

# 0425文献汇报：GRAPH-AWARE MODELING OF BRAIN CONNECTIVITY NETWORKS

脑网络是从 fMRI 信号构建的加权图，节点是脑区ROI，边表示两脑区之间的功能连接强度（Fisher转换相关系数）。当前研究存在两大问题：

- 1. **边级推断可解释性弱**：个体边级别推断缺乏可解释性，还应关注功能系统之间的连接变化；在神经影像学中称之为 图感知推断，但对功能区域进行聚合会导致大量信息丢失和准确性降低。
- 2. **边之间高度相关性被忽视**：若假设边独立，标准误将被严重低估，忽略受试者内部边权重之间的相关性，使得基于独立假设的推断不可靠。

针对这两类问题，文章提出了一种线性混合效应模型，该模型既考虑了功能系统，又考虑了边缘依赖性，同时仍然建模个体边缘权重以避免信息丢失。

模型允许比较两个群体，患者和健康对照组，在**功能区域层面**和**个体边缘层面**上进行对比，从而得出生物学上有意义的解释，并将此模型应用于精神分裂症患者和健康对照组的静息态fMRI数据，获得了与精神分裂症文献一致且可解释的结果。

## 图感知的LME

- **系统结构共享**（cell-level pooling）：系统之间存在共同规律；
- **边级特异性建模**（edge-level unpooling）：边对协变量响应各不相同；
- **个体差异与边依赖性建模**：引入随机效应与结构化协方差。

### 模型设定：

- 每位被试  $m$  的网络是对称加权图  $A^m$ ；
- 所有节点按功能系统划分为  $K$  个 community；  
Let  $c_i$  be the community label of node  $i$ , common across all networks and taking values in  $\{1, \dots, K\}$ . We refer to an unordered pair of communities  $(a, b)$ , where  $a, b \in \{1, \dots, K\}$  as a network cell; there are a total of  $K(K + 1)/2$  cells corresponding to  $K$  communities. We
- 每对系统  $(a, b)$  构成一个cell，内部所有边进入该 cell 的建模；  
Let  $n^{(a)} = |\{i : c_i = a\}|$  denote the number of nodes in community  $a$ , and let  $n^{(a,b)}$  be the number of edges in cell  $(a, b)$ , where
$$n^{(a,b)} = \begin{cases} n^{(a)}n^{(b)}, & \text{if } a \neq b, \\ n^{(a)}(n^{(a)} - 1)/2 & \text{if } a = b, \end{cases}$$
and let  $n^{(\cdot,\cdot)} = n(n - 1)/2$  be the total number of distinct edge weights.
- 每条边权作为响应变量建模： $y_{m,i}^{(a,b)}$  为个体  $m$  的 cell  $(a, b)$  中第  $i$  条边的权重。

### 建模：

- $$y_{m,i}^{(a,b)} = x_m^T \alpha^{(a,b)} + x_m^T \eta_i^{(a,b)} + \gamma_m^{(a,b)} + \epsilon_{m,i}^{(a,b)}$$
- $x_m$ ：个体协变量，是否患病；
  - $\alpha^{(a,b)}$ ：系统对的主效应（共享结构）——混合模型的 固定效应；
  - $\eta_i^{(a,b)}$ ：cell  $(a,b)$  内边  $i$  的特异响应（个体差异）——混合模型的cell级 随机效应，即使两个边都在 cell  $(a, b)$  内，也允许它们对疾病状态的响应强度不一样；
  - $\gamma_m^{(a,b)}$ ：该被试在该系统下的整体偏移（个体随机效应）——混合模型的个体级 随机效应， $\gamma_m^{(a,b)} \sim N(0, U)$
  - $\epsilon_{m,i}^{(a,b)}$ ：残差项，被试m的cell  $(a,b)$  内边  $i$  的随机扰动  $\epsilon_{m,i}^{(a,b)} \sim N(0, V)$
- $\eta_i^{(a,b)}$  的约束， $\sum_i \eta_{ij}^{(a,b)} = 0$

where  $m = 1, \dots, N$  is the subject index,  $i = 1, \dots, n^{(a,b)}$  is the edge index within the cell, and  $a, b \in \{1, \dots, K\}$  are community labels. For identifiability we require that  $\sum_i \eta_{i,j}^{(a,b)} = 0$  for all  $j$ .

考虑用被试 $m$ 的患病标签 $d_m$ 加入协变量协变量  $x$

含截距项:  $x_m = (1 \quad d_m)^T$ , 建模为

$$y_{m,i}^{(a,b)} = (1 \quad d_m)^T \alpha^{(a,b)} + (1 \quad d_m)^T \eta_i^{(a,b)} + \gamma_m^{(a,b)} + \epsilon_{m,i}^{(a,b)}$$

$$y_{m,i}^{(a,b)} = (\alpha_0^{(a,b)} + d_m \alpha_1^{(a,b)}) + (\eta_{i,0}^{(a,b)} + d_m \eta_{i,1}^{(a,b)}) + \gamma_m^{(a,b)} + \epsilon_{m,i}^{(a,b)}.$$

For every cell  $(a, b)$ , the term  $\alpha_0$  represents the cell-level mean for patients with no disease,  $\alpha_1$  the cell-level shift due to disease,  $\eta_{i,0}$  the edge-specific intercept for patients with no disease, and  $\eta_{i,1}$  the edge-specific disease effect. These are all fixed effects, whereas  $\gamma_m^{(a,b)}$  is the subject-specific random effect for the given network cell representing individual heterogeneity with mean 0 over the population of subjects.

矩阵形式

1. 将所有边拼接成向量  $y_m$ , 模型为:

$$E[y_m \mid \gamma_m] = X_m \beta + Z \gamma_m$$

$$\text{Var}(y_m \mid \gamma_m) = V$$

2. 随机效应设计矩阵  $Z$

- $Z$  用于将个体随机效应  $\gamma_m^{(a,b)}$  扩展到 cell  $(a, b)$  中的所有边。
- 每一块子矩阵是全 1 向量, 表示同一 cell 下的所有边都共享该系统层级的个体偏移量。

effects  $\gamma_m^{(a,b)}$  across edges and has the block-diagonal form

$$Z = \begin{pmatrix} 1_{n(1,1)} & 0_{n(1,1)} & 0_{n(1,1)} & \dots & 0_{n(1,1)} \\ 0_{n(2,1)} & 1_{n(2,1)} & 0_{n(2,1)} & \dots & 0_{n(2,1)} \\ 0_{n(2,2)} & 0_{n(2,2)} & 1_{n(2,2)} & \dots & 0_{n(2,2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0_{n(K,K)} & 0_{n(K,K)} & 0_{n(K,K)} & \dots & 1_{n(K,K)} \end{pmatrix},$$

where  $1_{n(a,b)}$  and  $0_{n(a,b)}$  are a vector of either all ones or all zeroes with length  $n^{(a,b)}$ . The fixed effects design matrix for cell  $(a, b)$  is given by

$$X_m^{(a,b)} = \begin{pmatrix} x_m^T & x_m^T & \dots & 0 \\ x_m^T & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ x_m^T & 0 & \dots & x_m^T \\ x_m^T & -x_m^T & \dots & -x_m^T \end{pmatrix},$$

3. 固定效应设计矩阵  $X$

- $X_m$  是第  $m$  个被试的固定效应设计矩阵。
- 它包含了所有 cell-level 的  $\alpha^{(a,b)}$  和 edge-level 的  $\eta_i^{(a,b)}$ 。
- 作用: 建模每条边在各系统和边特定的协变量效应 (如患病状态)。

where the initial columns hold the cell-level effects  $\alpha^{(a,b)}$  and subsequent columns capture the contribution of the  $\eta_i^{(a,b)}$ 's. The full design matrix for subject  $m$  is then given by “tiling” cell-level design matrices into a block-diagonal  $X_m = \text{diag}(X_m^{(a,b)})$ . Integrating out the random

4. 边向量  $y_m$  的边际分布, 用于推理

$$E[y_m] = X_m \beta, \quad \text{Var}(y_m) = V + Z U Z^T \equiv \Sigma$$

- $\beta$  包含所有 cell 的  $\alpha^{(a,b)}$  与  $\eta_i^{(a,b)}$

- $Z$  是 block-diagonal 矩阵，将每个  $\gamma_m^{(a,b)}$  广播到 cell 内所有边
- $V$  是边层级残差协方差矩阵
- $V$  可选为：
  - **Diagonal**
  - **Block-diagonal**: 允许同一 cell 内边相关（更灵活，但复杂）。

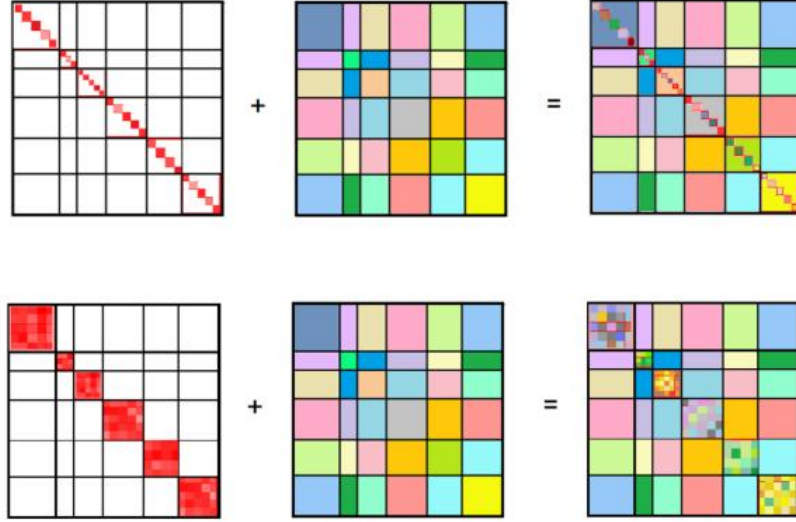


FIG. 1. Decomposition of the intrasubject covariance matrix of edge weights,  $V + ZU Z^T = \Sigma$  with diagonal  $V$  (top) and block-diagonal  $V$  (bottom).

- $U$  是 cell-level 随机效应  $\gamma_m$  的协方差矩阵，维度为  $K(K+1)/2 \times K(K+1)/2$
- $ZU Z^T$ : 来自 cell-level 的随机效应刻画边的依赖性（相关性）

#### 5. 估计

极大似然法联合估计  $V, U, \beta$

**2.5. Model fitting with the EM algorithm.** In practice, the GLS estimator (2.3) is not computable, since  $V$  and  $U$  are unknown. We take the approach of jointly estimating  $V, U$ , and  $\beta$  by maximum likelihood under the normal assumption on the random effects and the errors. Specifically, we assume that  $\gamma_m \sim N(0, U)$  and

$$y_m = X_m \beta + Z \gamma_m + \epsilon_m,$$

where  $\epsilon_m = \{\{\epsilon_{m,i}^{(a,b)}, 1 \leq i \leq n^{(a,b)}\} : 1 \leq a \leq b \leq K\} \sim N(0, V)$  is the concatenated error term independent of  $\gamma_m$ .