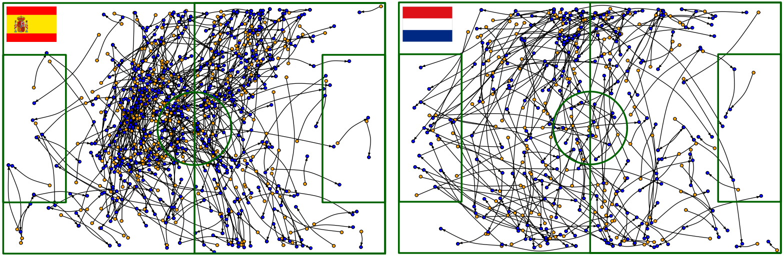
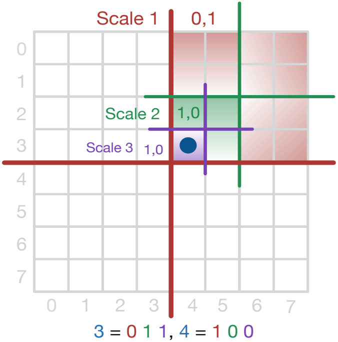


Figure: Spatial passing networks in a 2014 FIFA world cup match (Spain 1-5 Netherlands). Orange & blue nodes indicates origin-destination of pass. Team attack from left $\to$ right.

- Spatial replicated networks
  - Important to take spatial location into account
  - For brain nets, we used a pre-specified set of ROIs
  - Motivated by soccer passing, we develop multiresolution approaches

Fine-grained discretization



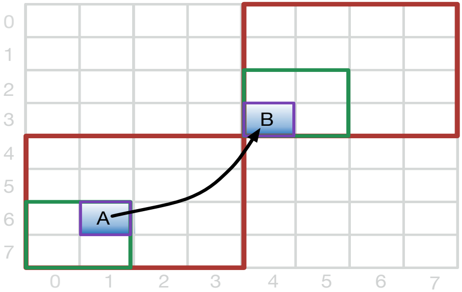


Figure: Coarse-to-fine dyadic partitioning

- *Binary coding* of each pass - according to sequence of partition set memberships of kicker & receiver

Fine-grained discretization

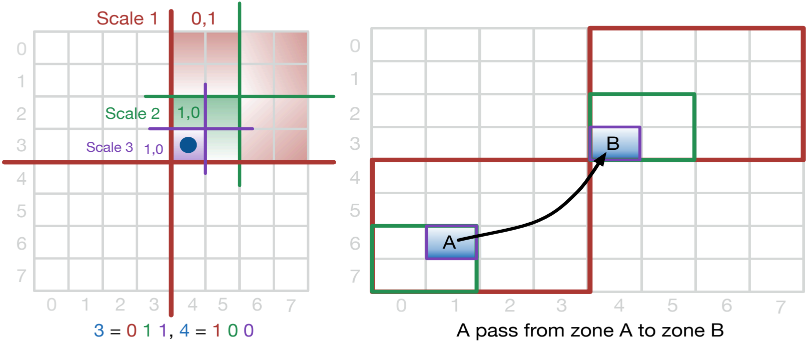


Figure: Coarse-to-fine dyadic partitioning

- *Binary coding* of each pass - according to sequence of partition set memberships of kicker & receiver
- Arrange the data as a *multiresolution adjacency tensor* $\mathcal{X}$

Fine-grained discretization

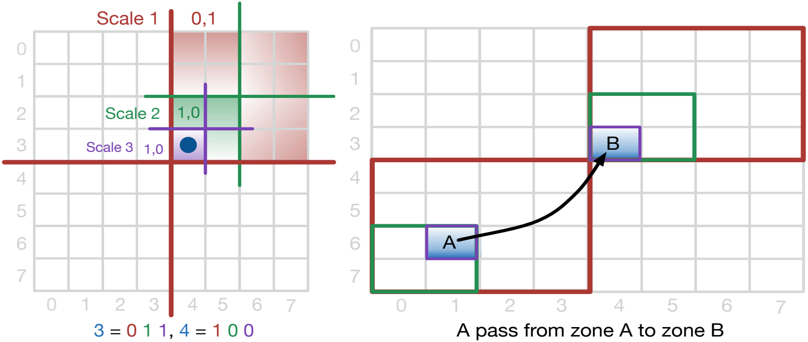


Figure: Coarse-to-fine dyadic partitioning

- *Binary coding* of each pass - according to sequence of partition set memberships of kicker & receiver
- Arrange the data as a *multiresolution adjacency tensor* $\mathcal{X}$
- Tensor is very large & sparse - we factorize using simpler pieces

Poisson block term decomposition

To represent the intensity of each weighted passing network as a superposition of H archetypal network motifs $\{\mathfrak{D}_h\}_{h=1:H}$, we propose the following model,

$$\mathfrak{X}_n \sim \mathrm{Pois}(\boldsymbol{\Lambda}_n), \quad \boldsymbol{\Lambda}_n = \sum_{h=1}^{H} \mathfrak{D}_h \upsilon_{h,n},$$

$$\mathfrak{D}_h = [\![\boldsymbol{\omega}_h; \boldsymbol{\Phi}_h^{(1)}, \boldsymbol{\Phi}_h^{(2)}, \boldsymbol{\Phi}_h^{(3)}, \boldsymbol{\Phi}_h^{(4)}, \boldsymbol{\Phi}_h^{(5)}, \boldsymbol{\Phi}_h^{(6)}]\!], \quad n = 1, \ldots, N.$$

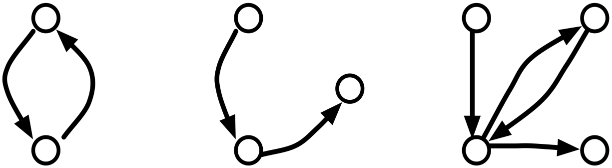


Figure: Three example *low-rank* passing network motifs involving 2–4 nodes

Block coordinate descent algorithm

Algorithm 1 Block nonlinear Gauss-Seidel algorithm for Poisson CP-BTD

Input: Multiresolution adjacency tensor $\mathfrak{X}$, the number of terms H, the CP rank R_h,
Initialize $\mathfrak{D}_h$
repeat
  % Given motifs $\{\mathfrak{D}_h : h = 1, \ldots, H\}$, update factor usage $\boldsymbol{\Upsilon}$;
  for $n = 1$ **to** N **do**
    Calculate $\boldsymbol{D}^{[n]}$ according to equation (4.5);
    $\boldsymbol{\upsilon_n} = \arg\min_{\boldsymbol{\upsilon_n} \geq 0} f_n(\boldsymbol{\upsilon_n}) \equiv \sum_{h=1}^{H} \upsilon_{h,n} - \sum_{j=1}^{J_n} x_{j,n} \log\left(\sum_{h=1}^{H} d_{j,h}^{[n]} \upsilon_{h,n}\right)$;
  end for
  Set $\boldsymbol{S} = \boldsymbol{\Omega\Upsilon}$, $\boldsymbol{\tau} = \boldsymbol{Se}$, $\boldsymbol{T} = \mathrm{diag}(\boldsymbol{\tau})$, $\boldsymbol{\Psi} = \boldsymbol{T}^{-1}\boldsymbol{S}^T$;
  for $p = 1$ **to** P **do**
    % Given $\boldsymbol{\Upsilon}$ and $\boldsymbol{A}^{(q)}$, $q = 1, \ldots, P$, $q \neq p$, update $\boldsymbol{\Phi}^{(p)}$;
    for $m = 1$ **to** I **do**
      Calculate $\boldsymbol{B}_m^{(p)}$ according to equation (4.8);
      $\boldsymbol{a}_m^{(p)} = \arg\min_{\boldsymbol{a}_m^{(p)} \geq 0} f_m(\boldsymbol{a}_m^{(p)}) \equiv \sum_{r=1}^{R} a_{r,m}^{(p)} - \sum_{j=1}^{J_m^{(p)}} x_{m,j}^{(p)} \log\left(\sum_{r=1}^{R} b_{j,r}^{(p)} a_{r,m}^{(p)}\right)$;
    end for
    Set $\boldsymbol{\rho} = \boldsymbol{A}^{(p)}\boldsymbol{e}$, update $\boldsymbol{\Phi}^{(p)} = \boldsymbol{A}^{(p)}[\mathrm{diag}(\boldsymbol{\rho})]^{-1}$;
    Update $\omega_{r_h,h} = \rho_{r_h,h} / \sum_{r_h=1}^{R_h} \rho_{r_h,h}$, $\forall\ (r_h, h)$;
  end for
until Convergence criterion is satisfied on all subproblems
Output: $\boldsymbol{\Omega}$, $\{\boldsymbol{\Phi}^{(p)}\}_{p=1:P}$, $\boldsymbol{\Upsilon}$

- The algorithm iterates between updating the tensor loading factor matrices and the factor usage; both steps boil down to a number of convex optimization subproblems

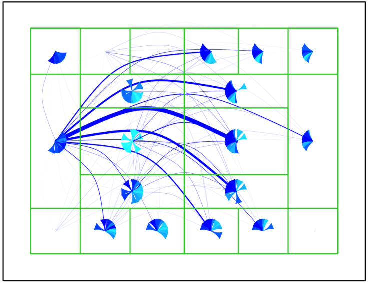
Block coordinate descent algorithm

Algorithm 1 Block nonlinear Gauss-Seidel algorithm for Poisson CP-BTD

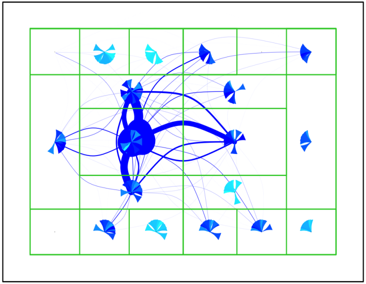
Input: Multiresolution adjacency tensor $\mathfrak{X}$, the number of terms H, the CP rank R_h,
Initialize $\mathfrak{D}_h$
repeat
  % Given motifs $\{\mathfrak{D}_h : h = 1, \ldots, H\}$, update factor usage $\boldsymbol{\Upsilon}$;
  for $n = 1$ **to** N **do**
    Calculate $\boldsymbol{D}^{[n]}$ according to equation (4.5);
    $\boldsymbol{\upsilon_n} = \arg\min_{\boldsymbol{\upsilon_n} \geq 0} f_n(\boldsymbol{\upsilon_n}) \equiv \sum_{h=1}^{H} \upsilon_{h,n} - \sum_{j=1}^{J_n} x_{j,n} \log\left(\sum_{h=1}^{H} d_{j,h}^{[n]} \upsilon_{h,n}\right)$;
  end for
  Set $\boldsymbol{S} = \boldsymbol{\Omega\Upsilon}$, $\boldsymbol{\tau} = \boldsymbol{Se}$, $\boldsymbol{T} = \mathrm{diag}(\boldsymbol{\tau})$, $\boldsymbol{\Psi} = \boldsymbol{T}^{-1}\boldsymbol{S}^T$;
  for $p = 1$ **to** P **do**
    % Given $\boldsymbol{\Upsilon}$ and $\boldsymbol{A}^{(q)}$, $q = 1, \ldots, P$, $q \neq p$, update $\boldsymbol{\Phi}^{(p)}$;
    for $m = 1$ **to** I **do**
      Calculate $\boldsymbol{B}_m^{(p)}$ according to equation (4.8);
      $\boldsymbol{a}_m^{(p)} = \arg\min_{\boldsymbol{a}_m^{(p)} \geq 0} f_m(\boldsymbol{a}_m^{(p)}) \equiv \sum_{r=1}^{R} a_{r,m}^{(p)} - \sum_{j=1}^{J_m^{(p)}} x_{m,j}^{(p)} \log\left(\sum_{r=1}^{R} b_{j,r}^{(p)} a_{r,m}^{(p)}\right)$;
    end for
    Set $\boldsymbol{\rho} = \boldsymbol{A}^{(p)}\boldsymbol{e}$, update $\boldsymbol{\Phi}^{(p)} = \boldsymbol{A}^{(p)}[\mathrm{diag}(\boldsymbol{\rho})]^{-1}$;
    Update $\omega_{r_h,h} = \rho_{r_h,h} / \sum_{r_h=1}^{R_h} \rho_{r_h,h}$, $\forall\ (r_h, h)$;
  end for
until Convergence criterion is satisfied on all subproblems
Output: $\boldsymbol{\Omega}$, $\{\boldsymbol{\Phi}^{(p)}\}_{p=1:P}$, $\boldsymbol{\Upsilon}$

- The algorithm iterates between updating the tensor loading factor matrices and the factor usage; both steps boil down to a number of convex optimization subproblems
- The algorithm is convergent with lower per-iteration cost and much greater memory efficiency.

Interpretable passing motifs: tactical styles & top 10 teams



Counter-Attack



Tiki-taka possession

Top 10 Counter-attack team-game: Algeria-54, Netherlands-3, Iran-12, Costa Rica-52, Colombia-37, Cameroon-33, Ecuador-26, Ecuador-42, Greece-22, Algeria-48

Top 10 Possession team-game: Spain-3, Bosnia-44, Italy-8, France-10, Italy-24, Spain-19, Switzerland-25, Brazil-63, Argentina-62, Bosnia-28

Supervised embedding of networks

- Interested in understanding how the usage of specific passing network motifs contribute to the outcomes, we take a supervised approach on the factor score

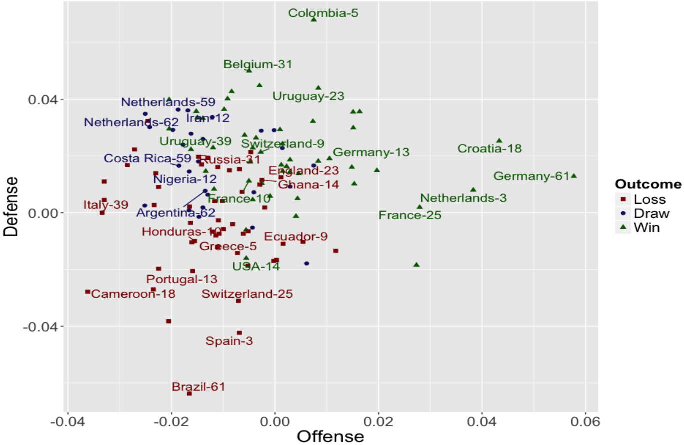
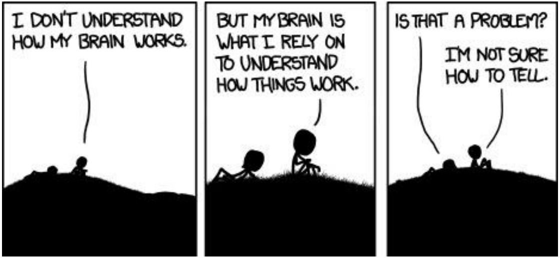


Figure: Embedding high-dimensional passing networks into a two-dimensional space.

Discussion

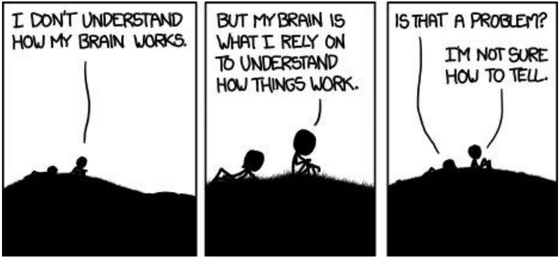
- Focus on interpretable predictive methods from replicated structured networks



Thank You

Discussion

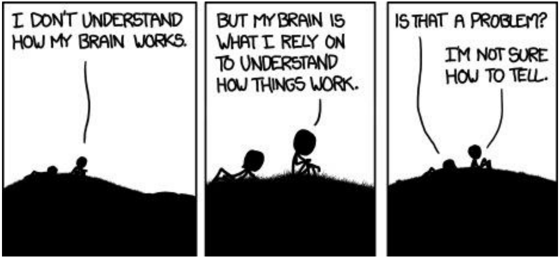
- Focus on interpretable predictive methods from replicated structured networks
- Little consideration of relevant methods in the literature



Thank You

Discussion

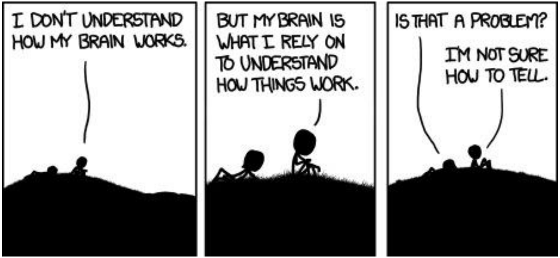
- Focus on interpretable predictive methods from replicated structured networks
- Little consideration of relevant methods in the literature
- We have been focusing on simple & fast algorithms motivated by concrete apps



Thank You

Discussion

- Focus on interpretable predictive methods from replicated structured networks
- Little consideration of relevant methods in the literature
- We have been focusing on simple & fast algorithms motivated by concrete apps
- Many, many more interesting directions - UQ, scalable Bayes, more theory, etc etc



Thank You

References

- Durante D, Dunson DB, and Vogelstein JT. "Nonparametric Bayes modeling of populations of networks". In: *Journal of the American Statistical Association* (2017), pp. 1–15.
- Wang L, Zhang Z, and Dunson D. "Symmetric Bilinear Regression for Signal Subgraph Estimation". In: *arXiv preprint arXiv:1804.09567* (2018).
- Wang R, Zhang Z, and Dunson D. "Common and individual structure of multiple networks". In: *arXiv preprint arXiv:1707.06360* (2017).
- Han S and Dunson D. "Multiresolution Tensor Decomposition for Multiple Spatial Passing Networks". In: *arXiv preprint arXiv:1803.01203* (2018).
- Zhang Z et al. "Mapping population-based structural connectomes". In: *NeuroImage* 172 (2018), pp. 130–145.
- Zhang Z et al. "Relationships between Human Brain Structural Connectomes and Traits". In: *bioRxiv* (2018),