arXiv:2305.01596v1 [stat.ME] 2 May 2023

空间连续性约束下的体素对级脑连接分析网络方法

Tong Lu[1], Yuan Zhang[2],Peter Kochunov[3],Elliot Hong[3],Shuo Chen[3,4]

[1]马里兰大学数学系

[2]俄亥俄州立大学统计系

[3]马里兰大学医学院马里兰精神病学研究中心

[4]马里兰大学医学院生物统计与生物信息学部，*shuochen@som.umaryland

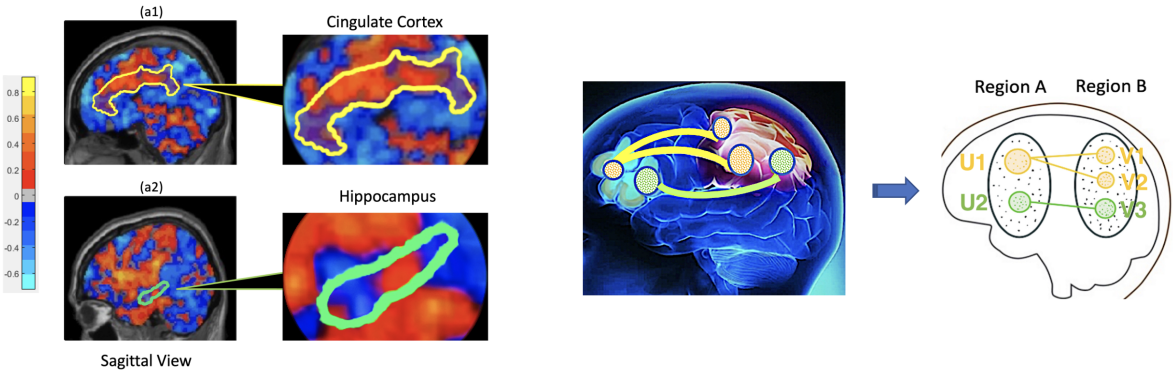
摘要

脑连接组分析通常会通过对感兴趣区域内的信号取平均，将高分辨率脑扫描（通常由数百万个体素组成）压缩为仅数百个感兴趣区（ROI）。这种大幅降维提高了计算速度，并改善了解剖结构的形态学特征；然而，这也会造成空间特异性和灵敏度的大量损失，尤其是在信号表现出高度感兴趣区域内异质性时。脑部疾病导致一对感兴趣区域之间功能连接异常表达，往往主要由该感兴趣区对内少量体素对驱动。本文提出一种在空间约束下检测体素对层级神经连接异常的新网络方法。具体而言，针对一对感兴趣区域，我们的模型旨在提取包含异常体素对连接的密集子区域，同时确保涉及的体素在空间上连续。此外，我们开发了用于实现该模型的子社群检测算法，并证明了这些算法的一致性。综合模拟研究表明，该方法能够有效降低假阳性率，同时提高统计功效、检测可重复性和空间特异性。我们将该方法用于揭示：（i）一项精神分裂症研究中，来自330名参与者的显著性网络和颞叶-丘脑网络内受精神分裂症影响的体素级功能连接模式；（ii）利用英国生物样本库数据，从3269名参与者中发现基底神经节、海马与岛回之间与尼古丁成瘾相关的受损体素级功能连接模式。检测结果与既往医学发现一致，但提供了更精细的局部信息。

关键词：关键词：脑连接组、空间连续性、体素对层级连接、功能性磁共振成像

1 引言

统计网络分析和图论在研究人脑错综复杂的神经回路（即“人类连接组”）方面发挥着基础性作用（Bullmore和Sporns，2009年；Rubinov和Sporns，2010年）。大量文献揭示，人类连接组是一个组织良好但不显眼的网络，并具有智能网络（如社交网络和互联网）的图属性（Bahrami等人，2019；Cao等人，2014年）。基于图论，脑网络分析将脑连接组描述为一个图，其中皮层区域表示为节点，区域之间的连接表示为边。在这一框架下，研究者开发了大量统计模型，用于研究复杂神经连接与实验状况或临床状况之间的关联（例如，Simpson等人（2013年）；Fornito等人（2016））。这些模型有助于加深我们对脑部疾病（如阿尔茨海默病和帕金森病）潜在病理生理机制的理解，并辅助疾病诊断和治疗方案选择方面的临床预测。



(a) A demonstration of voxel-level variability within ROIs　　(b) An example of covariate-related sub-area connections

图 1： *(a)* 展示了种子点至体素分析中 *ROI* 内体素之间功能连接（*FC*）的异质性。此处使用岛叶作为种子感兴趣区。*(a)* 中的热图刻画了种子感兴趣区与扣带皮层（*a1*）或海马（*a2*）中体素之间的功能连接。扣带皮层和海马都是众所周知的感兴趣区域，但它们内部与岛叶的功能连接差异显著。*(b)* 展示了较大感兴趣区对（区域 *A*，*B*区域）中子区域对 (U_1, V_1)、(U_1, V_2) 和 (U_2, V_3) 之间协变量相关功能连接的简化示例。

在脑网络研究中，感兴趣区（ROIs）通常被视为分析的基本单位，相当于图论中的节点/顶点。区域级脑网络（RBN）分析之所以广受欢迎，是因为它具有较高的解剖一致性和计算可处理性。在考察全脑连接组时，RBN分析会将搜索维度从数万亿（$10^6 \times 10^6$）大幅减少至数千（$10^2 \times 10^2$）。然而，RBN分析依赖于ROI内

体素之间信号同质性的假设，而这一假设在现实中往往并不成立。当ROI内存在显著异质性时，RBN分析可能导致若干分析缺陷：忽视变异。简单地对ROI内体素的时间序列取平均，可能会造成体素级信息变异的损失（例如，图 1（a））；空间特异性损失。临床协变量可能仅通过扰动少部分ROI内体素对，便改变ROI对连接。在这种情况下，RBN分析无法准确区分局部变化；统计功效损失。平均处理会混合显著与不显著的体素级连接，往往会降低效应量和统计功效。

近年来，许多脑网络研究已将关注重点从RBN分析转向体素水平网络分析（Loewe等，2014；Wu等，2013）。传统的多重检验方法（如错误发现率（FDR）和族错误率（FWER）控制）不适用于高维多变量体素对，因为这些方法无法考虑感兴趣区域（ROI）中的解剖学限制以及疾病相关体素固有的系统性模式。其他一些现有方法也可能存在局限，例如未利用丰富的体素级信息来补充区域级连接特征描述，或者由于各种原因（如神经生物学结构代表不足，或种子选择过程存在偏差）而产生较难解释的结果。为解决这些局限，研究者提出了若干先进的统计方法。例如，Xia和Li，2017）和（Xia和Li，2019）通过考虑网络属性提供了局部统计推断。（Chen等，2016）提出了一种贝叶斯层次模型，用于识别与临床协变量相关的体素级连接模式，随后利用体素水平功能连接（vFC）模式推断区域级连接。这些新方法改善了推断结果并提高了局部特异性。然而，它们并不直接适用于我们关注的输入数据（即“双聚类”而非 $n \times n$ 邻接矩阵），也没有对所涉及的体素施加空间连续性约束。与RBN分析不同，空间连续性对vFC分析至关重要，原因有二：（i）它能够保持解剖学同质性，从而保留vFC结果的可解释性（Thirion等，2006年）；（ii）它能更好地控制FDR和FWER，因为表型相关的体素水平功能连接通常与脑连接组的拓扑结构内在相关（Fan等，2012）。

本研究旨在从较大的区域对中，识别空间连续子区域对之间发生改变的 vFC 模式。更具体地说，对于给定的目标区域对，我们希望提取内部子区域对，使其能够最大程度覆盖空间相邻、且与协变量相关的 vFC，同时将 FDR 和 FWER 控制在合理范围内（例如 图 1(b) ）。我们的子区域 提取 方法与其他常用的脑区划分 方法有根本区别，例如基于解剖结构的方法和数据驱动方法（如基于梯度或基于相似性的映射）（Wig 等人，2014年；Craddock 等人，2012）；这些脑区划分方法旨在将感兴趣区域划分为不同的子区域，并将每个体素分配到相应的子区域。与处理每个体素的脑区划分方法不同，我们的子区域提取方法只选择与协变量相关、且受空间连续性约束的体素子集。所有其他未选中的体素均视为协变量无关。子区域提取更适合本研究，原因如下：(i) 临床组之间与协变量相关的差异可能集中在区域 A 的某个子区域与区域 B 中由现有脑区划分方法划分出的多个子区域交集之间的 vFC 中；(ii) 通常发现区域 A 和 B 中只有一小部分体素受到影响，因此没有必要对整个感兴趣区域进行全面划分（

临床组可能聚集在区域 A 中某个子区域与 B区域 中由现有脑区划分方法划分的多个子区域交集之间的体素功能连接中；(ii) 通常发现区域 A 和 B 中只有一小部分体素受到破坏，因此无需对整个感兴趣区域进行全面脑区划分(Cao 等，2014年)。

为实现所需的子区域提取，并解决上述讨论的局限性，我们提出了一种新的统计网络框架，用于提取**S**空间 **C**约束且**C**连通的 **N**网络，以下简称 SCCN。SCCN 是一种两步法（图2），聚焦于一对感兴趣区域（例如 A 区和 B 区）之间的连接；我们认为，这些区域包含由脑部疾病导致的异常功能连接。在 **步骤 1**中，SCCN 提取空间上连贯的子区域对，使其最大限度地包含 A 区和 B 区之间受疾病影响的体素功能连接。在 **步骤 2**中，我们对每个提取出的子区域对进行正式检验，以确定在多重检验控制下其是否与协变量显著相关。如果没有发现任何子区域对达到显著性，我们则将该区域对连接视为与协变量无关。如果观察到显著结果，则可将所关注协变量与感兴趣区域对连接之间的关联追溯至由所提取体素构成的更小但精确得多的子区域。这些体素功能连接结果或可为理解疾病潜在的神经生理机制提供启示。

本文表明，SCCN 能够对真实社区结构给出一致的估计：对于较大的区域对，边分配误差可以忽略不计。我们还通过两个真实数据示例和大量模拟研究，对 SCCN 的性能进行了实证评估。具体而言，两个真实数据示例分别使用了来自一项精神分裂症研究的功能性磁共振成像（fMRI）数据（330 名参与者），以及来自一项尼古丁成瘾研究的英国生物样本库[1] 数据（3269 名参与者）。对于精神分裂症数据，我们关注显著性网络和颞叶-丘脑网络，这两个网络均被认为易受精神分裂症影响。SCCN 得到空间连贯的子区域对，其中大多数内部体素功能连接在精神分裂症中受到破坏，尤其是（背侧岛叶，前扣带皮层）对和（前部/内侧颞回，内侧丘脑）对。对于英国生物样本库数据，我们关注基底神经节、海马和岛回，这些区域被认为容易受到尼古丁影响。SCCN 在此得到空间连贯的子区域对，其中包含与成瘾相关的体素功能连接，尤其是（基底神经节内侧下部，岛叶内侧下部）对。相比之下，传统的多重检验校正方法（例如错误发现率和 FWER 控制）以及若干现有的 RBN分析，均未能通过体素级拓扑结构揭示这些发现。此外，模拟研究结果还表明，SCCN 在控制假阳性率的同时，能够有效提高统计功效和空间特异性。此外，SCCN 可轻松扩展至小型和大型数据集。

[1]英国生物样本库是一个大型生物医学数据库和研究资源，包含来自 50 万名英国参与者的深入遗传和健康信息。

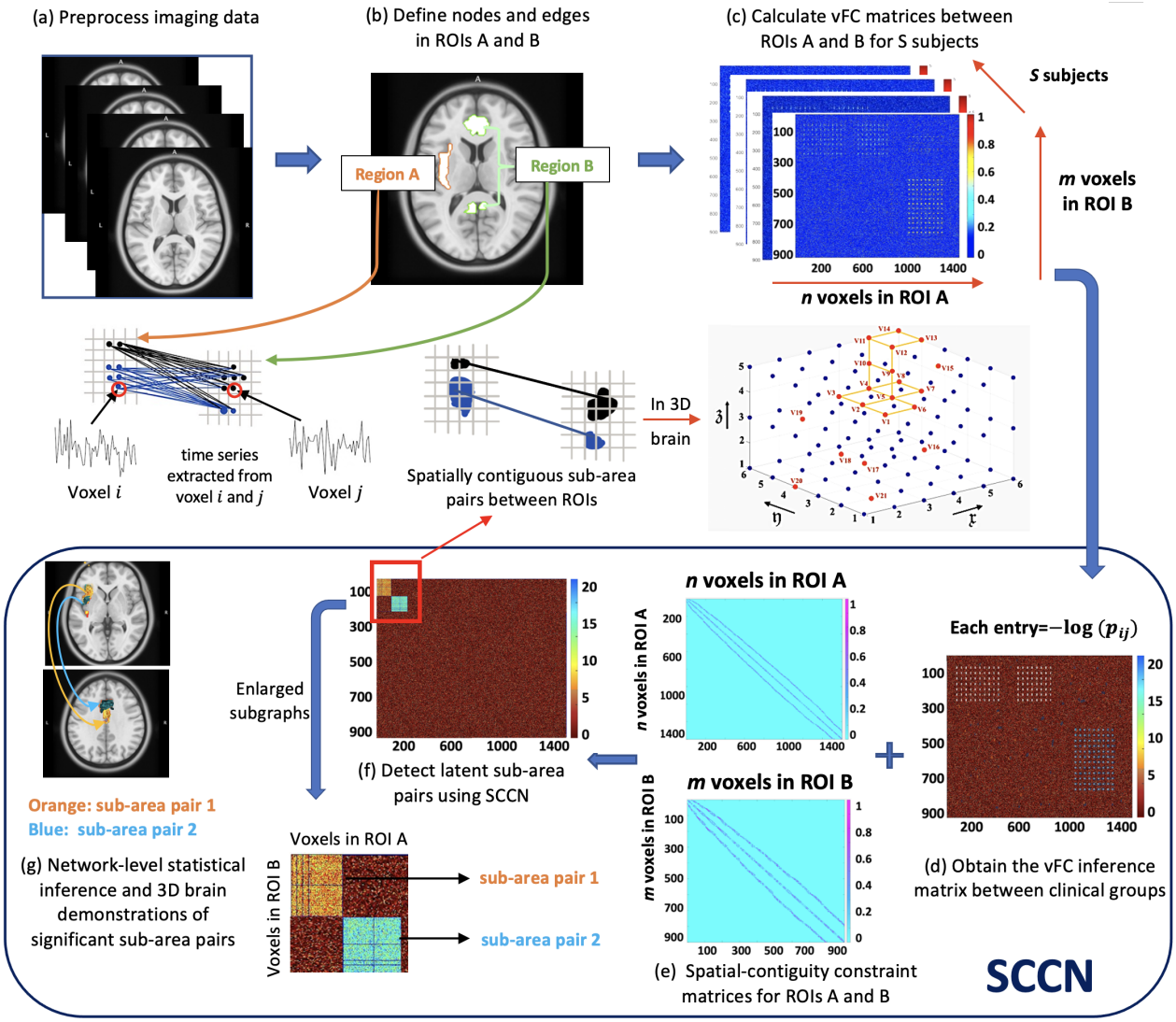


图2： *vFC*模式提取与推断流程。*(a)* 对功能性磁共振成像数据进行预处理，以去除不需要的伪影，并将数据转换至标准脑模板。*(b)* 将感兴趣区域中的体素定义为节点，并将体素之间的连接定义为边。提取每个体素的脑信号时间序列。*(c)* 针对每名受试者，计算区域 *A* 与区域 *B* 中体素之间的连接矩阵。*(d)* 计算连接推断矩阵，其中每个元素表示临床组间每条边的检验统计量。热图中颜色越热的点表示组间差异越大。*(e)* 计算感兴趣区域 *A* 和 *B* 的空间连续性约束矩阵（矩阵构建的详细说明见第 *2.1* 节）。在 *(e1)* 中，每个点表示 *3D* 坐标中的一个体素，其中红点表示感兴趣体素。由黄色线连接的体素构成空间连续区域。*(f)* 联合依据 *(d)* 和 *(e)*，检测子区域对中包含的疾病相关连接。*(f)* 通过对 *(d)* 中的节点重新排序得到，使密集改变的子网络移至顶部（即，*(d)* 和 *(f)* 是同构图）。*(g)* 进行基于*MDL*的网络层面统计推断。通过统计检验的子区域对将以高亮显示。

2 方法

2.1 背景

2.1.1 数据结构

聚焦于功能性磁共振成像数据，我们希望研究两个感兴趣区域 A 和 B 之间发生改变的体素功能连接模式，它们分别由 n 个和 m 个体素组成。对于受试者 $s \in [S] := \{1, \ldots, S\}$，令 $\mathbf{Z}^{A,s}_{n\times T}$ 和 $\mathbf{Z}^{B,s}_{m\times T}$ 分别表示感兴趣区域 A 和 B 在 T 个不同时间点的体素级血氧水平依赖（BOLD）信号矩阵。结局变量是通过 A 和 B 中体素时间序列之间的相似度矩阵量化的功能连接度量。例如，Y^s_{ij}（即 A 中体素 i 与 B 中体素 j 之间的连接强度）可通过 $Y^s_{ij} = f(Z^{A,s}_{i\cdot}, Z^{B,s}_{j\cdot})$ 计算，其中 $Z^{A,s}_{i\cdot}$ 和 $Z^{B,s}_{j\cdot}$ 分别是体素 i 和 j 的 BOLD 时间序列，而 f 是一种相似性度量（例如，经 Fisher z 变换的 Pearson 相关系数）。收集每个体素对 $(i, j) \in [n] \times [m]$ 的所有 Y^s_{ij}，即可得到区域间连接矩阵 $\mathbf{Y}^s_{n\times m}$。此外，还会观测每位受试者 s 的协变量向量 $\mathbf{X}^s_{1\times p}$，其中包含人口统计学和临床信息。

我们的目标是在体素层面识别临床/行为相关功能连接（功能连接，FC）模式。这是因为体素级发现能够以更高的统计功效以及更强的空间特异性和分辨率揭示功能连接的变化。为此，需要对具有空间约束的 $n \times m$ 体素功能连接（vFC）结果进行多变量统计推断（通常处于高维空间，例如数百万维）。我们首先检验每个结果 Y^s_{ij} 与主要关注的回归变量 $x^s_1 \in \mathbf{X}^s$（在我们的应用中为临床状态，例如患者或对照）之间的关联：

$$\mathbb{E}(Y^s_{ij}|\mathbf{X}^s) = \alpha_0 + x^s_1\beta_{ij} + \mathbf{X}^s_{1\times(p-1)}\boldsymbol{\alpha},$$

其中，β_{ij} 是 x^s_1 的系数，$\boldsymbol{\alpha}$ 是其余协变量$\mathbf{X}^s_{1\times(p-1)}$ 的系数向量（例如年龄、族裔等）。我们记为 $\boldsymbol{\beta} := \{\beta_{ij}\}_{i\in[n],j\in[m]}$，并旨在系统地提取 $\boldsymbol{\beta} \neq 0$ 具有高准确性的体素功能连接。我们进一步用连接推断矩阵 $\mathbf{W}_{n\times m}$总结 $\boldsymbol{\beta}$ 的显著性水平。$\mathbf{W}_{n\times m}$ 的每个条目均通过$\mathbf{W}_{ij} = -$的对数p_{ij}计算，其中p_{ij}是 β_{ij}的 p值。在神经影像学统计中，$\boldsymbol{\beta} \neq 0$ 的选择不仅由统计显著性水平决定，也受空间约束影响。除这两个因素外，$\boldsymbol{\beta}$ 还与 ROI A 和 B 之间潜在的二部图 $n \times m$ 内在相关。因此，我们需要图结构信息和空间信息共同协助识别 $\boldsymbol{\beta} \neq 0$ 的体素功能连接。下面给出详细的图结构和空间构建方法。

2.1.2 图表示

为了阐释复杂的体素对连接组，我们考虑推断矩阵 $\mathbf{W}_{n\times m}$所依托的二部图结构 $G = \{U, V\}$。节点集 U和 V分别表示感兴趣区 A 和感兴趣区 B 中的体素，其中 $|U| = n$ 和 $|V| = m$。我们假设，在对功能性磁共振成像数据进行空间标准化和配准后，所有受试者共享相同的节点集，即 $(U^s, V^s) \equiv (U, V)\ \forall s \in [S]$。

2.1.3 空间连续性

我们数据集中的每个节点都对应于三维脑成像中的某个空间位置上的体素（例如，图 2(e1)）。当我们将检测到的每个体素子组映射回三维脑空间时，希望这些体素呈现为空间相邻簇（即连通分量）。这种预期用形式化语言表述，即称为空间连续性。具体而言，我们在感兴趣区 A 内的所有节点之间定义一个“基础结构图” $\mathcal{S}_A$ ，以体现空间连续性。$\mathcal{S}_A$ 中的每个条目 $S_{ii'}$ 都是感兴趣区 A 中体素 i 与 i' 之间的空间邻接指示变量；当 $d_{ii'} \leq \varepsilon$时， $S_{ii'} = 1$ ，否则为 $S_{ii'} = 0$（$d_{ii'}$ 是体素 i 与 i'之间的欧几里得距离）。例如，在三维网格空间中，当将 ϵ 设为 $\sqrt{3}$时，立方体中的一个质心体素 i 将有 26 个周围体素 i' ，满足 $S_{ii'} = 1$。对于感兴趣区B 内的节点，我们以类似方式定义并解释 $\mathcal{S}_B$ 。在实现 SCCN 时，将使用$\mathcal{S}_A$ 和 $\mathcal{S}_B$ 来规定空间连续性约束。我们在附录 A.1 中给出了空间连续性、$\mathcal{S}_A$和$\mathcal{S}_B$ 更为严格的数学定义。

我们提出 SCCN 模型，通过综合考虑体素对层面的统计显著性、底层图结构和空间约束的信息，系统地筛选 $\beta_{ij} \neq 0$ 的体素功能连接。我们将这些信息整合为加权图 $\mathcal{G} = \{W, \mathcal{S}_A, \mathcal{S}_B\}$，作为本方法的输入。

2.2 从感兴趣区对中检测密集改变的子区域对

2.2.1 空间连续性约束的目标函数

据报道，与感兴趣区 A 中体素对应的节点集 U 可划分为彼此互不重叠的子区域 $\{U_c\}$，记为 $U = \bigoplus_{c=1}^{C} U_c$ （Eickhoff 等人，2015）。类似地，对于感兴趣区 B，我们有 $V = \bigoplus_{d=1}^{D} V_d$。在本文中，我们旨在提取主要包含疾病相关体素对的子区域对{（U_c、 V_d）}，并将其称为“密集改变”的子区域对。形式上，如果 $\sum$（i、 j）$\in$（U_c、 V_d） I（$\beta_{ij}\neq 0$） $|U_c|\ |V_d|$ $\gg \sum$（i、 j）$\in$（U_c'、 V_d'） I（$\beta_{ij}\neq 0$） $|U_c'||V_d'|$，则子区域对（U_c、 V_d）被视为密集改变，其中 U_c' 和 $\overline{V_d'}$ 分别是节点集 U_c 和 $\overline{V_d}$ 的补集。因此，我们受到启发，设计一个正则化目标函数，以生成连接推断矩阵 W 所对应的棋盘状网络结构。该网络结构对 W 进行重新排列，并揭示出来自（U、 V）的密集改变 {（U_c、 V_d）}对。此外，我们对 U_c 和 V_d 施加空间连续性约束，以提高生物学可解释性，并禁止孤立的假阳性边。最终，目标函数表述如下：

$$\underset{\substack{C,\ D,\ U=\bigoplus_{c=1}^{C} U_c, V=\bigoplus_{d=1}^{D} V_d \\ (U_c, V_d \text{ subject to spatial contiguity})}}{\operatorname{argmax}} \int \sum_{c=1}^{C}\sum_{d=1}^{D} \left\{ \log \frac{\sum_{i\in U_c, j\in V_d} W_{ij}\cdot I(W_{ij} > r)}{|U_c||V_d|} + \lambda \log(|U_c||V_d|) \right\} g(r)dr, \tag{1}$$

其中 $\lambda \in [0,1]$ 是一个调节参数，r 是一个阈值，低于该阈值时，W_{ij}上不存在疾病相关效应，且 g（r）是 r的分布函数。（r）和 λ 均可根据先验知识或第2.2.2节提出的数据驱动方法进行选择。

调节参数 λ的取值范围为 $[0,1]$：当 $\lambda = 0$时，最大化 (1) 等价于最大化 $f_1 = \frac{\sum_{i \in U_c, j \in V_d} W_{ij} \cdot I(W_{ij} > r)}{|U_c||V_d|}$，这是连接密度的一种常用定义；当 $\lambda = 1$时，最大化 (1) 就是最大化 $f_2 = \sum_{i \in U_c, j \in V_d} W_{ij} \cdot I(W_{ij} > r)$，该函数量化了子区域对 (U_c, V_d) 所包含的显著体素对的数量。直接优化连接密度 f_1 往往会检测到规模极小的稠密子图，而优化 f_2 则可能产生规模过大的子图。定理 1 表明，函数 (1) 能够一致地估计目标拓扑结构（边诱导的子区域对集合），其意义在于，在较大的区域对中，边分配误差可以忽略不计。大量模拟研究还表明，在检测子图时，函数 (1) 能够很好地平衡规模与密度。

2.2.2 给定(1) 时目标函数的优化 $g(r)$ 和 λ

在本节中，我们聚焦于在给定 $g(r)$ 和 λ 的配置下优化函数 (1)，其中它们分别是阈值 r 的密度函数和式 (1) 中的调节参数。下一节将讨论如何确定 $g(r)$ 和 λ 。遗憾的是，即使给定 $g(r)$ 和 λ，直接优化式 (1) 仍然是一个 NP难问题。因此，由于该问题具有非凸性，无法使用梯度下降等传统优化方法。这里，我们提出一种求解式 (1) 的替代策略。其核心思路是将 W 与空间连续性约束相结合，然后通过迭代过程，利用改进的谱聚类算法估计目标社区结构。如前所述，目标网络结构是从 (U, V) 中划分出的 $\{U_c, V_d\}$（即边诱导的子区域对的集合，换言之，也就是 U_c 和 V_d 的体素归属），其中 $U = \bigoplus_{c=1}^{C} U_c$ 和 $V = \bigoplus_{d=1}^{D} V_d$。

根据谱聚类算法，对$\mathbf{W} = \mathbb{U}\mathbf{\Sigma}\mathbb{V}^\top$ 的拉普拉斯矩阵进行奇异值分解，然后对 $\mathbb{U}$ 和 $\mathbb{V}$ 进行聚类，即可分别得到区域 A 和 B 的划分。现在，由于 $\mathbb{V}$ 是$\mathbf{W}^\top\mathbf{W}$的特征向量，对 $\mathbf{W}^\top\mathbf{W}$ 的拉普拉斯矩阵进行谱聚类即可得到 B区域的划分。类似地，对$\mathbf{W}\mathbf{W}^\top$ 的拉普拉斯矩阵进行谱聚类即可得到区域 A 的划分。因此，我们的社区检测算法可以基于$\mathbf{W}\mathbf{W}^\top$和$\mathbf{W}^\top\mathbf{W}$进行。接下来，为了将空间连续性约束纳入优化过程，我们使用前文介绍的两个区域内“基础结构图” $\mathcal{S}_A$ 和 $\mathcal{S}_B$。具体而言，我们定义

$$\mathbf{W_A} = \mathbf{W}\mathbf{W}^\top \odot \mathcal{S}_A \quad \text{and} \quad \mathbf{W_B} = \mathbf{W}^\top\mathbf{W} \odot \mathcal{S}_B, \tag{2}$$

其中 $\odot$ 是逐元素乘积。正如Kamvar 等人（2003年）和Craddock 等人（2012）所指出的，$\mathcal{S}_A$ 和 $\mathcal{S}_B$ 会将所有非相邻体素对之间的相似性置为零，从而切断图中连接孤立体素的边。基于此，n 通过 n矩阵$\mathbf{W}_A$（ii'）（其中 i 和 i' 是 A 中的两个体素）在 A 中的体素空间相邻且具有相似的特征谱（该特征谱描述其与 B区域中体素的连接情况）时取值更大。空间连续性

约束使我们的方法能够产生更符合神经生物学背景的结果，即相邻神经元群体具有连贯性（Thirion 等人，2006年）。

现在，我们可以使用谱聚类算法为$\mathbf{W}_A$拟合随机区块模型（并为$\mathbf{W}_B$拟合另一个随机区块模型），然后通过网格搜索找到函数（1）的最优解。我们还进一步考察估计得到的U_c和V_d值是否满足空间连续性约束；实证结果表明，这些约束通常都能得到满足。因此，无需再对约束进行任何修改。我们在算法1中正式介绍聚类过程。

算法1 给定1的目标函数优化，其中给定 λ

. 1: 过程 ALGORITHM **（输入：** λ、$\mathcal{G} = \{\mathbf{W}\ \mathcal{S}_A$、$\mathcal{S}_B\}$**） 2: 函数 SCCN.partition（**λ**，** $\mathcal{G}$ **） 3: 对于** $C = 1$**，2，...，**$|U|$**，执行 4: 将** $\mathbf{W}_A$ **进行比率割谱聚类，划分为** C **个网络：** $U = \bigoplus_{c=1}^{C} U_c$**5: 对于** $D = 1$**，2，...，**$|V|$**，执行 6: 将** $\mathbf{W}_B$ **进行比率割谱聚类，划分为** D **个网络：** $V = \bigoplus_{d=1}^{D} V_d$**7: 将网络集合** U **和** V **代入目标函数（1），并获得输出值 8: 结束循环 9: 结束循环 10: 返回使输出值最大的** C、D、$U = \bigoplus_{c=1}^{C} U_c$ **和** $V = \bigoplus_{d=1}^{D} V_d$ **11: 结束函数 12: 结束过程**

子图检测的一致性。在引理1中，我们首先证明，给定真实的子区域数量C^*和D^*，对目标函数（1）进行优化所得的解，能够对目标社区（$\{U_c, V_d\}$）（即边诱导的子区域对集合）的拓扑结构给出一致估计；具体而言，在非常大的二部图中，假阳性边分配可以忽略不计 $G =$（U，V），$|U| \to \infty$，且 $|V| \to \infty$。随后，我们基于定理1和1 以及2，证明算法1在优化函数（1）时的收敛性。在定理1中，我们证明该算法能够对子区域数量C和D给出一致估计。在定理2中，我们证明算法1的实现会收敛到目标函数（1）的最优解。

为展示理论结果，我们考虑以下设定。设 $\{e_{ij}^1\}$ 和$\{e_{ij}^0\}$分别为正边（例如疾病相关边）和负边的集合。对于邻接矩阵 $\mathbf{W}$，我们假设 $w_{ij}|e_{ij}^1 \overset{\text{iid}}{\sim} f_1$ 且 $w_{ij}|e_{ij}^0 \overset{\text{iid}}{\sim} f_0$，其中 f_1 和f_0 是两个概率密度函数，其均值和方差分别为 (μ_1, σ_1^2) 和 (μ_0, σ_0^2)。此外，设 $\mathcal{M}^*$ 为边的真实归属（落在子区域对 (U_c, V_d) 中的边的社区索引 ^）。进一步地，设 $\mathcal{M}_{(\hat{C}, \hat{D})}$ 为归属

ˆ ˆ由函数 (1) 估计， C 区域 A 中的子区域数，以及 D B区域中的子区域数。

引理 *1.*（已知子区域数量 C^* 和 D^*时的一致性）。假设 $\mathbb{E}(WW^{\mathbf{T}})$ 的秩为 C^*，其最小绝对非零特征值至少为 Λ_A；且 $\mathbb{E}(W^{\mathbf{T}}W)$ 的秩为 D^*，其最小绝对非零特征值至少为 Λ_B。进一步假设，对于某个 $d \leq max(log n/n, log m/m)$，$max(\mu_0, \mu_1, \sigma_0^2, \sigma_1^2) \leq d$。那么，若存在 $(2+\varepsilon_A)\frac{ndCD}{\Lambda_A^2} < \tau_A$ 和 $(2+\varepsilon_B)\frac{mdCD}{\Lambda_B^2} < \tau_B$，其中 $\tau_A, \tau_B, \varepsilon_A, \varepsilon_B > 0$，使得使函数 *(1)* 取最大值的输出 $\hat{\mathcal{M}}_{(\hat{C},\ \hat{D})}$ 与潜在社区结构所对应的真实成员关系 $\mathcal{M}^*_{(C^*,\ D^*)}$ 在置换意义下具有一致性。

ˆ ˆ等价地，令 S_c、S_d 分别为子图 G_c、G_d（由ˆU_c 和 V_d诱导）的估计节点集。则 $S_c \cap U_c$ 表示 G_c 中其归属可以得到保证的节点。ˆ $S_d \cap V_d$遵循相同的定义。以至少 $1 - \max(n, m)^{-1}$ 的概率，在允许置换的情况下，我们有

$$\sum_{c=1}^{C}\sum_{d=1}^{D}\left[1 - \frac{\left|\ (\hat{S}_c \cap U_c) \bigotimes (\hat{S}_d \cap V_d)\ \right|}{|U_c|\ |V_d|}\right] \leq \tau_A^{-1}(2+\varepsilon_A)\frac{ndCD}{\Lambda_A^2} + \tau_B^{-1}(2+\varepsilon_B)\frac{mdCD}{\Lambda_B^2},$$

其中，$\bigotimes$ 表示连接其左右两侧两个节点集的边集。

定理 1。*(*网格搜索 C、D 时的一致性*)*。设子图对$|U_c| \times |V_d|(\forall c = [C^*], d = [D^*])$ 的大小由概率为 $\pi = (\pi_1, \ldots, \pi_{C^* \times D^*})$的多项分布生成。假设 $\exists \delta > 0$，使得

$$\mu_1 > \mu_0 \frac{1+\delta}{1-\delta}\left(1 + \sqrt{1 + \frac{\pi_{\min}^2}{\pi_1^2 + \cdots + \pi_{C^* \times D^*}^2}}\right),$$

那么，在引理 *1* 和调节参数 $\lambda = 0.5$的条件下，误分类的边数 N_{edge} 满足

$$N_{\text{edge}} = o_p(n_{\min} * m_{\min}) \text{as } |U|, |V| \to \infty,$$

其中 n_{m}在 $m_{\min}$ 是 U和 V中可能的最小子图的大小，分别 y.

˜ ˜**定理 2.** *(*算法 *1*的收敛 $\tilde{U} = \bigoplus_{c=1}^{C} U_c, \tilde{V} = \bigoplus_{d=1}^{D} V_d$ 是对 $\mathbf{W}_A$ 和 $\mathbf{W}_B$ 进行比率割谱聚类后得到的划分，且使函数˜ ˜ ˜ ˜*(1)* 在聚类数为 C, D时达到最大值。那么 U, V几乎必然收敛到真实的社区结构，其中分配给每个子双聚类的假阳性边可忽略不计。

证明。引理 1及定理 1 和 2 的证明见附录 B .

总之，上述结果为所提出的目标函数（1）的解及算法 1 收敛至目标社区结构（$\{U_c, V_d\}$）提供了理论依据。此外，在多种设定下开展的大量模拟分析表明，面对范围广泛的不同样本量，SCCN 均能准确揭示真实的社区结构，同时保持较低的假阳性率和假阴性率。

2.2.3 确定 $g(r)$ 和 λ

确定 g（r）。根据（Efron，2012），我们可以将 g（·）选为阈值 $\{r_1, \ldots, r_p\}$上的离散分布。一个简单的例子如下。假设由预先选定的阈值 r_1、r_2和 r_3 得到的体素对层面的假发现率分别为 0.20、0.10 和 0.05。然后，我们可以对产生较低错误发现率的 r_p 赋予更大的概率质量，例如，g（r_1）$= 0.1$，g（r_2）$= 0.3$，以及 g（r_3）$= 0.6$。此外，r_1、r_2和 r_3可以从磁共振成像研究中常用的阈值中选取，例如 –log(0.005) 和–log(0.001)（W_{ij} 是筛选后的 –log，而 r 是 W_{ij}的阈值）。

选择 λ。如前所述，调节参数 λ 用于平衡子图大小与连接密度，因此在我们的方法中起着至关重要的作用。较大的 λ 会促使 $|U_c|$ 和 $|V_d$|较大，而较小的 λ 则会更严格地限制（U_c, V_d）对的连接密度。从本质上说，λ 的选择与 β_{ij}的网络结构有关。实践中，我们从许多数据集中观察到，系数 $\beta_{ij} \neq 0$ 通常呈现块模型。为体现这一点，我们假设以下层次模型。假设存在一个非随机的潜在超参数 $\boldsymbol{\beta} \in \mathbb{R}^{n \times m}$，其所有元素均非零。我们可以从具有块 $\{(U_c, V_d)\}$ 和相应连接概率 $\{\pi_{ij}\}$的二部随机块模型中生成二部相似性矩阵 $\eta \in \{0,1\}^{n \times m}$，使得 η_{ij} ~伯努利分布(π_{ij})彼此独立，其中

$$\pi_{ij} = \begin{cases} \pi_{cd}(\lambda) & i \in U_c, j \in V_d, \\ \pi_0(\lambda) & \text{otherwise.} \end{cases}$$

我们选择使该块模型的似然达到最大的 λ 值。在实际应用中，η_{ij} 值无法直接观测，因此我们用$\eta_{ij}(r_0) := I(w_{ij} > r_0)$替代它们。$\lambda$ 的对数似然函数为：

$$l_\lambda(\pi_{cd}, \forall c \in [C], d \in [D] | \eta_{ij}(r_0)) = \sum_{c,d} \sum_{(i,j) \in U_c \times V_d} \eta_{ij}(r_0) \log \pi_{cd} \quad + (1 - \eta_{ij}(r_0)) \log (1 - \pi_{cd}) .$$

为消除选择阈值 r_0时的任意性，我们根据上述方法确定的先验分布 $g_0(r_0)$，对似然函数关于 r_0 进行积分。由此得到以下准则：

$$\lambda_{\text{optimal}} = \text{argmax}_\lambda \Big\{ \int \max_{U_c, V_d, \pi_{cd}} \quad l_\lambda^r \left(\pi_{cd}(\lambda), \forall c = [C], d = [D] \mid \eta_{ij}(r_0)\right) \; g_0(r_0) dr_0 \Big\}.$$

我们正式给出选择调节参数的流程 λ，见算法 2。算法的总体复杂度为 $O(Knm)$，其中 K 是 λ的一个充分搜索范围，$n = |U|$和 $m = |V|$。由于不同临床组之间、跨 S名受试者的推断结果均由 **W**捕获，算法复杂度不再涉及样本量 S，这表明 SCCN 可扩展至大型数据集。此外，聚类算法通常需要计算一个可能具有高通量的相似性矩阵的前 K 个特征向量。经过筛选并施加空间连续性约束后，我们的输入相似性矩阵 $\mathbf{W}$ 是稀疏的（处理后通常仅有 0.2%–5.0% 的边为非零项），从而显著降低了计算开销。不过，值得注意的是

由于我们的算法基于单个区域对，因此在研究多个不同区域对时，计算负担可能会很重，尤其是在需要进行全脑分析时。

算法2 针对 λ的网格搜索。

过程 ALGORITHM
  对于 $0 \leq \lambda \leq 1$ **执行**
    返回 $U = \bigoplus_{c=1}^{C} U_c$ 和 $V = \bigoplus_{d=1}^{D} V_d$ 通过算法1
    对于 $r_0 = (r_0)_1$ 到 $(r_0)_q$ **do**
      计算对数似然： $l_\lambda(\hat{\pi}_{c\times d}^{\mathrm{MLE}}, \quad \forall c = [C], \quad d = [D] \mid \eta_{ij}(r_0))$
    结束循环
    对 r_0 积分，计算对数似然：
    $l_\lambda = \sum_{i=1}^{p} L_\lambda \left(\hat{\pi}_{c\times d}^{\mathrm{MLE}}, \quad \forall c = [C], \quad d = [D] |\ \eta ij(r)\right)\ g((r_0)_i)$
  结束循环
  返回 $\hat{\lambda}$ 使 l_λ 最大化
结束过程

2.3 $\{(U_c, V_d)\}$ 对的统计推断

回顾一下，我们的最终目标是从 $\{(U_c, V_d)\}$ 中，基于当前已获得的块划分 $\{U_c, V_d : c \in [C], d \in [D]\}$ 提取少数连接最密集的子图对。一个自然的思路是检查每个（U_c, V_d）对，并对其进行统计检验，备择假设为子图 $U_c \otimes V_d$ 的连接密度异常高。这里，我们设计了一种具有 FWER 控制的聚类级置换检验（Nichols 和 Holmes，2002）。假设如下：

$$H_0 : \text{Subgraph } U_c \otimes V_d \text{ is not unusually dense,}$$
$$H_a : \text{Subgraph } U_c \otimes V_d \text{ is unusually dense.}$$

更具体地说，在 H_0下，$U_c \otimes V_d$ 的连接密度应接近通过随机打乱二部图中的边所得到的子区域对的连接密度。类似于（Gr¨unwald，2007），我们设计了以下最小描述长度（MDL）检验统计量：

$$\mathrm{MDL}\,(U_c, V_d) = \log_2 \left[\binom{n}{|U_c|} \binom{m}{|V_d|} \right] + \left(\frac{1-\mu_1^2}{2\ln 2} - L_\zeta \right) |U_c| \times |V_d|,$$

其中，μ_1 是 ζ_{ij} 内各条边的逐边检验统计量的均值 $U_c \times V_d$，而 $L_\zeta = -\int \phi(\zeta_{ij}) \log_2 (\phi(\zeta_{ij}))\, d\zeta_{ij} + C$ 是基于标准正态分布的信息熵度量 ϕ，针对 ζ_{ij}。基于MDL的检验统计量及其与我们推断目标之间联系的详细推导见附录 B.4。我们在算法 3中，正式给出了针对每个观测到的子区域对（U_c, V_d）的聚类级置换检验。该算法中的置换次数 H 可根据样本量、目标计算开销和检验精度确定。例如，$H = 1000$。

与传统的多重检验校正方法（例如错误发现率和家族错误率）相比，基于MDL的簇级置换检验在真实数据示例和模拟中抑制了假阳性结果，并展现出更高的统计功效。

算法3 针对每个 (U_c, V_d) 对的基于MDL的簇级置换检验 1：过程 ALGORITHM**2：使用真实协变量标签，对每个生成的 (U_c, V_d) 对计算** $T_{c,d}^0 =$ **MDL** (U_c, V_d) **3：对于** $h = 1, \ldots, H$**，执行 4：置换协变量标签并获得新的推断连接矩阵** $\mathbf{W}^h$**5：将** $\mathbf{W}^h$ **代入算法1和2，获得** $U^h = \bigoplus_{c=1}^{C} U_c^h$ **和** $V^h = \bigoplus_{d=1}^{D} V_d^h$**6：返回** $T^h =$ **max** $\left($ **MDL** $\left(\mathbf{U}_c^h, \mathbf{V}_d^h\right)\right)$**7：结束循环 8：计算每个观测到的 ($U_c, V_d$) 对的** p**-值：** $P_{c,d} = \sum^{I}$ $(T^h > T_{c,d}^0)$ $_H$**9：根据** $P_{c,d}$ **在预定的** α**水平下，返回每个观测到的 (U_c, V_d) 对的显著性 10：结束过程**

3 实验

在本节中，我们将 SCCN 应用于两个真实数据集，以研究特定临床情境下的体素级连接异常。数据集1包含一项精神分裂症（SZ）研究中的330名参与者，其功能性磁共振成像数据采集于美国马里兰州巴尔的摩。数据集2包含一项尼古丁成瘾研究中的3269名参与者，其功能性磁共振成像数据来自英国生物样本库数据库。

3.1 精神分裂症研究

我们的主要数据集包含330名个体，其中包括148名精神分裂症患者（男/女 84/64，年龄 37.5 ± 14.4）和182名健康对照者（男/女 80/102，年龄 37.0 ± 16.1）。这些参与者是为一项正在进行的大型研究招募的，该研究考察精神分裂症认知缺陷的影响。具体而言，该研究探讨认知缺陷如何导致精神分裂症患者出现功能障碍，以及认知缺陷与服务于认知的功能网络改变有何关联。2004年至2016年间，所有受试者均按照统一的招募标准，在巴尔的摩大都会区的当地研究中心接受评估，包括神经学和临床评估。有关参与者人口统计学信息、招募过程、影像采集和功能性磁共振成像预处理流程的详细信息，见附录C.1。

3.1.1 显著性网络连接受损

临床背景 显著性网络主要由双侧岛叶和扣带皮层构成，与精神分裂症的若干核心症状有关。大量神经影像学研究文献表明，精神分裂症患者在信息处理过程中，显著性网络的连接受到扰乱（Palaniyappan 等人，2012）。因此，我们打算聚焦于双侧岛叶和扣带皮层，研究二者之间受精神分裂症影响的vFC模式。具体而言，我们希望提取受精神分裂症影响的边，这些边连接岛叶内空间连贯子区域中的体素与扣带皮层内空间连贯子区域中的体素。通过数据驱动方式提取由精神分裂症中的vFC异常所致的子区域，或可为更有效的临床治疗提供见解（例如，经颅磁刺激或深部脑刺激疗法）。

我们基于 Brainnetome图谱（Fan 等，2016）标注了双侧岛叶和扣带皮层（左侧岛叶：1762 个体素；右侧岛叶：1577 个体素；扣带皮层：5768 个体素）。首先，我们为每名受试者计算了左侧/右侧岛叶与扣带皮层之间的体素功能连接矩阵。体素功能连接矩阵中的每个条目均为 Fisher z变换后的 Pearson 相关系数。接下来，我们基于所有受试者获得了群体水平统计推断矩阵$\mathbf{W}^L_{1762\times5768}$ 和 $\mathbf{W}^R_{1577\times5768}$ ；$\mathbf{W}^L$ 和 $\mathbf{W}^R$ 中的每个条目均赋予一个负对数 p值，用于量化 SZ 组与健康对照组之间体素功能连接差异的证据。随后，我们将 SCCN 和基于MDL的检验分别应用于 $\mathbf{W}^L$ 和 $\mathbf{W}^R$，在空间连续性约束下分别提取左侧/右侧岛叶与扣带皮层之间的异常子区域对。最后，我们将结果与采用现有常用方法获得的结果进行了比较。

网络层面结果 vFC 推断矩阵 W^L 中的每个元素为 W^L_{ij} =–log(p^L_{ij})，其中 p^L_{ij} 是 p值，用于检验左侧岛叶与扣带皮层之间（i, j）体素对的病例对照组间体素功能连接差异（图 3(L1)）。随后，我们使用预先选定的阈值（例如，$p = 0.05$）对 W^L 进行筛选：$\mathrm{W}^L_{ij} = (\mathrm{W}^L)_{ij} \cdot I\left((\mathrm{W}^L)_{ij} \leq -\log(0.05)\right)$.
筛选后的推断矩阵 $\mathbf{W}^L$ 能够有效排除大多数无信息的假阳性边，同时保留较高比例的真阳性边（Fan 和 Lv，2008；Li 等人，2012b）。类似的设置适用于 $\mathbf{W}^R$ （图 3(R1)）。运行算法2得到 ˆ 的最大似然估计（MLE） $\lambda_L = 0.625$ ，对应于 $\mathbf{W}^L$ ，以及 ˆ $\lambda_R = 0.75$, 对应于 $\mathbf{W}^R$。

ˆ ˆ ˆ 给定估计得到的 λ，算法1为 W^L返回的簇数量为 $C_L = 135, D_L$ =107，而为 W^R返回的为 ˆ $C_R = 225$, ˆ$D_R = 226$ 。基于MDL的检验在 W^L 中检出九个异常子区域对，在 W^R 中检出十个异常子区域对（在图3（L2）和（R2）中以红色标出）。图4展示了 W^L 的检测结果三维展示（采用基于MDL的置换检验，显著性水平为0.05）。图4还标明了精确的大小、 p值和位置等信息。所有提取出的子区域对均呈现出良好组织的拓扑结构。总体而言，W^L 中异常的vFC模式集中分布于背侧岛叶与前扣带皮层（ACC）之间。W^R 的详细检测结果见附录C.2。

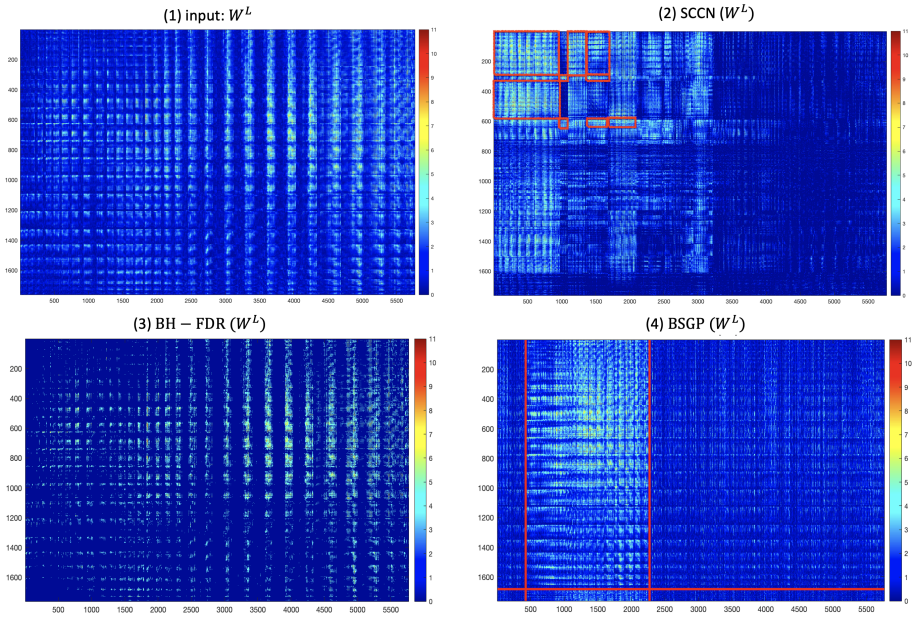


图 3： *(1)* $\mathbf{W}^L$ 的热图：行和列分别对应左侧岛叶和扣带皮层中的体素。颜色越热，表示经其他协变量校正后，临床组之间差异表达的体素对越多。*(2) SCCN* 得出的结果：通过基于*MDL*的置换检验的正向子区域对以红框标出。红框外有许多 p 值较小的边（例如左下角），因为它们与框内的边不具备空间连续性，因此被 *SCCN* 自动排除。*(3) BH-FDR* 得出的结果：对于 $q = 0.05$，未检测到子区域对。*(4) BSGP* 得出的结果：仅检测到一个有信息量但密集程度低得多的子区域对。该子区域对也缺乏空间连续性和特异性。

检测到的子区域的生物学解释 检测到的子区域由若干广为人知的脑区组成，人们认为这些脑区常与精神分裂症相关，其中尤为突出的是前岛叶（AI）和前扣带皮层（ACC）。最能引发AI活动的情绪，如愤怒和恐惧，恰恰是精神分裂症患者往往最难识别的情绪（Wylie和Tregellas，2010年）。此外，研究发现，精神分裂症患者的前扣带皮层中神经元、轴突和突触的密度异常（Arnold和Trojanowski，1996年）。检测到的所有异常边在精神分裂症患者中均表现为连接减弱或未发生变化。这与医学发现一致：精神分裂症是一种“连接异常”障碍，主要表现为显著性网络中的功能连接（FC）减弱（Lynall等，2010年），不过不能完全排除药物影响。施加的空间连续性约束有助于识别双侧岛叶和扣带皮层中能够最大程度涵盖疾病相关vFC的脑子区域。这些新发现提高了显著性网络中精神分裂症相关连接异常的空间特异性，并可能为未来的治疗提供指导。

与现有方法的比较 为进行比较，我们进行了 Benjamini–Hochberg 错误发现率（BH-FDR）逐边校正，以及一种常用的双聚类校正。

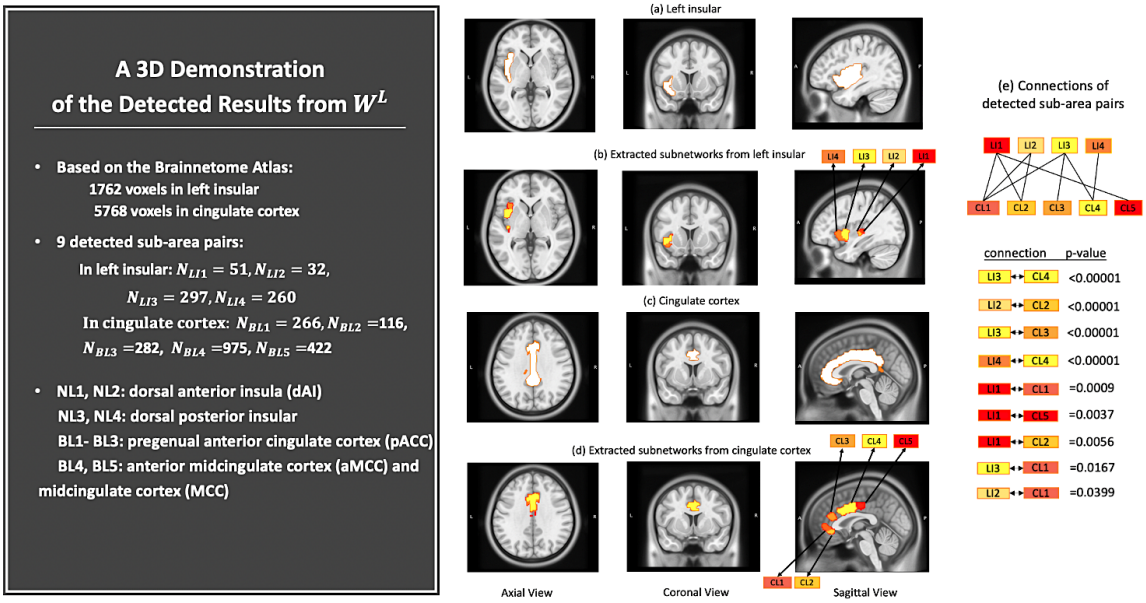


图4： 从$\mathbf{W}^L$中检测到的子区域对的三维展示。设 LI_i 为从左侧岛叶中检测到的第 i个断连子区域，它与扣带皮层中的第 j个子区域相连， CL_j。设 N_{LI_i} 表示子区域 LI_i中的体素数量，类似地，N_{CL_j}表示 CL_j的体素数量。*(a)(c)*展示了原始左侧岛叶和扣带皮层的图像；*(b)*展示了精神分裂症影响的左侧岛叶子区域，这些子区域与*(d)*中高亮显示的扣带皮层子区域相连；*(e)*展示了从$\mathbf{W}^L$ 中检测到的子区域之间的互连结构，以及基于*MDL*的置换检验得到的关联 p值。附录*C.2*中提供了$\mathbf{W}^R$ 检测结果的三维展示。

聚类算法，即二分谱图划分（BSGP），以簇为单位进行。首先对体素功能连接（vFC）与精神分裂症状态进行初步相关性分析，发现 $\mathbf{W}^L$ 中有 17.84% 的边具有 $p < 0.005$ 显著性，其中 $p = 0.005$ 是神经影像学研究中常用但未经校正的阈值（Derado等人，2010年）。应用 BH-FDR校正后，使用 $q = 0.01$ 作为阈值时发现 9.45% 的边具有显著性（图 3(L3)），且未揭示出社区结构。对于 $\mathbf{W}^R$，有 13.50% 的边的 p 值小于 0.005；应用 BH-FDR校正并采用 $q = 0.01$ 后，仅发现 3.61% 的边具有显著性（图 3(R3)）；同样，$\mathbf{W}^R$ 中未发现社区结构。对 $\mathbf{W}^L$ 和 $\mathbf{W}^R$ 应用 BSGP 时，仅检测到一个异常子区域对（图 3(L4) 和 (R4)），其中包含了超过 36.80% 的 $p > 0.005$ 边，比例高于 SCCN。与现有方法相比，SCCN 得到的与精神分裂症相关的体素功能连接更为密集，且包含在空间连续的子区域对中，并具有更强的拓扑结构。

3.1.2 颞叶-丘脑连接受损

与精神分裂症患者显著性网络连接减少相反，许多研究表明，精神分裂症患者与多个感觉运动区域的丘脑连接增强，其中最显著的是颞回（Ferri等人，2018；Cetin等人，

2014年）。具体而言，丘脑与颞中回的连接与许多精神分裂症的核心特征（如幻觉和妄想）呈正相关。因此，我们旨在利用SCCN，在精神分裂症患者右半球颞中回与双侧丘脑之间发现一些新特征。根据Brainnetome图谱，右侧颞中回包含3566个体素（标记为82、84、86和88），双侧丘脑包含3275个体素（标记为231–246）。我们计算了临床组之间的vFC连接推断矩阵$\mathbf{W}^{(\mathrm{Tem}_{\mathrm{right}},\text{丘脑}_{\mathrm{left}})\ 3566\times1727}$和$\mathbf{W}^{(\mathrm{Tem}_{\mathrm{right}},\text{丘脑}_{\mathrm{right}})\ 3566\times1548}$，随后实施了SCCN。由于篇幅有限，我们在附录C.3中提供了所有参数选择及密集改变的子区域对的结果。

3.2 尼古丁成瘾研究

我们的主要数据集包含来自英国生物样本库数据库的3269名参与者，其中包括1353名持续吸烟者（男/女：2653/616，年龄：48.6 ± 15.3）和1916名既往轻度吸烟者（男/女：1187/729，年龄：32.9 ± 18.1）。具体而言，我们将目前每天吸烟超过十支香烟的参与者定义为当前吸烟者（即尼古丁成瘾者）。我们将那些表示过去只尝试过几支香烟、但目前不依赖尼古丁产品的人定义为既往轻度吸烟者（即对照组）。通过考察两组之间不同的神经连接模式，我们或许能够更深入地了解尼古丁依赖的内在神经学机制，从而帮助吸烟者抵御尼古丁渴求。

临床背景

大量文献表明，基底神经节（BG）、海马（Hippo）和岛回（Ins）在尼古丁成瘾中发挥着重要作用（Ersche 等，2011；Gaznick 等，2014年；McClernon 等，2016）。因此，我们打算考察这三个双侧感兴趣区之间所有可能的成对连接模式（共12种）：（BG，Hippo）、（BG，Ins）和（Hippo，Ins）。为简洁起见，我们将在此展示（左侧BG，左侧Ins）这一对的结果，其余11种情况见附录D。同样，我们使用Brainnetome图谱标注左侧BG和左侧Ins（左侧BG：2345个体素；左侧Ins：1762个体素）。我们采用与数据集1相同的计算流程，获得了各临床组的逐边连接推断矩阵$\mathbf{W}^{(\mathrm{BG}_{\mathrm{left}},\mathrm{Ins}_{\mathrm{left}})\ 2345\times1762}$。随后，我们将SCCN和基于MDL的检验应用于$\mathbf{W}^{(\mathrm{BG}_{\mathrm{left}},\mathrm{Ins}_{\mathrm{left}})}$ </style}，在空间连续性约束下提取异常子区域对。最后，我们将所得结果与使用两种常见现有方法获得的结果进行比较。

网络层面结果

推断矩阵$\mathbf{W}^{(\mathrm{BG}_{\mathrm{left}},\mathrm{Ins}_{\mathrm{left}})}$中的每个条目都带有一个 $-\log\ p$值，用于检验临床组之间的体素功能连接差异。（图5(1)）。执行算法2 返回了最大似然估计 $\lambda = 0.75$。给定估计的 λ后，算法1 返回了$\mathbf{W}^{(\mathrm{BG}_{\mathrm{left}},\mathrm{Ins}_{\mathrm{left}})}$的簇数量$\hat{C} = 306, \hat{D} = 210$。基于MDL的检验返回了六个异常

子区域对，在图5(2)中以红色标记。来自$\mathbf{W}^{(\mathrm{BG}_{\mathrm{left}},\mathrm{Ins}_{\mathrm{left}})}$的检测到的子区域对的三维展示如图5(a)–(e)所示（为基于MLD的置换检验选择的显著性水平为0.05）。所有提取的子区域对都呈现出组织良好的拓扑结构。来自$\mathbf{W}^{(\mathrm{BG}_{\mathrm{left}},\mathrm{Ins}_{\mathrm{left}})}$的大多数异常vFC模式集中在左侧基底神经节的内侧下部与左侧岛叶之间。

检测到的子区域的生物学解释

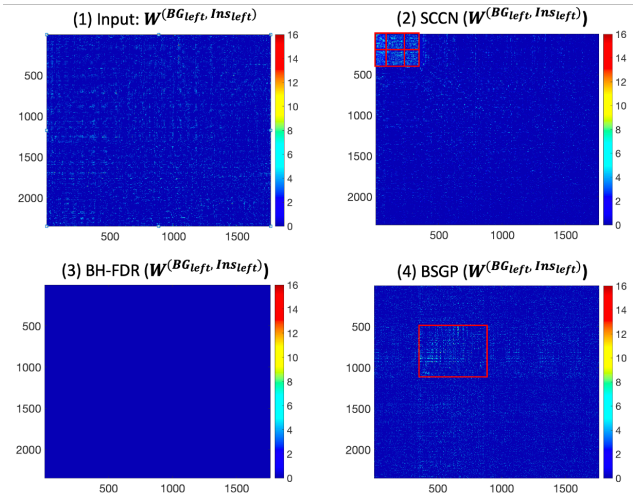
检测到的子区域包括若干被认为与尼古丁成瘾频繁相关的位置，包括基底神经节的内侧下部和岛叶后部。我们还观察到当前吸烟者这些区域内的连接减弱，这与此前的医学发现一致：静息态功能连接减弱与尼古丁成瘾严重程度增加相关（Fedota andStein，2015；Sutherland and Stein，2018）。纳入空间连续性约束有助于揭示基底神经节、海马和岛叶内的子区域，这些子区域最大限度地覆盖了与成瘾相关的体素功能连接。这些新发现提高了三个脑区中成瘾相关位置的空间特异性，并可能为今后抵制使用尼古丁产品的冲动提供指导。

与现有方法的比较

为了进行比较，我们再次对 $\mathbf{W}^{(\mathrm{BG}_{\mathrm{left}},\text{岛叶}_{\mathrm{left}})}$进行了逐边 BH-FDR 校正，并进行了逐簇 BSGP 校正。首先，在当前吸烟者组和既往轻度吸烟者组之间进行初步的逐边显著性检验，结果发现仅有 7.29% 的边达到显著（$p < 0.005$）。然而，应用 BH-FDR 校正后，没有边达到显著性，校正阈值为 $q = 0.01$ (图5(3))。将 BSGP 应用于 $\mathbf{W}^{(\mathrm{BG}_{\mathrm{left}},\text{岛叶}_{\mathrm{left}})}$时，仅检测到一个异常子区域对（图5(4)），该对中包含了 *49*.5% 的 $p > 0.005$边，而 SCCN 得出的比例为 3.12%。与这两种现有方法相比，SCCN 检出的空间连续子区域对包含了更多密集改变的体素功能连接，并具有更强的拓扑结构。

4 模拟

在模拟研究中，我们探究了与常见的现有方法相比，SCCN 能否以更优的性能提取密集改变的子区域对。具体而言，我们从两个角度评估其性能。(i) 多变量边级推断：提取的体素对是否具有较高的真正例率（TPR）和较低的假正例率（FPR）；(ii) 网络级推断：与其他未提取的子区域相比，提取的子区域是否包含最多的真正例体素。



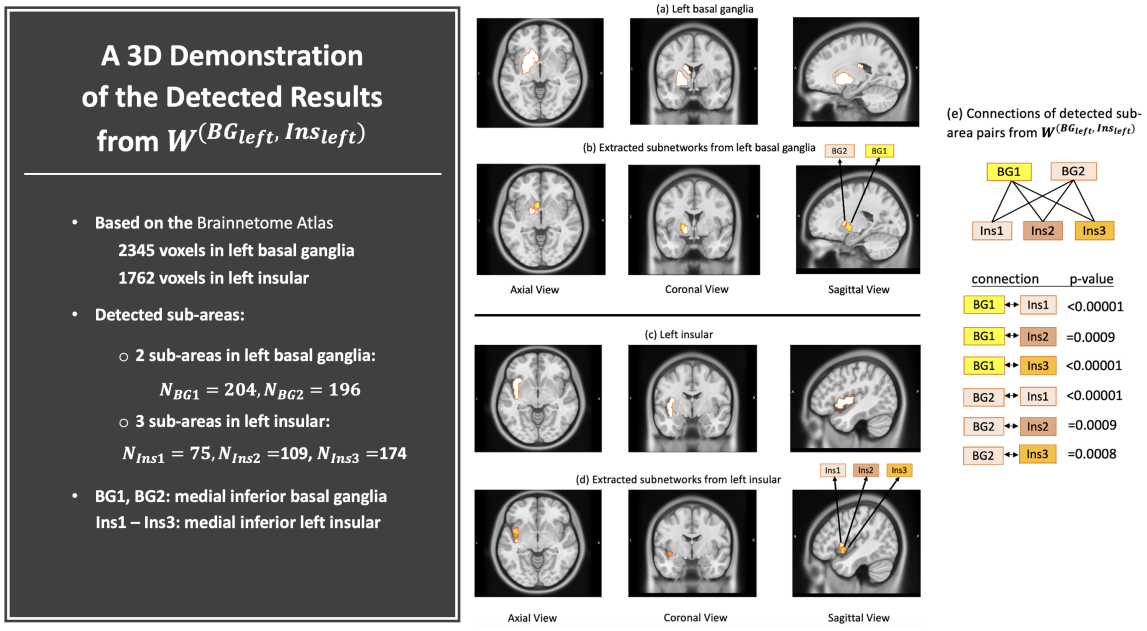


图5：*(1)* $\mathbf{W}^{(\mathrm{BG}_{\mathrm{left}}, \mathrm{Ins}_{\mathrm{left}})}$的热图：行和列分别对应左侧基底神经节和左侧岛叶中的体素。*(2) SCCN* 得出的结果：通过基于*MDL*的置换检验的变化子区域以红框标出。*(3) BH-FDR* 得出的结果：假设检验误差度量的截断值设为 $q = 0.05$ 。未检测到子区域对。*(4) BSGP* 得出的结果：仅检测到一个阳性但密集程度低得多的子区域对。检测到的子区域对也缺乏空间连续性和特异性。*(a)-(d)* 展示了从 $\mathbf{W}^{(\mathrm{BG}_{\mathrm{left}}, \mathrm{Ins}_{\mathrm{left}})}$ 检测到的 *6* 个变化子区域的三维展示。*(a)-(e)* 展示了从 $\mathbf{W}^{(\mathrm{BG}_{\mathrm{left}}, \mathrm{Ins}_{\mathrm{left}})}$ 检测到结果的三维展示。根据 *(e)* 中所示的基于*MDL*的置换检验的 *p*值，大多数阳性子区域对位于左侧基底神经节和左侧岛叶的内侧下部。

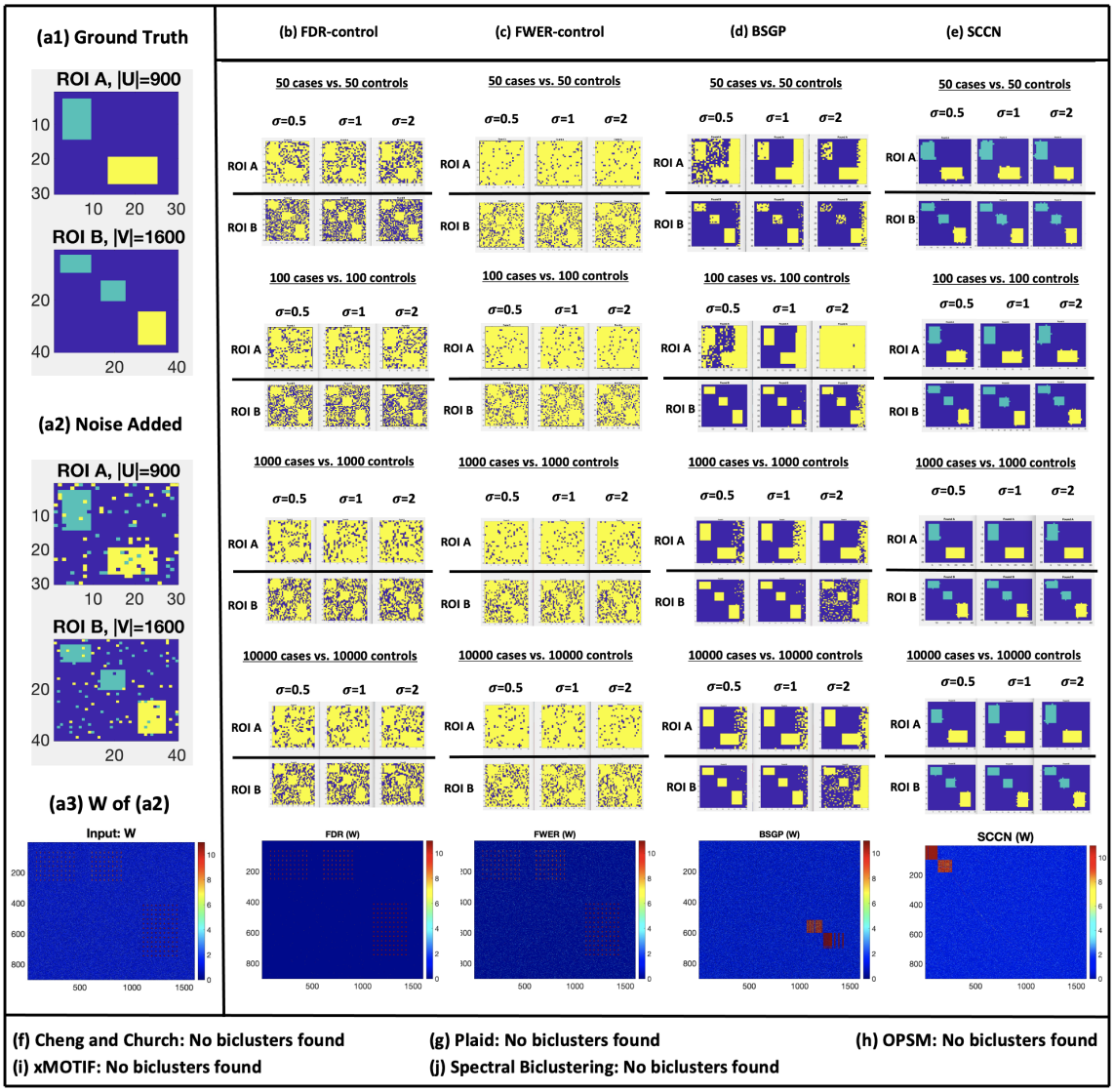


图 6： 模拟结果的可视化。*(a1)*显示了具有空间连续性的疾病相关子区域对（U_1, V_1）、（U_1, V_2）和（U_2, V_3）的真实位置（只有颜色相同的区域才表现出从 A 到 B 的连接异常）。*(a2)*在*(a1)*的基础上加入了额外噪声和彼此孤立但为正的体素对，以模拟脑连接组中的真实 *vFC*模式。*(a3)*显示了基于*(a2)*获得的连接推断矩阵$\mathbf{W}$ 。*(b)–(e)*显示了在不同方差 σ 和样本量 S下检测到的疾病相关体素对（同样，只有颜色相同的区域构成一对）。最后一行显示了*(a3)*的同构图，其中提取出的子区域对被置于顶部。我们突出显示了由错误发现率控制和家族错误率控制产生的超过阈值的体素对中的体素，以及由 *BSGP* 和 *SCCN* 提取的子区域对中的体素。采用错误发现率控制和家族错误率控制进行多重检验往往会提取过多体素，并导致较高的假阳性错误率。*BSGP*能够更好地控制错误率，但它提取体素对时无法区分正确的区域级连接，即（U_1, V_1）、（U_1, V_2）和（U_2, V_3）。相比之下，*SCCN*能够同时恢复 A 和 B 中空间连续的子区域，并揭示正确的疾病相关 *vFC*模式。*(f)–(j)*显示，所列双聚类算法均未提取出任何单个差异表达的子区域对。

4.1 主要分析

我们首先构建了一个二部图 $G = \{U, V\}$，用于表示 S 名受试者的两个脑区 A 和 B 之间的脑连接组（图 6(a1)），其中 U 对应于区域 A 中的体素集，V 对应于区域 B 中的体素集。我们假设所有S 名受试者在空间标准化和配准后共享相同的节点集，即 $(U^s, V^s) \equiv (U, V)\ \forall s \in [S]$。接下来，我们模拟感兴趣的协变量 $\{\mathrm{X}^1、\ldots\ldots、\mathrm{X}^S\}$，其中包含所有 S 名受试者的临床信息。最后，我们模拟了所有受试者在区域 A 和 B 之间的 Fisher z-变换连接矩阵 $\{\mathrm{Z}^1、\ldots\ldots、\mathrm{Z}^S\}$，其中 $\mathrm{Z}^s \in \mathbb{R}^{n\times m}, n = |U|, m = |V|$。具体而言，将 Z^s 中的每个元素 z_{ij}^s 设定为服从 $\mathcal{N}(h(z_{ij}^s)\ \sigma^2)$，其中 $h(z_{ij}^s) = \mathrm{X}^s\beta_{ij}$ 在区域 A 和 B 内具有位置特异性。

在 the 接下来，我们展示上述模拟框架下的数值设置。 k:

1. 对于两个预先定义的感兴趣脑区，我们在区域 A 中模拟了 $|U| = 900$ 个体素，在B区域中模拟了 $|V| = 1600$ 个体素。在 $|U|$ 和 $|V|$ 内，我们还随机模拟了三个疾病相关子区域对（U_1，V_1）、（U_1，V_2）和（U_2，V_3）。并非每个可能的对子 $\{(U_c, V_d)\ \ c = [2],\ d = [3]\}$ 都与疾病相关；只有颜色相同的区域才表现出从 A 到 B 的连接异常（见图 6(a1)）。这些子区域对的大小分别为 $|U_1||V_1| = 84 \times 70 = 5880$、$|U_1||V_2| = 84 \times 64 =$5376 和 $|U_2||V_3| = 96 \times 117 = 11\,232$。此外，我们还在区域 A 和 B 中加入了空间上孤立的异常体素以及噪声，以模拟更真实的神经连接（图 6(a2) ）。

2. 对于经 Fisher z变换的连接矩阵 $\{\mathbf{Z}^s, s \in S\}$，我们设定 $h(z_{ij}^s) = \beta_0 + \beta_{ij,1}x_1^s + \beta_{ij,2}x_2^s + \beta_{ij,3}x_3^s$，其中 x_1^s 和 x_2^s 存储受试者 s的年龄和性别信息，而 x_3^s 表示其临床状态（若患者 s 患有精神障碍，则为 x_3^s=1 ；健康对照则为 0）。此外，虽然 $\beta_{ij,1}$ 和 $\beta_{ij,2}$ 通常不随空间变化，但 $\beta_{ij,3}$ 被认为具有脑区特异性：

$$\beta_{ij,3} = \begin{cases} 0.9, & \text{if} \quad (i,j) \in (U_1, V_1) \bigcup (U_1, V_2), \\ 0.13, & \text{if} \quad (i,j) \in (U_2, V_3), \\ 0, & \text{if} \quad (i,j) \in U/\{(U_1, V_1) \bigcup (U_1, V_2) \bigcup (U_2, V_3)\}. \end{cases}$$

3. 为控制标准化效应量，我们在 $\mathbf{Z}^s \sim \mathcal{N}(h(z_{ij}^s)\ \sigma^2)$ 中设置 $\sigma^2 = 0.5, 1.0, 2.0$。此外，采用了四种样本量：$S = 100$、200、2000年和 20,000，且每种样本量下健康对照者与患者人数均衡。针对不同的 (σ, S) 设置，分别进行了 1000 次模拟，以评估真正例率和假正例率的变异性。

我们实施了 SCCN 的算法1 和 2 ，从每个模拟数据集中识别子区域对；随后应用算法3，对检测到的子区域对进行簇级推断。为评估多变量边级推断的性能，我们考虑了两种常规的多重检验控制方法（错误发现率和家族错误率）。具体而言，我们使用体素级置换检验（进行 1000 次置换）控制家族错误率，并使用 Benjamini–Hochberg程序（以 $q = 0.05$ 作为截断值）控制错误发现率（Benjamini和Hochberg，1995年）。为评估簇级

性能，我们旨在将真实的疾病相关子图 $\{(U_c,\ V_d)\}$ 与五种常用双聚类算法（即 Cheng and Church、Plaid、OPSM、xMOTIF 和谱双聚类（J. K.Gupta，2013年））生成的估计子图 $\{(U_c,\ V_d)\}$ 进行比较。

边级推断结果见表1，结果的图示见图6。对于所有不同（$\sigma,\ S$）下的边级推断性能，与两种传统多重检验校正方法（即 FDR 和 FWER 控制）相比，SCCN 的 TPR 更高，而其控制 FPR 的能力介于两者之间。SCCN 在控制 FPR 方面的表现相对较差（与灵敏度相比），有时会受到以下缺点的影响：在采用统一阈值的传统多重检验方法中，一项假阳性发现恰好对应一条假阳性边。然而，SCCN 通过对每个感兴趣区域内的体素进行分区来检测改变的边；因此，SCCN 的一项假阳性发现对应一个假阳性体素，例如 $v_i \in U_c$，这会产生 n 条假阳性边，当 V_d （$|V_d| = n$）被发现与 $|U_c|$相连时。 V_d的规模越大，产生的假阳性边就越多。尽管检测一个假阳性体素会受到如此严重的惩罚，SCCN 仍能控制 FPR，并且在综合考量 TPR 和 FPR 时表现更佳。更重要的是，传统 FDR 和 FWER 校正方法发现的假阳性边几乎覆盖了所有 ROI 内的体素，导致在识别协变量相关的体素功能连接模式时，空间特异性显著下降。

关于网络级推断性能，除 BSGP 外，所有常见的双聚类方法均未能检测到任何阳性双聚类（差异表达的子区域对）。然而，BSGP 仍未能确保空间连续性，也未能正确揭示所提取子区域之间的确切连接关系。也就是说，与 SCCN（图 6（e））产生的结果不同，BSGP（图 6（d））无法有效区分黄色簇和蓝色簇。相比之下，SCCN 在检测社区结构并纳入空间连续性方面展现出卓越的网络级性能。

4.2 阴性对照分析

我们进一步进行了阴性对照分析，以评估我们方法的假正例率。我们考虑这样一种情形：预先选定的感兴趣区对之间的连接与关注的临床状况无关。我们在区域 A 和 B 中生成了 $|U| = 900$ 和 $|V| = 1600$ 个体素。我们将患者组和对照组分别标记为 1 和 0，但由于组间不存在异常子区域对 $\{(U_c,\ V_d)\}$，因此我们仅为所有 S 名受试者设置了覆盖整个区域的连接矩阵 $\mathbf{Z}^s \sim \mathcal{N}(0,\ \sigma^2)$。基于 $\mathbf{Z}^s$，我们获得了跨临床组的推断矩阵 $\mathbf{W^0}$ 。由于网络检测已验证可扩展至不同的样本量和样本方差，我们评估了配置（$S = 1000,\ \sigma = 1$）作为概念验证。最后，我们在 $\mathbf{W^0}$ 上实施了 SCCN。由于假阳性体素对倾向于随机分布，因此没有子区域对达到显著性。因此，子区域级别的假阳性发现为 0。1000 次迭代中的边级假阳性率（超过阈值的体素对）为 6.82×10^{-5}（标准差 1.29×10^{-5}），这与预先设定的 α 水平（$\mathbb{E}(p) = 0.00005$）一致。我们已提供

Sample Size: 50 cases vs. 50 controls				
Methods	σ	**0.5**	**1**	**2**
FDR-control	TPR	0.918 (0.229)	0.832 (0.314)	0.637 (0. 328)
	FPR	0.060 (0.007)	0.065 (0.053)	0.073 (0.158)
	Network Detection	No	No	No
FWER-control	TPR	0.994 (0.169)	0.893 (0.177)	0.842 (0. 179)
	FPR	0.762 (0.142)	0.764 (0.152)	0.771 (0.193)
	Network Detection	No	No	No
SCCN	TPR	1(0)	1(0)	1(0)
	FPR	0.087 (0.011)	0.093 (0.012)	0.096 (0.014)
	Network Detection	Yes	Yes	Yes
BSGP	TPR	0.946 (0.106)	0.893 (0.122)	0.667 (0.278)
	FPR	0.858 (0.258)	0.773 (0.173)	0.845 (0.439)
	Network Detection	Yes	Yes	Yes
Cheng and Church Plaid OPSM xMOTIF Spectral Biclustering	TPR	0(0)	0(0)	0(0)
	FPR	0(0)	0(0)	0(0)
	Network Detection	No biclusters detected		

Sample Size: 100 cases vs. 100 controls				
Methods	σ	**0.5**	**1**	**2**
FDR-control	TPR	0.958 (0.215)	0.892 (0.249)	0.746 (0. 427)
	FPR	0.065 (0.069)	0.069 (0.048)	0.070 (0.164)
	Network Detection	No	No	No
FWER-control	TPR	0.994 (0.157)	0.989 (0.182)	0.842 (0. 163)
	FPR	0.767 (0.139)	0.762 (0.140)	0.761 (0.187)
	Network Detection	No	No	No
SCCN	TPR	1(0)	1(0)	1(0)
	FPR	0.081 (0.010)	0.082 (0.012)	0.097 (0.013)
	Network Detection	Yes	Yes	Yes
BSGP	TPR	1 (0.105)	1 (0.120)	0.993 (0.262)
	FPR	0.844 (0.228)	0.764 (0.172)	0.940 (0.369)
	Network Detection	Yes	Yes	Yes
Cheng and Church Plaid OPSM xMOTIF Spectral Biclustering	TPR	0(0)	0(0)	0(0)
	FPR	0(0)	0(0)	0(0)
	Network Detection	No biclusters detected		

Sample Size: 1000 cases vs. 1000 controls				
Methods	σ	**0.5**	**1**	**2**
FDR-control	TPR	0.957 (0.209)	0.878 (0.287)	0.789 (0. 363)
	FPR	0.061 (0.007)	0.061 (0.058)	0.071 (0.159)
	Network Detection	No	No	No
FWER-control	TPR	0.993 (0.169)	0.904 (0.180)	0.823 (0. 187)
	FPR	0.749 (0.140)	0.727 (0.145)	0.754 (0.176)
	Network Detection	No	No	No
SCCN	TPR	1(0)	1(0)	1(0)
	FPR	0.081 (0.008)	0.087 (0.010)	0.081 (0.012)
	Network Detection	Yes	Yes	Yes
BSGP	TPR	0.968 (0.106)	0.903 (0.122)	0.875 (0.278)
	FPR	0.374 (0.438)	0.405 (0.360)	0.737 (0.447)
	Network Detection	Yes	Yes	Yes
Cheng and Church Plaid OPSM xMOTIF Spectral Biclustering	TPR	0(0)	0(0)	0(0)
	FPR	0(0)	0(0)	0(0)
	Network Detection	No biclusters detected		

Sample Size: 10000 cases vs. 10000 controls				
Methods	σ	**0.5**	**1**	**2**
FDR-control	TPR	0.959 (0.229)	0.873 (0.357)	0.771 (0. 391)
	FPR	0.060 (0.007)	0.061 (0.061)	0.069 (0.160)
	Network Detection	No	No	No
FWER-control	TPR	0.992 (0.174)	0.907 (0.176)	0.819 (0. 183)
	FPR	0.753 (0.145)	0.734 (0.145)	0.759 (0.157)
	Network Detection	No	No	No
SCCN	TPR	1(0)	1(0)	1(0)
	FPR	0.081 (0.007)	0.091 (0.007)	0.087 (0.011)
	Network Detection	Yes	Yes	Yes
BSGP	TPR	0.972 (0.208)	0.884 (0.122)	0.873 (0.290)
	FPR	0.574 (0.426)	0.368 (0.499)	0.679 (0.407)
	Network Detection	Yes	Yes	Yes
Cheng and Church Plaid OPSM xMOTIF Spectral Biclustering	TPR	0(0)	0(0)	0(0)
	FPR	0(0)	0(0)	0(0)
	Network Detection	No biclusters detected		

表*1*：模拟结果。四个子表展示了不同样本量和方差下的推断结果，其中 *TPR* 和 *FPR* 分别对应逐边真正例率和逐边假正例率。网络检测结果表明，该算法能否成功提取疾病相关子区域对之间正确的连接模式。

这些结果的图可视化见附录 E。

总之，我们已证明，子区域检测不受方差 σ^2、样本量 S 取值差异或其他噪声来源的影响。SCCN 还能够产生灵敏度高、FPR 较低的 vFC模式。空间连续性约束使正边能够在数据驱动的子区域内相互借力，因此灵敏度显著提高。具有这些约束的数据驱动子区域还能够排除连接感兴趣区域中随机散布体素的假阳性边。因此，假阳性发现得到了大幅抑制。此外，灵敏度（以及统计功效）的共同提升和对 FPR 的控制，使所有模拟数据集中的体素集合几乎完全相同。因此，可重复性显著提高。

5 讨论

精神疾病和神经系统疾病通常与脑连接组受损有关。为了提高检测受疾病影响的脑连接组时的空间特异性和灵敏度，本研究聚焦于体素级连接网络分析。我们开发了统计模型，重点从感兴趣的区域对中提取异常体素对，并可进一步扩展至全脑连接组分析。我们尝试同时解决多体素对检验中控制假正例率以及体素功能连接分析中的空间连续性约束这两项挑战。事实上，疾病相关的体素级连接也可能发生在一个区域内部。我们可以将类似方法应用于预先获得的区域内邻接矩阵 $\mathbf{W}_{n\times n}$，方法是整合相应的空间连续性基础结构图，例如 $\mathcal{S}_A$。如此一来，SCCN 便适用于区域内体素级连接分析。相关详细步骤见附录 A.3。此外，提取子区域的脑区划分通常基于常用脑图谱（例如布罗德曼图或国际脑图谱联盟图谱），这些图谱建立在经过全面研究的皮层解剖结构之上，例如复杂的脑回和脑沟褶皱模式。被脑回和脑沟分隔的不同区域往往呈现不同的神经生物学结构和功能，因此这些图谱可作为研究子区域社区结构的良好基础。不过，为了进一步克服使用现有脑区划分的局限，可以考虑合并来自一对空间相邻区域的任意已提取空间相邻子区域，前提是这种合并在统计上具有一致性，在生物学上也有意义。

我们提出的方法的核心是识别包含异常高密度表型相关体素对的子区域对。借助这种高密度特征，我们可以通过排除孤立的假阳性边，有效控制假正例率，从而大幅减少假阳性节点的数量。因此，我们提高了在体素层面提取的疾病相关模式的空间特异性。在此，我们提出了一种新的非参数目标函数来实现这一目标，并通过高效算法加以实现。我们还开发了推断方法，用于评估所提取的每个子区域对的统计显著性。

我们数据示例中的生物学发现颇具新意；SCCN 揭示了精神分裂症在广为人知的显著性网络中的体素功能连接组模式。我们发现，显著性网络连接功能异常主要由背侧岛叶与前扣带皮层之间的连接受损所致，而非综合性的区域层面发现所显示的结果。我们进一步通过大量模拟验证了这些发现，并表明我们的方法能够在控制假正例率的同时提高灵敏度，并保持空间连续性。

总之，SCCN 提供了一套新的体素功能连接（vFC）分析工具，在保持假阳性错误率得到良好控制的同时，提高了空间分辨率和特异性。因此，SCCN 的发现有望转化为针对脑部疾病的更有效治疗方案。由于 SCCN 的输入数据是体素对层面的推断结果，因此它适用于所有能够进行有效统计推断的连接测量方法和数据模态（例如白质纤维束追踪）。SCCN 还可能为全脑连接组的体素对网络分析提供一种有前景的策略。所有示例代码均可在此处找到：https://github.com/TongLu-bit/DecodingNetwork。

参考文献

Ardila, A., Bernal, B., and Rosselli, M. (2016). 语言脑区的定位有多精确？关于布罗德曼分区参与口语的综述。临床神经心理学档案, 31(1):112–122. Arnold, S. and Trojanowski, J. (1996). 精神分裂症神经病理学界定方面的最新进展。神经病理学学报, 92(3):217–231. Bahrami, M., Laurienti, P. J., and Simpson, S. L. (2019). 在全脑网络背景下分析脑子网络。*Human brain mapping*, 40(17):5123– 5141. Bellec, P., Perlbarg, V., Jbabdi, S., P´el´egrini-Issac, M., Anton, J.-L., Doyon, J., and Benali, H. (2006). 使用功能性磁共振成像识别脑中的大规模网络。*Neuroimage*, 29(4):1231–1243. Benjamini, Y. and Hochberg, Y. (1995). 控制假发现率：一种实用且有效的多重检验方法。英国皇家统计学会杂志：*B*辑（方法论）, 57(1):289–300. Bolla, M. (2013). 谱聚类与双聚类：学习大型图和列联表. John Wiley & Sons.¨Boltzmann, L. (1866). 关于热理论第二定律的力学意义¨（于*1866*年*2*月*8*日会议上提交）. 国立印刷局。Bradley, P. S. and Fayyad, U. M. (1998). 优化k均值聚类的初始点。发表于*ICML*, 第98卷, 第91–99页。Citeseer.

Bullmore, E. 和 Sporns, O.（2009）。复杂脑网络：结构和功能系统的图论分析。*Nature Reviews Neuroscience*，10(3)：186–198。Cano, C.、Adarve, L.、L´opez, J. 和 Blanco, A.（2007）。用于微阵列数据双聚类的可能性方法。*Computers in Biology and Medicine*，37(10)：1426–1436。Cao, M.、Wang, J.-H.、Dai, Z.-J.、Cao, X.-Y.、Jiang, L.-L.、Fan, F.-M.、Song, X.-W.、Xia, M.-R.、Shu, N.、Dong, Q. 等（2014）。贯穿整个生命周期的人脑功能连接组的拓扑组织。*Developmental Cognitive Neuroscience*，7：76–93。Cetin, M. S.、Christensen, F.、Abbott, C. C.、Stephen, J. M.、Mayer, A. R.、Ca˜nive, J. M.、Bustillo, J. R.、Pearlson, G. D. 和 Calhoun, V. D.（2014）。在精神分裂症患者中，丘脑和颞叶后部在静息状态及不同感觉范式下表现出更强的网络间连接。*NeuroImage*，97：117–126。Chen, S.、Bowman, F. D. 和 Mayberg, H. S.（2016）。用于神经影像数据脑连接建模的贝叶斯层级框架。*Biometrics*，72(2)：596–605。Chen, S.、Bowman, F. D. 和 Xing, Y.（2020）。利用 k 部网络拓扑检测和检验发生改变的脑连接网络。*Computational Statistics & Data Analysis*，141：109–122。Chen, S.、Wu, Q. 和 Hong, L. E.（2019）。基于图组合学的组水平网络推断。*bioRxiv*，第758490页。Chen, S.、Zhang, Y.、Wu, Q.、Bi, C.、Kochunov, P. 和 Hong, L. E.（2023）。识别用于全脑连接组分析的协变量相关子网络。*Biostatistics*，第kxad007页。Cheng, H.、Hua, K. A. 和 Vu, K.（2008）。约束局部加权聚类。*VLDB Endowment* 会议录，1(1)：90–101。Craddock, R. C.、James, G. A.、Holtzheimer III, P. E.、Hu, X. P. 和 Mayberg, H. S.（2012）。通过空间约束谱聚类生成的全脑功能性磁共振成像图谱。*Human Brain Mapping*，33(8)：1914–1928。Derado, G.、Bowman, F. D. 和 Kilts, C. D.（2010）。功能性磁共振成像数据中时空依赖的建模。*Biometrics*，66(3)：949–957。Efron, B.（2008）。微阵列、经验贝叶斯与双组模型。*Statistical Science*，23(1)：1–22。Efron, B.（2012）。大规模推断：用于估计、检验和预测的经验贝叶斯方法，第1卷。剑桥大学出版社。Eickhoff, S. B.、Thirion, B.、Varoquaux, G. 和 Bzdok, D.（2015）。基于连接的脑区划分：批判与启示。*Human Brain Mapping*，36(12)：4771–4792。

Ersche, K. D., Barnes, A., Jones, P. S., Morein-Zamir, S., Robbins, T. W. 和 Bull more, E. T.（2011）。额纹状体脑系统的结构异常与可卡因依赖者的某些冲动性和强迫性表现相关。《*Brain*》, 134(7):2013– 2024. Fan, J., Han, X. 和 Gu, W.（2012）。在任意协方差依赖下估计错误发现比例。《美国统计学会会刊》, 107(499):1019– 1035. Fan, J. 和 Lv, J.（2008）。针对超高维特征空间的确定性独立筛选。《英国皇家统计学会会刊：*B*辑（统计方法学）》, 70(5):849–911. Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A. R. 等（2016）。人类脑网络组图谱：基于连接架构的新脑图谱。《大脑皮层》, 26(8):3508–3526. Farahani, F. V., Karwowski, W. 和 Lighthall, N. R.（2019）。图论在识别人类脑网络连接模式中的应用：系统综述。《神经科学前沿》, 13:585. Fedota, J. R. 和 Stein, E. A.（2015）。静息态功能连接与尼古丁成瘾：生物标志物开发前景。《纽约科学院年鉴》, 1349(1):64–82. Feng, X. 和 He, X.（2014）。基于稳健低秩数据矩阵近似的统计推断。《统计年鉴》。, 42(1):190–210. Ferri, J., Ford, J., Roach, B., Turner, J., Van Erp, T., Voyvodic, J., Preda, A., Belger, A., Bustillo, J., O'Leary, D. 等（2018）。精神分裂症患者的静息态丘脑连接异常及其与症状的关系。《心理医学》, 48(15):2492– 2499. Fornito, A., Zalesky, A. 和 Bullmore, E.（2016）。脑网络分析基础。学术出版社。Gao, C., Ma, Z., Zhang, A. Y. 和 Zhou, H. H.（2017）。在随机区块模型中实现最优错误分类比例。《机器学习研究杂志》, 18(1):1980–2024. Gaznick, N., Tranel, D., McNutt, A. 和 Bechara, A.（2014）。基底神经节与岛叶同时受损，比单独基底神经节受损对吸烟成瘾造成更严重的破坏。《尼古丁与烟草研究》, 16(4):445–453. Govaert, G. 和 Nadif, M.（2008）。基于伯努利混合模型的区块聚类：不同方法的比较。*Computational Statistics & Data Analysis*, 52(6):3233–3245。

Gr¨unwald, P. D.（2007）。最小描述长度原理。MIT出版社。Hartigan, J. A.（1972）。数据矩阵的直接聚类。《美国统计学会会刊》, 67(337):123–129. J. K. Gupta, S. Singh, N. K. V.（2013）。Mtba：用于双聚类分析的 Matlab 工具箱。第94–97页。IEEE。Jiang, X., Zhang, T., Zhang, S., Kendrick, K. M. 和 Liu, T.（2021）。脑回与脑沟之间的基本功能差异：对脑功能、认知和行为的启示。*Psychoradiology*, 1(1):23–41. Kamvar, S., Klein, D. 和 Manning, C.（2003）。谱学习。技术报告2003-25，斯坦福InfoLab。Li, R., Zhong, W. 和 Zhu, L.（2012a）。通过距离相关学习进行特征筛选。《美国统计学会会刊》。, 107(499):1129–1139. Li, R., Zhong, W. 和 Zhu, L.（2012b）。通过距离相关学习进行特征筛选。《美国统计学会会刊》, 107(499):1129–1139. Loewe, K., Grueschow, M., Stoppel, C. M., Kruse, R. 和 Borgelt, C.（2014）。快速构建体素级功能连接图。*BMC neuroscience*, 15(1):1– 13. Luxburg, U., Bousquet, O. 和 Belkin, M.（2004）。谱聚类的局限。神经信息处理系统进展, 17. Lynall, M.-E., Bassett, D. S., Kerwin, R., McKenna, P. J., Kitzbichler, M., Muller, U. 和 Bullmore, E.（2010）。精神分裂症中的功能连接与脑网络。*Journal of Neuroscience*, 30(28):9477–9487. McClernon, F. J., Conklin, C. A., Kozink, R. V., Adcock, R. A., Sweitzer, M. M., Addicott, M. A., Chou, Y.-h, Chen, N.-k, Hallyburton, M. B. 和 DeVito, A. M.（2016）。海马和岛叶对与吸烟相关环境的反应：尼古丁依赖中药物—情境效应的神经影像学证据。*Neuropsychopharmacology*, 41(3):877– 885. Moran, L. V., Sampath, H., Kochunov, P. 和 Hong, L. E.（2013）。将精神分裂症与高吸烟风险联系起来的脑回路。*Schizophrenia bulletin*, 39(6):1373–1381. Nichols, T. E.（2012）。多重检验校正、非参数方法与随机场理论。***NeuroImage***, 62(2):811–815. Nichols, T. E. 和 Holmes, A. P.（2002）。功能神经影像学的非参数置换检验：含实例入门。*Human brain mapping*, 15(1):1–25。

Palaniyappan, L., White, T. P. 和 Liddle, P. F.（2012）。精神分裂症中显著性网络功能障碍的概念：从神经影像学观察到治疗机遇。《*Current topics in medicinal chemistry*》, 12(21):2324–2338。Rubinov, M. 和 Sporns, O.（2010）。脑连接的复杂网络测度：用途与解读。*Neuroimage*, 52(3):1059–1069。Shi, J. 和 Malik, J.（2000）。归一化割与图像分割。《*IEEE*模式分析与机器智能汇刊》, 22(8):888–905。Simpson, S. L., Bowman, F. D. 和 Laurienti, P. J.（2013）。复杂功能性脑网络分析：融合统计学与网络科学以理解大脑。《*Statistics Surveys*》, 7:1–36。St. Jacques, P. L., Dolcos, F. 和 Cabeza, R.（2009）。衰老对杏仁核功能连接及负性图片后续记忆的影响：一项功能性磁共振成像数据的网络分析。《*Psychological science*》, 20(1):74–84。Sutherland, M. T. 和 Stein, E. A.（2018）。烟草使用障碍的功能性神经环路与神经影像学生物标志物。《*Trends in molecular medicine*》, 24(2):129–143。Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P. 和 Poline, J.-B.（2006）。应对空间标准化的不足：功能性磁共振成像数据集的多被试脑区划分。*Human brain mapping*, 27(8):678–693。Wetherill, R. R., Jagannathan, K., Shin, J. 和 Franklin, T. R.（2014）。尼古丁依赖型吸烟者静息态神经网络的性别差异。*Addictive behav-iors*, 39(4):789–792。Wig, G. S., Laumann, T. O. 和 Petersen, S. E.（2014）。一种利用静息态相关性划分人类皮层区域的方法。*Neuroimage*, 93:276–291。Woo, C.-W., Krishnan, A. 和 Wager, T. D.（2014）。功能性磁共振成像分析中基于簇范围的阈值化：陷阱与建议。*Neuroimage*, 91:412–419。Wu, G.-R., Stramaglia, S., Chen, H., Liao, W. 和 Marinazzo, D.（2013）。绘制静息态功能性磁共振成像中的体素水平有效连接组。《*PloS one*》, 8(9):e73670。Wu, Q., Huang, X., Culbreth, A. J., Waltz, J. A., Hong, L. E. 和 Chen, S.（2021a）。通过自适应稠密子图发现提取脑疾病相关连接组子图。*Biometrics*。Wu, Q., Zhang, Y., Huang, X., Ma, T., Hong, L. E., Kochunov, P. 和 Chen, S.（2021b）。一种用于体素水平全基因组关联分析的多变量到多变量方法。*bioRxiv*。

Wylie, K. P. 和 Tregellas, J. R.（2010）。岛叶在精神分裂症中的作用。*Schizophre-nia research*, 123(2-3):93–104。Xia, Y. 和 Li, L.（2017）。矩阵图模型的假设检验及其在脑连接分析中的应用。*Biometrics*, 73(3):780–791。Xia, Y. 和 Li, L.（2019）。矩阵图假设检验及其在脑连接变化检测中的应用。*Statistica Sinica*, 29(1):303–328。Zalesky, A., Fornito, A. 和 Bullmore, E. T.（2010）。基于网络的统计方法：识别脑网络中的差异。*Neuroimage*, 53(4):1197–1207。