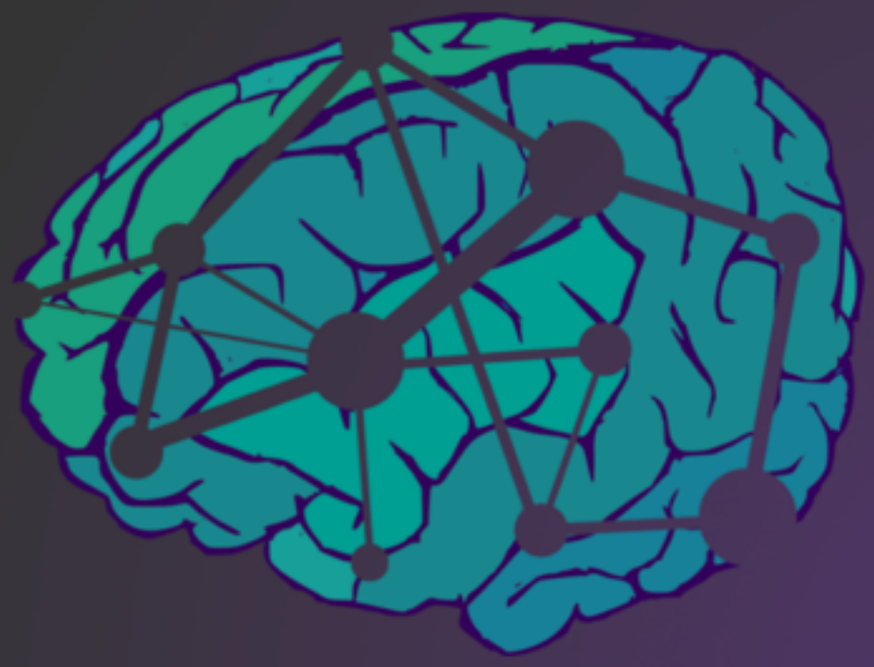
neurodata.io

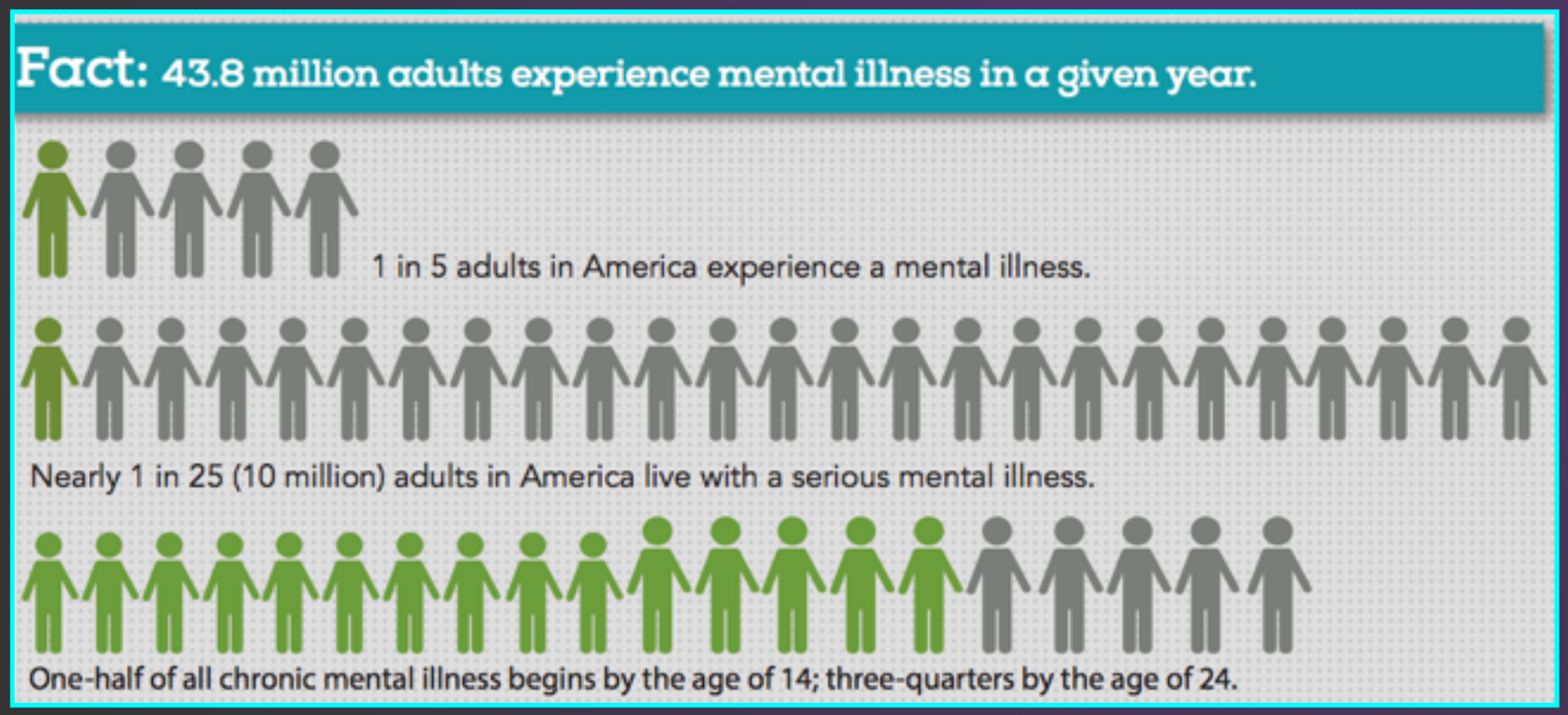


HOCKEY: How to Observe Connectomes for gaining Knowledge and Estimating ‘Y’

Greg Kiar

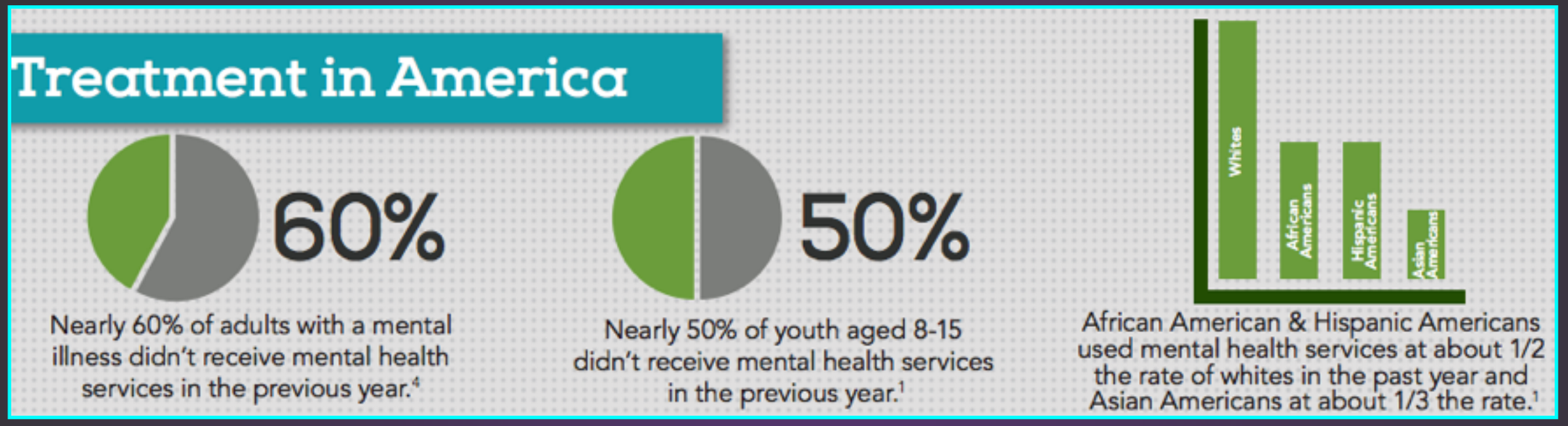
March 7th, 2016

Significance



NAMI: https://www.nami.org/NAMI/media/NAMI-Media/Infographics/GeneralMHFacts.pdf

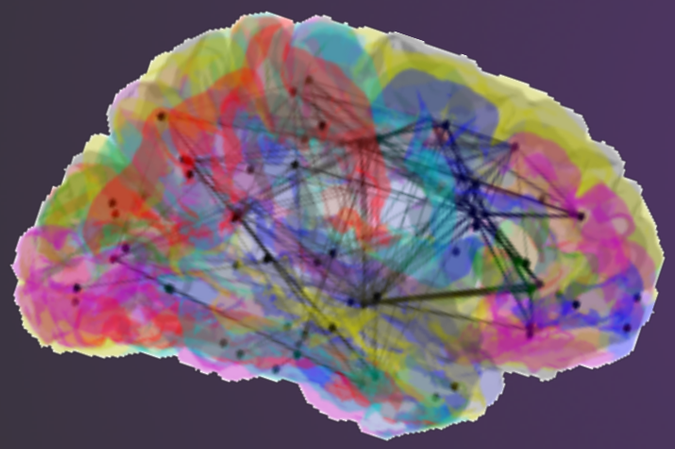
Gap



- "50% of people show signs at 14 years old diagnosed with a mental disorder show signs of the disease by age 14, 75% by age 25."
- "76-85% of serious cases went untreated in low and middle income countries, 35-50% of cases in high income countries."

Prevalence, Severity, and Unmet Need for Treatment of Mental Disorders, World Health Organization World Mental Health Surveys, June 2004, Journal of the American Medical Association

Challenge



???

Formal Statement of Problem

$$G_i, Y_i \sim \mathcal{F} = \{F_{G,Y}(\cdot;\theta) : \theta \in \Theta)$$

Graphs and labels observed

$$F_{G,Y} = F_{X,Y}$$

They are graph matched

$$X_i = \prod_{u,v}^{\epsilon} A_{uv}; \epsilon \subset V \times V$$

X is our adjacency matrix

$$Y_i = \{0, 1\}$$

Y is our binary label vector

$$\hat{p_i} = \frac{\sum_i^{\epsilon} X_i}{|\epsilon|}$$

We estimate edge probability

CLASSIFY

$$E(l) = \sum \mathbb{I}(\hat{Y}_i \neq Y_i)$$

We evaluate based on correct assignment

Model Assumptions

$$G_i, Y_i \overset{iid}{\sim} \mathcal{F}$$

Graphs are i.i.d.

$$F_{X|0} = ER(p_0) = Bern(p_0)^{V \times V}$$
$$F_{X|1} = ER(p_1) = Bern(p_1)^{V \times V}$$

Edges are i.i.d.

$$p_0 \neq p_1$$

A class conditional edge probability exists

Formal Statement of Algorithm

Algorithms used:

- Linear Discriminant Analysis (LDA)
- Quadratic Discriminant Analysis (QDA)
- K-Nearest Neighbours (KNN)
- Support Vector Machine (SVM)
- Random Forrest (RF)

QDA:

Quadratic Score Function:

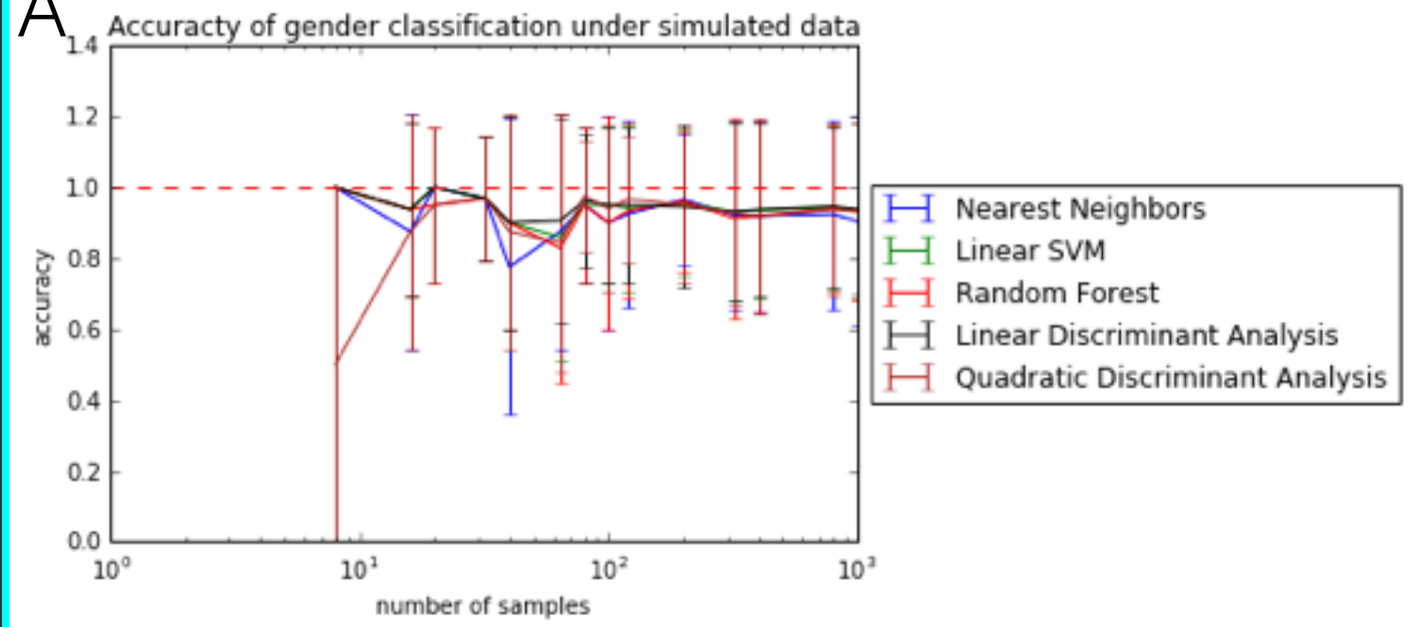
$$s_i^Q(\mathbf{x}) = -\frac{1}{2}\ln|\Sigma_i| - \frac{1}{2}(\mathbf{x}-\mu_i)^T\Sigma_i^{-1}(\mathbf{x}-\mu_1) + \ln p_i$$

Becomes Decision Rule:

$$s_i^Q(\mathbf{x}) = -\frac{1}{2}\ln|S_i| - \frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^T S_i^{-1}(\mathbf{x}-\bar{\mathbf{x}}) + \ln p_i$$

Results

A



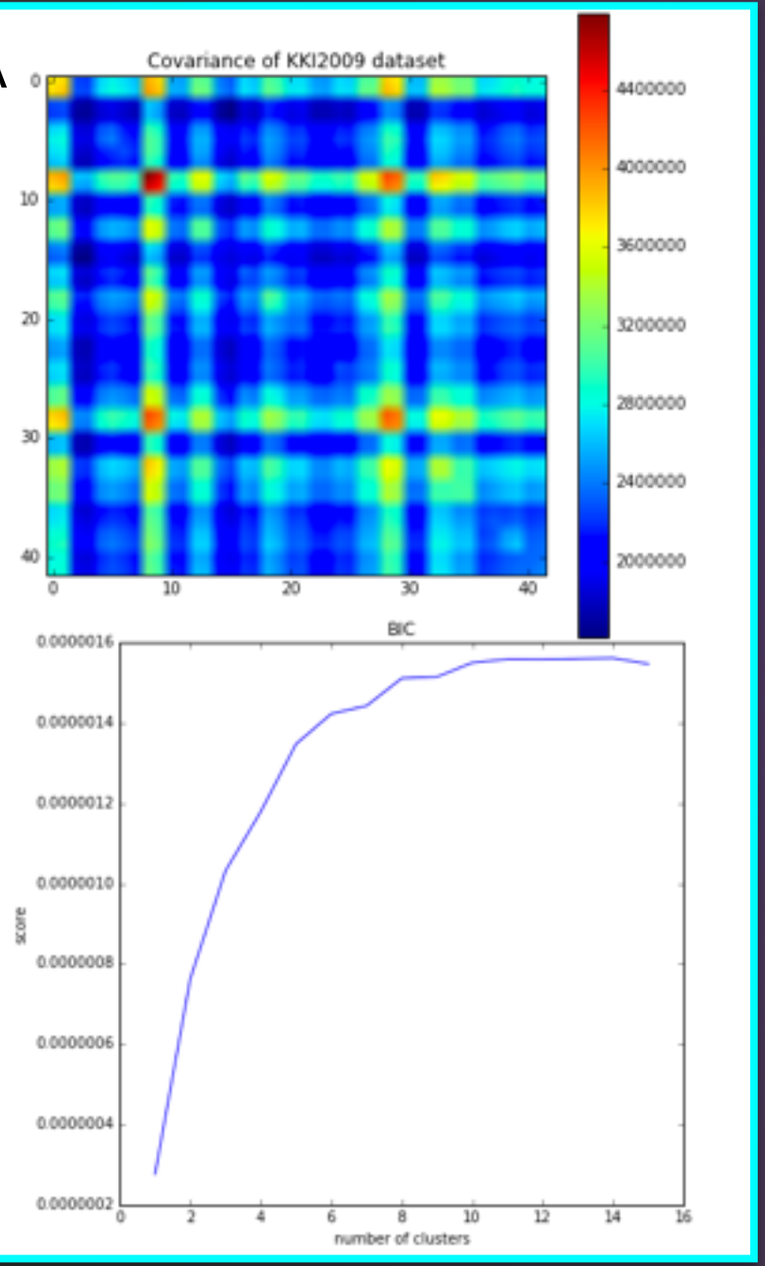
B

Algorithm	Classification Accuracy
Nearest Neighbors	0.48 (+/- 1.00)
Linear SVM	0.55 (+/- 1.00)
Random Forest	0.57 (+/- 0.99)
Linear Discriminant Analysis	0.45 (+/- 1.00)
Quadratic Discriminant Analysis	0.71 (+/- 0.90)

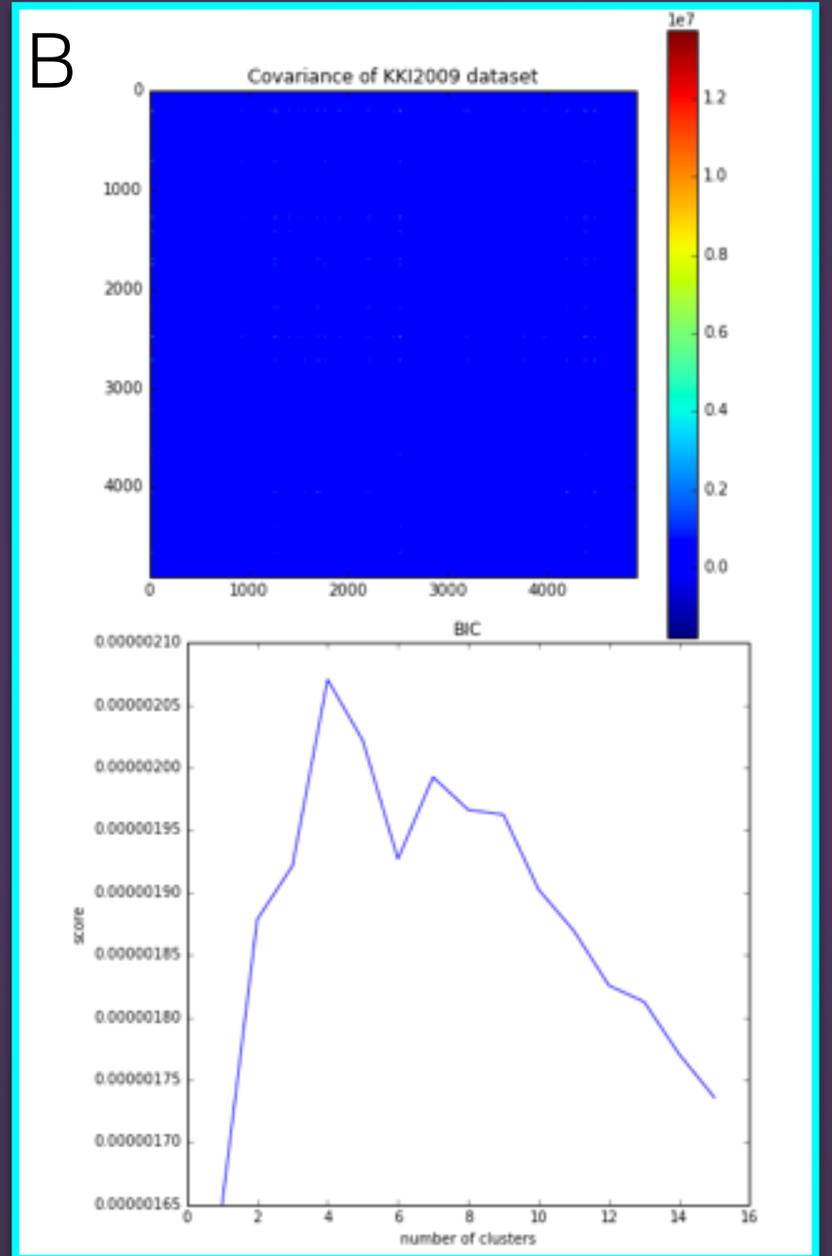
A. The performance of several classifiers on simulated data from our model, showing performance as the number of samples scaled. All algorithms scaled well with N and achieved near perfect accuracy in classification under these conditions. B. Performance of the same classifiers on the given dataset consisting of 42 graphs with 20 and 22 members of class 0 and 1, respectively, and each graph containing 70 nodes. The best classifier used here was the QDA method, which achieved just over 70% classification accuracy.

Model Checking

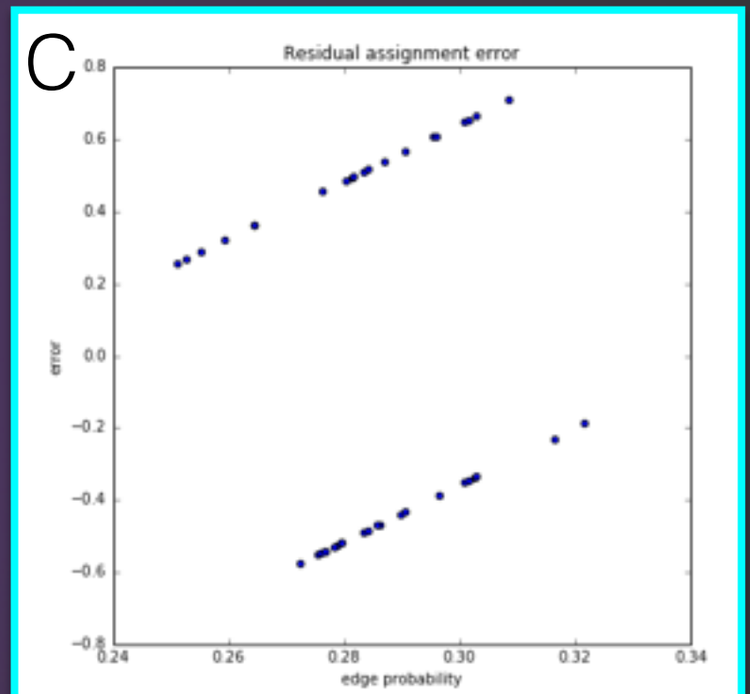
A



B

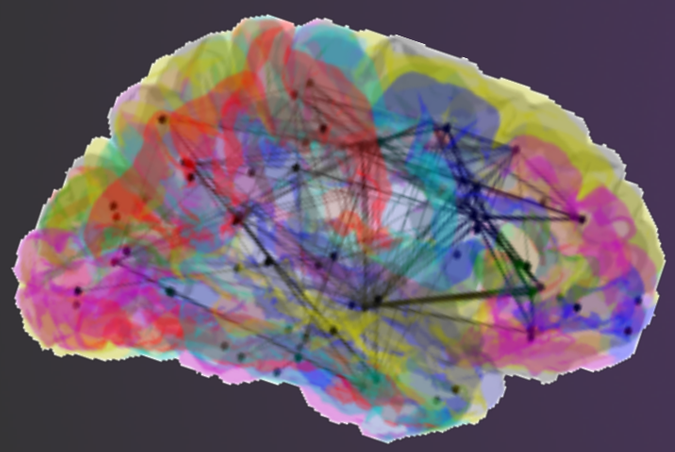


C



A. We show that the covariance across subjects is non zero and the ideal number of clusters is > 1, indicating both our graph i.i.d. assumptions were false. B. Like A, the same is true for the edge i.i.d. assumption. C. Performing linear regression over edge probability and class labels failed to separate the subjects successfully 100% of the time, as can be seen in the residual plot; this shows us the class conditional probability difference assumption is also false.

Resolution



!!!

Rx