

# 1 Introduction

## 2 Methods

## 3 Simulation Study

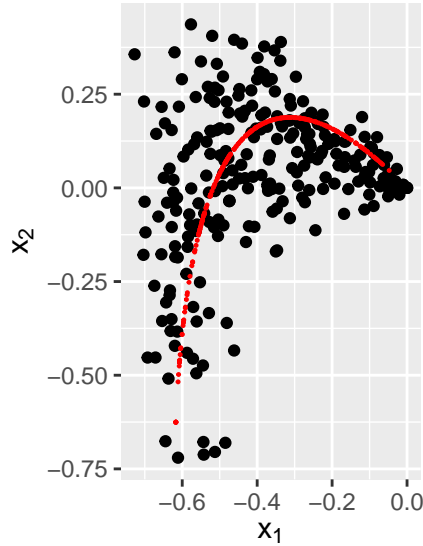
### 3.1 Classification

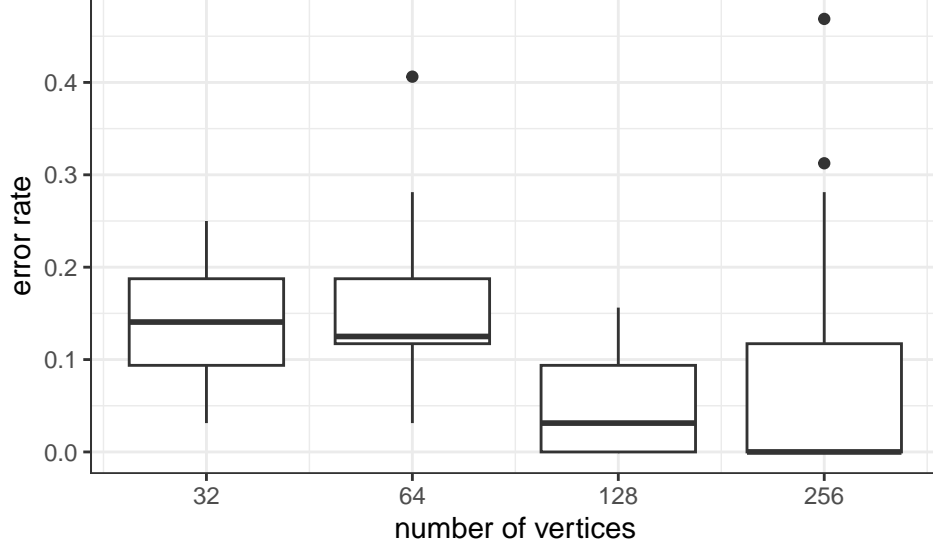
In this simulation experiment, the latent vectors were sampled along a Bezier curve defined by  $g(t) = \begin{bmatrix} t^2 & 2t(1-t) \end{bmatrix}^\top$ . The timepoints  $t_i$  were sampled as iid Beta random variables with two sets of parameters,  $(\alpha_1, \beta_1)$  and  $(\alpha_2, \beta_2)$ . The setup is as follows:

1. Draw response variables  $y_1, \dots, y_n \stackrel{\text{iid}}{\sim} \text{Multinomial}(1/2, 1/2)$ .
2. For each  $i = 1, \dots, n$ , draw  $t_i \mid y_i \stackrel{\text{ind}}{\sim} \text{Beta}(\alpha_{y_i}, \beta_{y_i})$ , such that
  - $\alpha_1 = 1$
  - $\beta_1 = 2$
  - $\alpha_2 = 2$
  - $\beta_2 = 1$
3. Construct each latent vector as  $x_i = g(t_i)$  and compile them in data matrix  $X = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^\top$ .
4. Sample graph and its adjacency matrix as  $A \sim \text{RDPG}(X)$ .

For each graph, we constructed the ASE, which was used to estimate the parameters  $(\hat{\alpha}, \hat{\beta})$  for the graph, using the maximum likelihood method. The estimated parameters were then used as predictors  $y_1, \dots, y_n$ , setting aside half for training and half for testing. We investigated graphs of size  $|V| = 32, 64, 128, 256$ . The number of graphs for each experiment was set to  $n = 64$ . For each (number of graphs, size of graph) pair, we performed 32 replicates. Figure ?? shows the ASE of one graph.

Figure ?? shows the boxplots of the classification error rates.

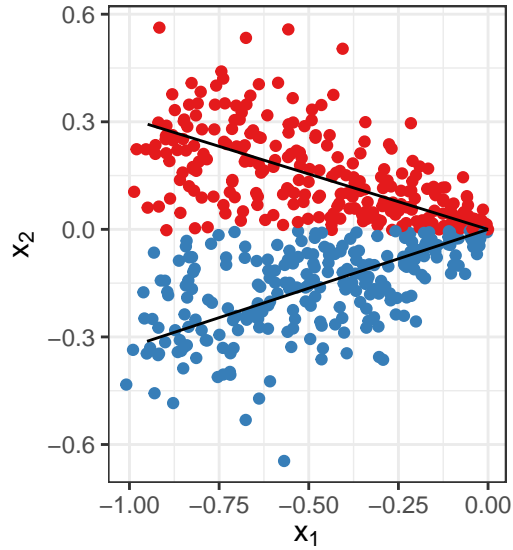


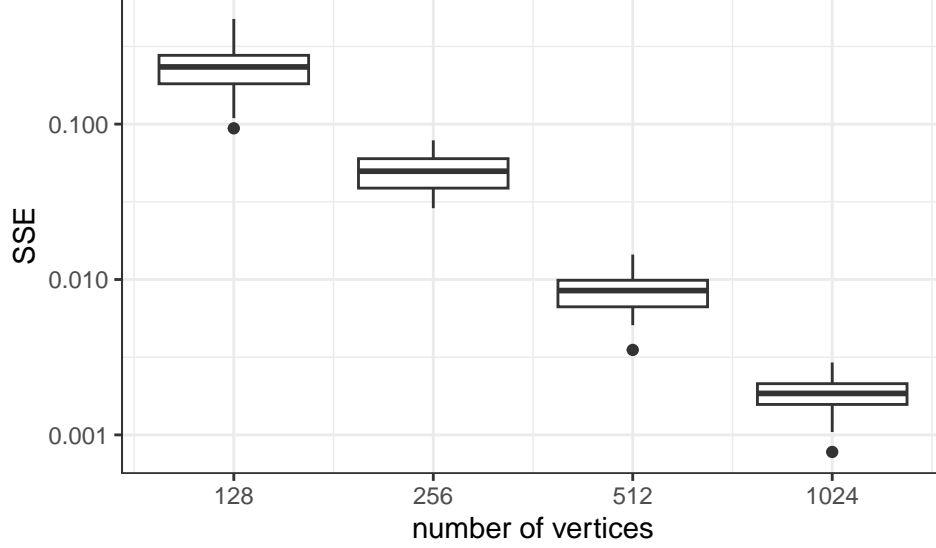


### 3.2 Regression

1. Draw angles  $\theta_1, \dots, \theta_N \stackrel{\text{iid}}{\sim} \text{Uniform}(\pi/6, \pi/3)$ .
2. For each  $k = 1, \dots, N$ ,
  - i. Draw  $t_1, \dots, t_n \stackrel{\text{iid}}{\sim} \text{Uniform}(0, 1)$ ;
  - ii. Draw  $z_1, \dots, z_n \stackrel{\text{iid}}{\sim} \text{Multinomial}(1/2, 1/2)$ ;
  - iii. For each  $i = 1, \dots, n$ , set  $x_i = \begin{cases} \begin{bmatrix} t_i & 0 \end{bmatrix}^\top & z_i = 1; \\ \begin{bmatrix} t_i \cos \theta_k & t_i \sin \theta_k \end{bmatrix}^\top & z_i = 2 \end{cases}$ ;
  - iv. Collect  $X = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^\top$  and draw  $A \sim \text{RDPG}(X)$ ;
  - v. Set the response  $y_k = \beta_0 + \beta_1 \theta_k$ .

In this simulation, we set  $\beta_0 = \beta_1 = 1$ . The number of graphs was set to  $N = 64$ , and the number of vertices per graph was set to  $n = 128, 256, 512, 1024$ . For each  $n$ , we simulated 32 replicates.





## 4 Applications

In the first example, we analyzed fiber count data between brain regions from the Human Connectome Project (HCP). When analyzing these data as graphs, we denote the regions as vertices and the fiber counts between pairs of regions as weighted edges. A plausible statistical model for these data is to assume that the edge weights between pairs of vertices is Poisson distributed, i.e., the adjacency matrix is sampled as  $A_{ij} \stackrel{\text{ind}}{\sim} \text{Poisson}(\Theta_{ij})$ , where  $\Theta \in \mathbb{R}_+^{n \times n}$  is a symmetric matrix of Poisson parameters.

In this dataset, there are  $N = 136$  graphs (corresponding to individual subjects), each with  $n = 68$  vertices (corresponding to brain regions). Analyzing these graphs as RDPGs reveals that the DCBM is a good candidate for these data (fig 1).

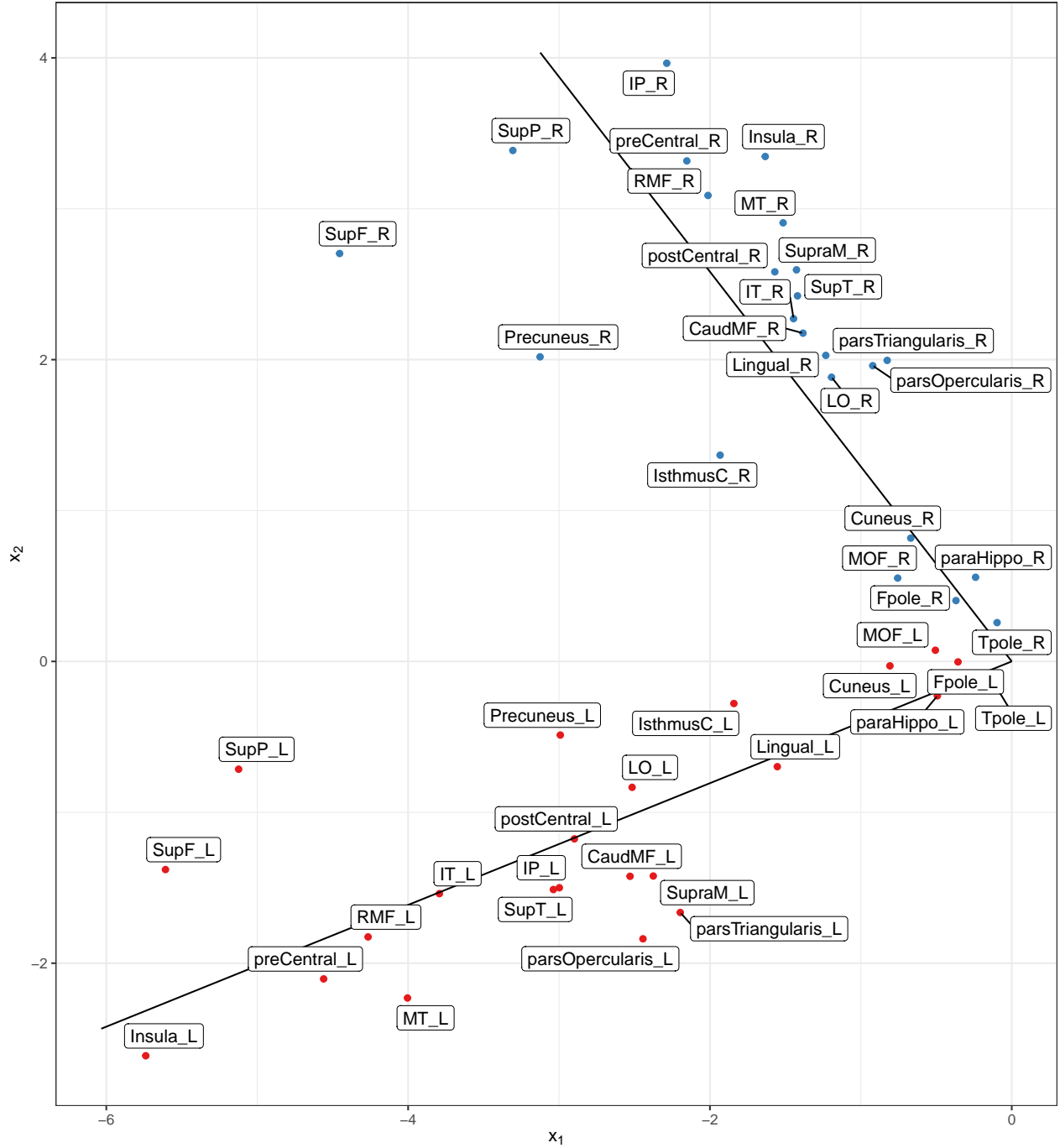
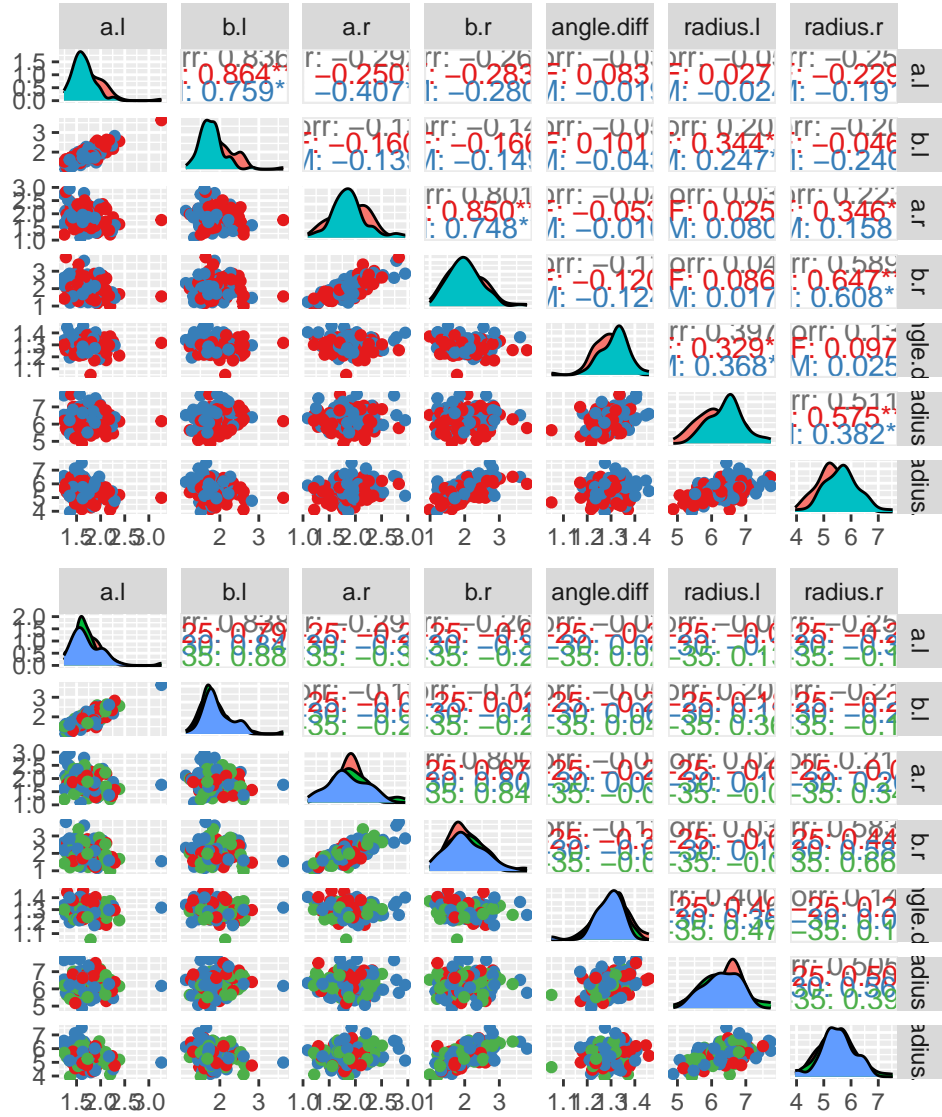


Figure 1: ASE of one graph from the HCP dataset. The lines are fitted via  $K$ -curves clustering using degree = 1. The outputted clusters correspond exactly to the left (red) and right (blue) hemispheres.

For each graph, the



## 5 Discussion