

高阶网络读书会

高阶脑网络拓扑特征提取及分析

报告人：阴桂梅

太原师范学院计算机学院副教授

更多“人工智能&复杂系统”的知识，请关注官网：pattern.swarma.org

脑网络特征分析流程

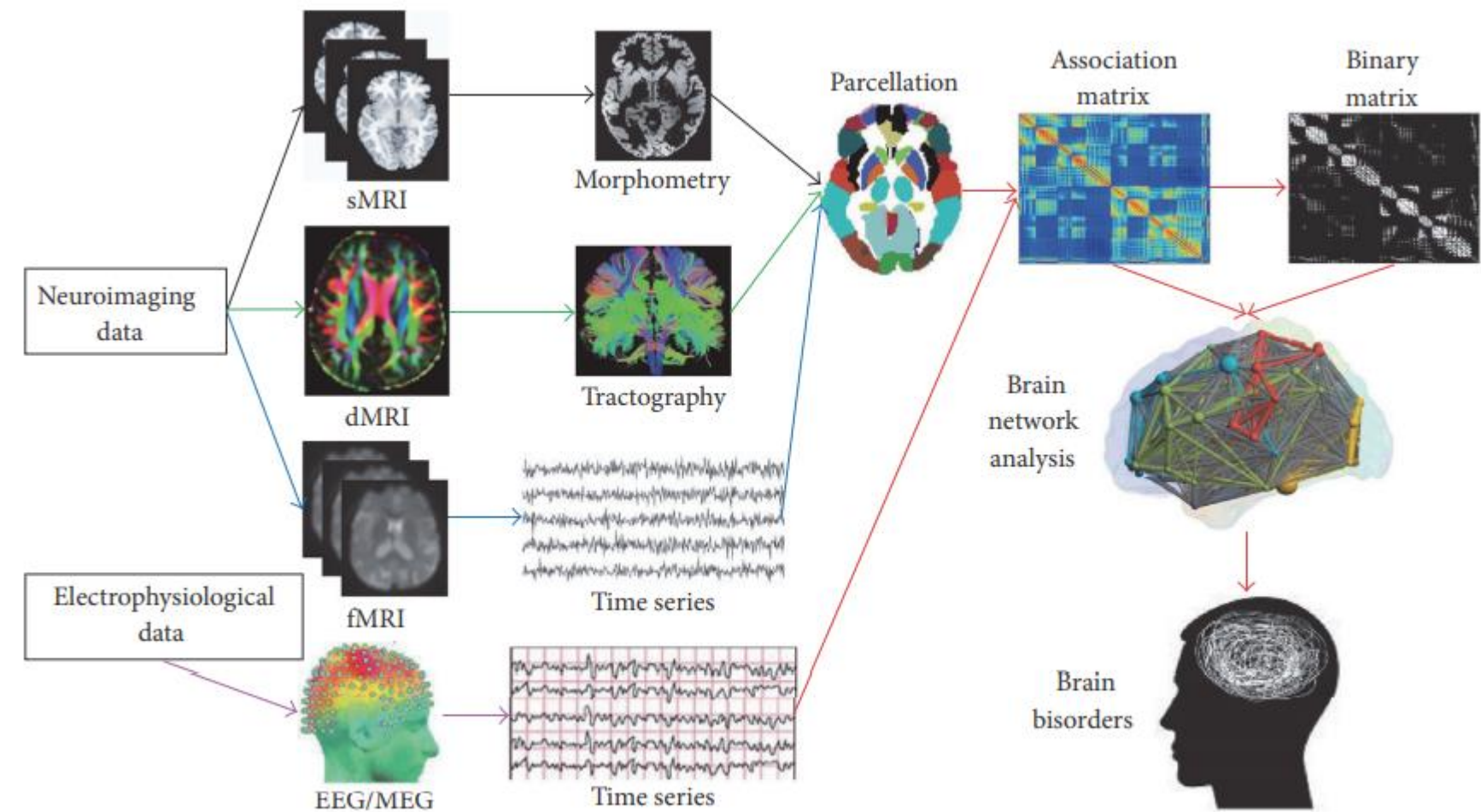
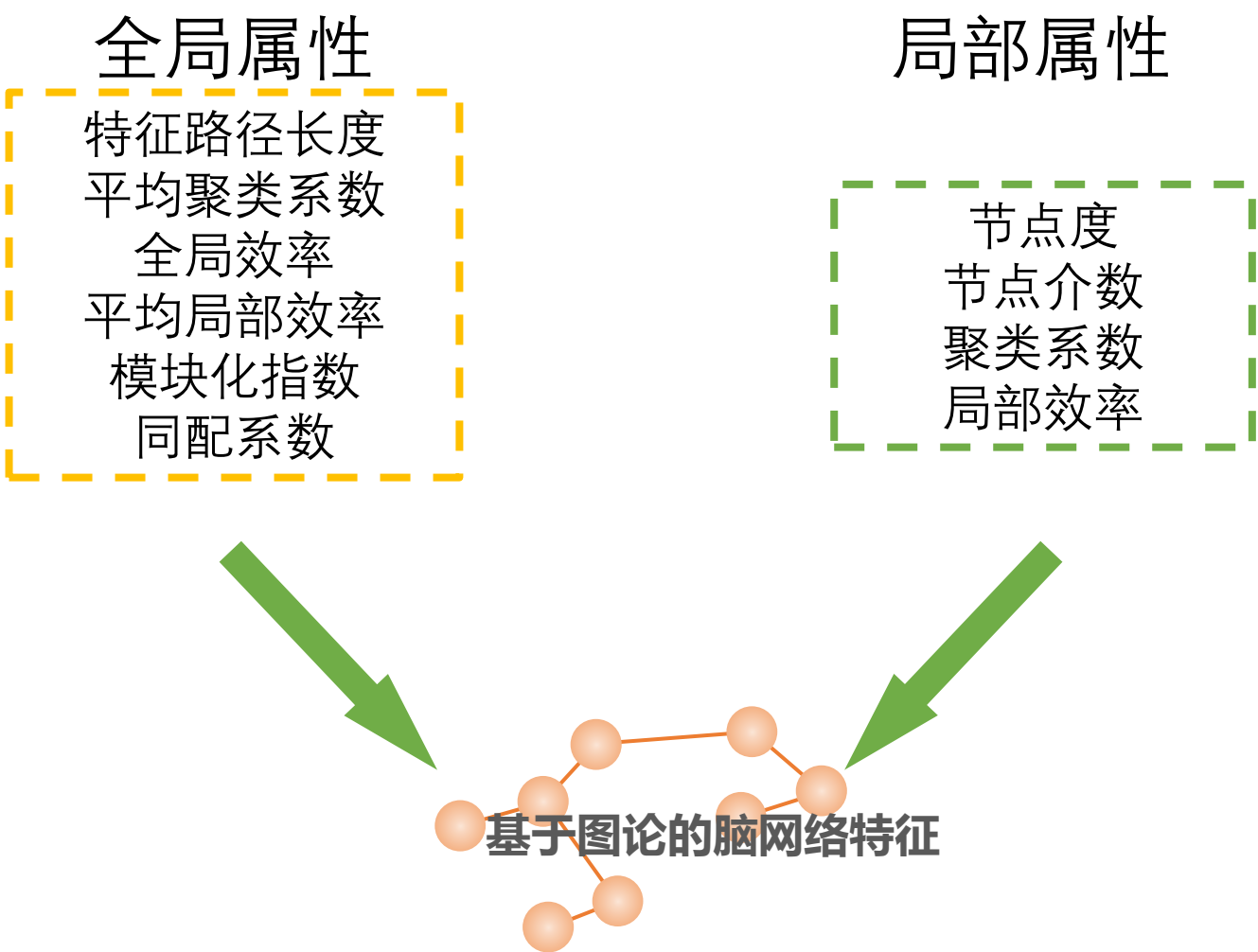


FIGURE 1: A general perspective for complex brain network analysis methods and its applications in brain disorders.



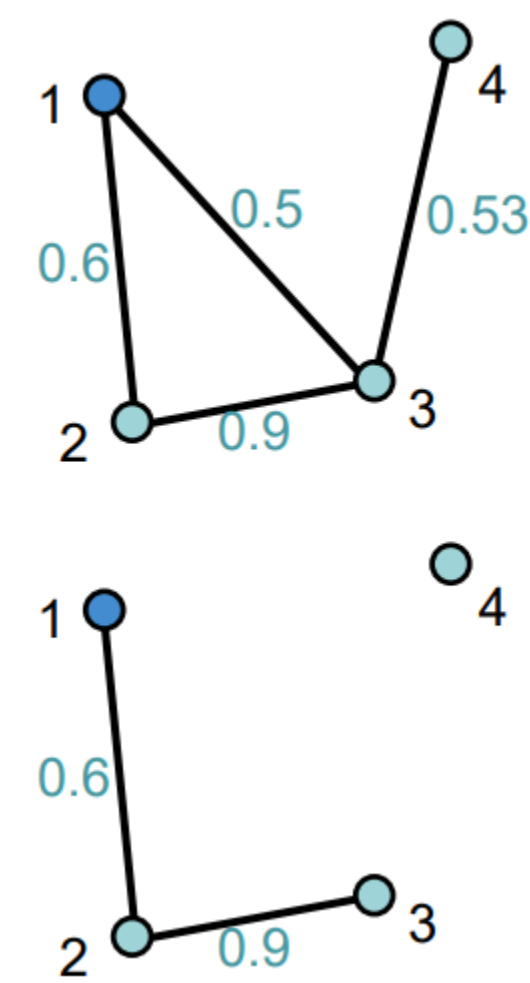
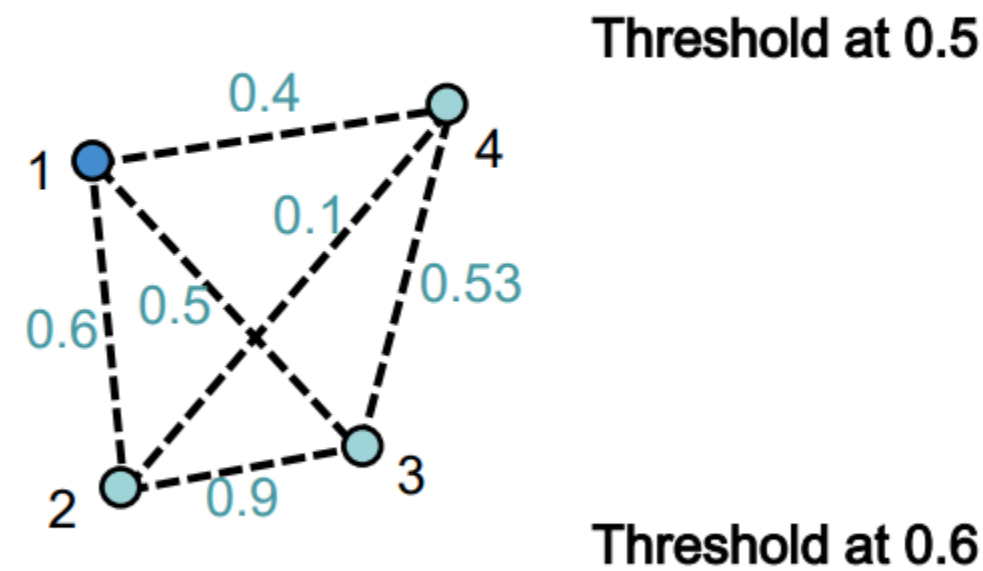
脑网络特征分析存在的问题

- 其构建的连接不保证绝对可靠，可根据自定义阈值选择连接，因此存在虚假连接
 - 由于其连接网络是两两关系的映射，因此不能解决大脑信息传递的高阶互动过程
- 阈值选择
 - (1) 选取单个阈值；预定义阈值空间；条件限制下的阈值空间，最终满足选取的阈值下所构建的脑网络具有小世界属性，如果构建随机网络则要求与原始网络具有相同的节点数和节点度分布，且节点的平均度 $>2\lg N$ 。
 - 存在如下问题，如：只适应于用随机网络，在实际网络中没有意义；数据变化时确定阈值过程复杂不具备普适性；在删除的弱连接中可能会包含重要的信息传递的路径等问题。
 - (2) 使用网络的最小生成树（MST）**构建无偏网络的方法**，该方法中虽然MST对阈值和密度值不敏感，可以认为是网络图形二值化的良好技术，但是由于网络的最小生成树非常稀疏，造成许多重要的本地连接被忽略；
 - (3) 一种基于**热核高斯核的无窗口**的方法，该方法可减少系统中虚假的快速变化大脑连接的状态空间，解决了滑动窗口法运用于动态脑网络分析时存在高频噪声问题。但还属于低阶网络分析。

基于阈值标准脑网络构建的缺点

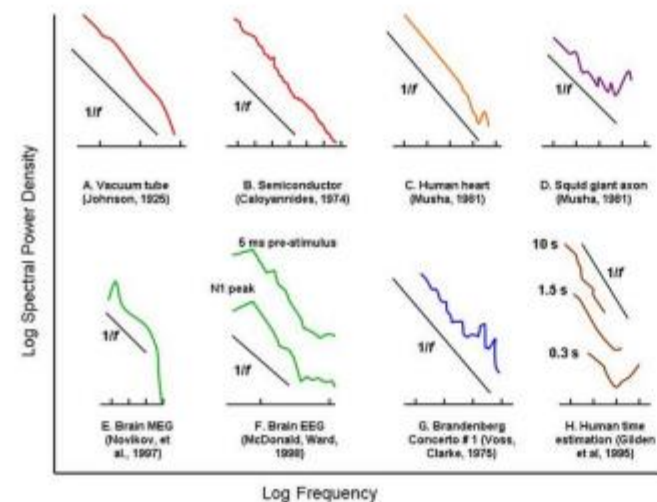
- **Shortcomings** of standard methods

- Thresholding issue

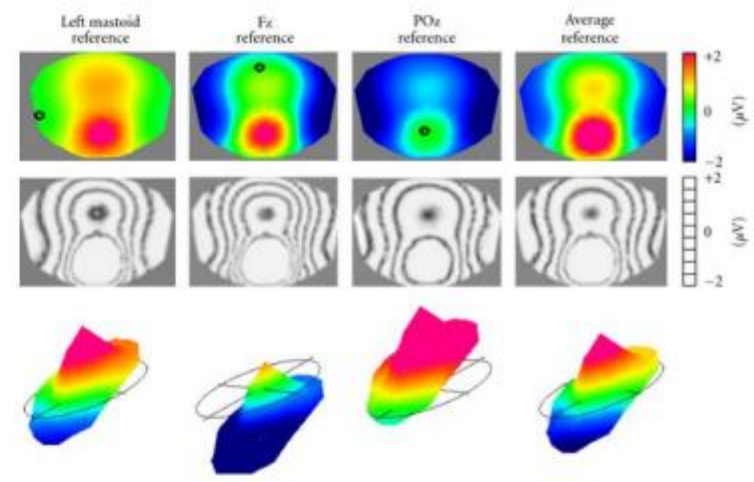


基于阈值标准脑网络构建的缺点

- **Shortcomings** of standard methods
 - Thresholding issue
 - Data perturbation across ...
1/f spectral distribution, subject by subject, condition to condition, ...



1/f noise – Scholarpedia,



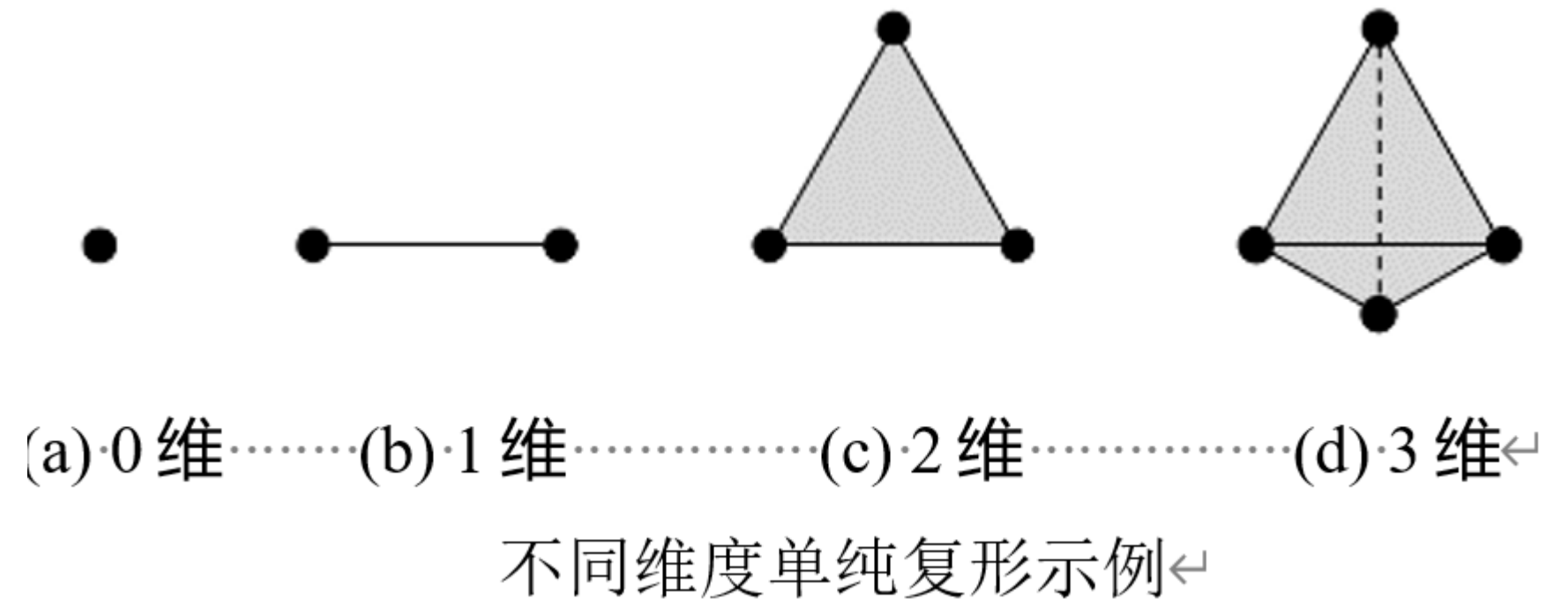
Brunet et al 2011

基于持续同调的脑网络构建

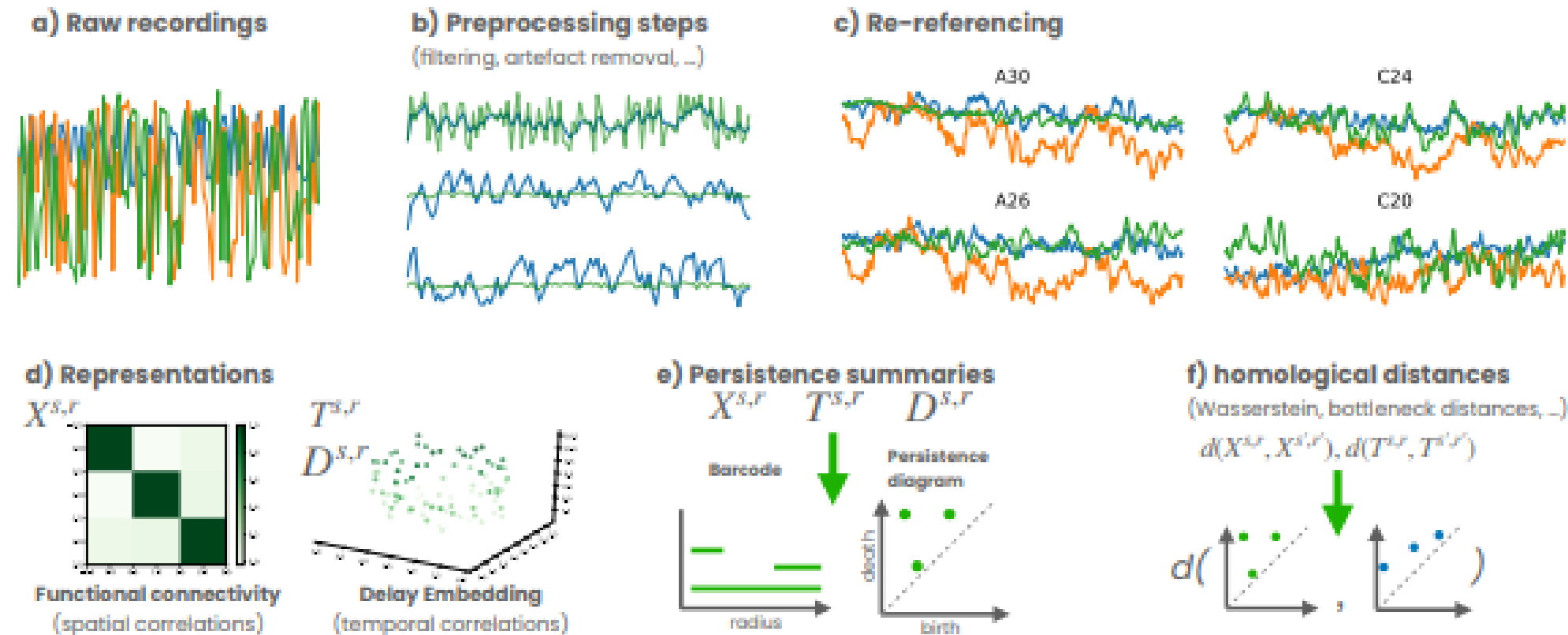
- 改进：将拓扑数据分析中的持续同调方法引入脑网络的构建
- 构建网络时不需确定阈值，而是在度量空间的全尺度范围内，通过不断增加阈值分析网络，动态提取不同尺度下网络的拓扑特征观察大脑网络的拓扑变化，最终选择合适的度量性标准分析拓扑特征的稳定性，提取持续拓扑特征来揭示不同脑网络之间的差异。
- 优势：对信号中的噪声不敏感。

持续同调

- 持续同调 (Persistent Homology)
- 是拓扑数据分析的一种重要工具，它通过持续过滤的方法捕获度量空间中的持续变化的高维拓扑信息。
- **持续同调**包含同源性和持续性两部分内容。
- **同源性**在群论称作同调，它是拓扑集合分类的工具，可以度量一个单纯复形的特定结构,如连通组件、孔、洞；
- **持续性**是指给定一个 ε ，在 ε 所有可能的值下计算哪些结构是持续存在的，即获得持续稳定的拓扑特征。
- 特征中能够保持时间长的就是有用的特征，而寿命较短的可能是噪声，这个过程就称为持续同调。
- **贝蒂数**可以作为点云在复形结构上的拓扑不变量。

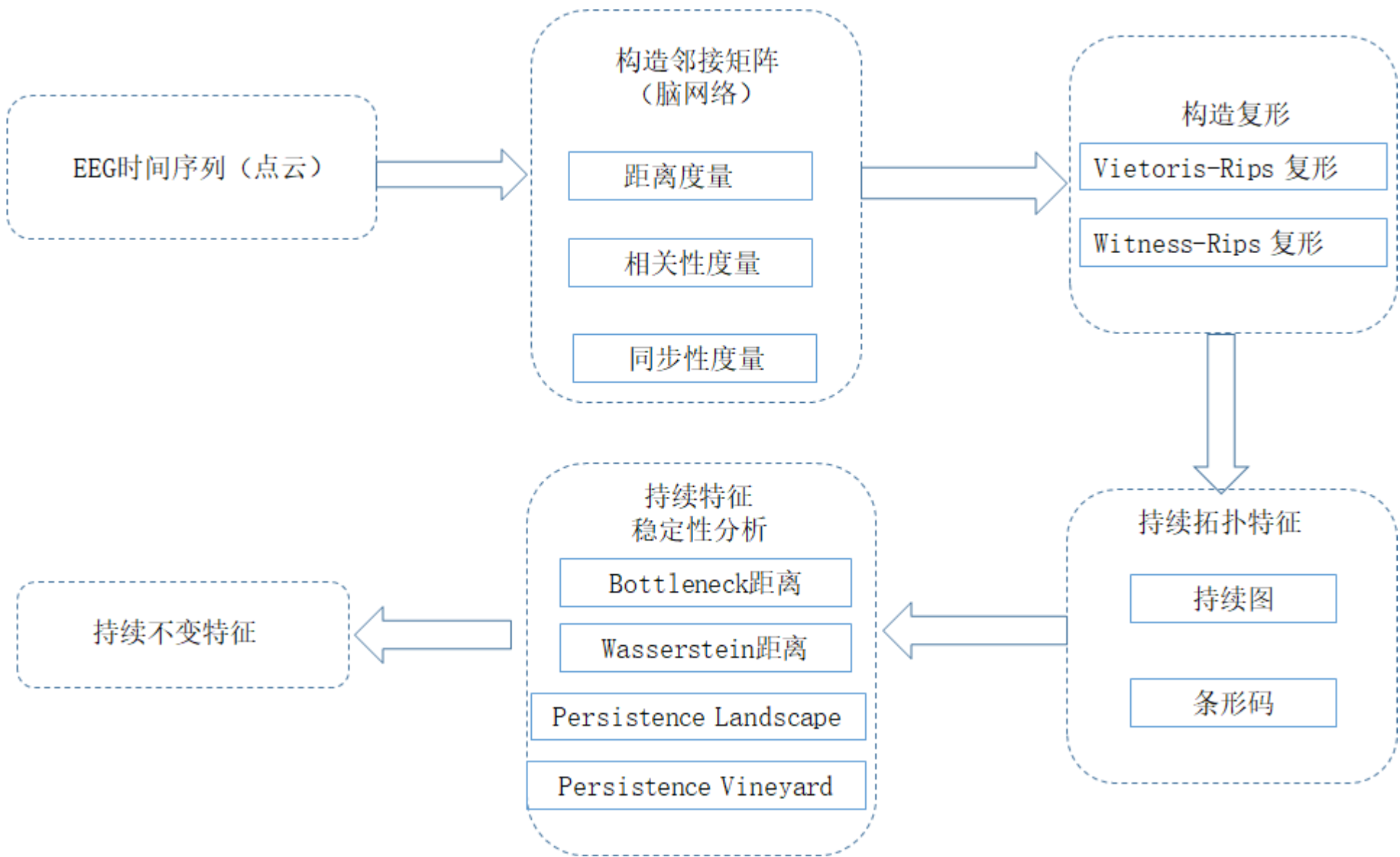


基于持续同调的脑网络构建和分析流程

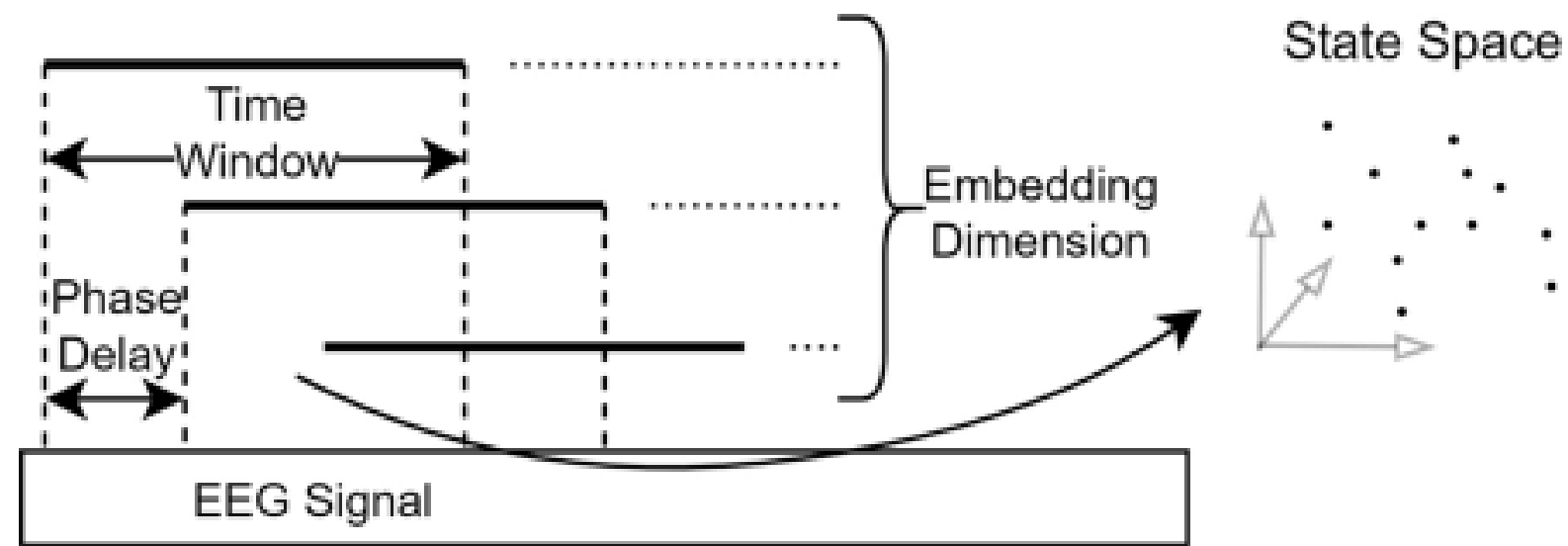


In this study we investigate three representations: 1) X is obtained by considering the Pearson correlations among channels and results in a description of spatial correlations, 2) T is the Takens (or delay) embeddings starting from the multivariate EEG timeseries and explicitly encoding temporal correlations within the signals, 3) D is a variant of T wherein the EEG timeseries is directly embedded, that is, without the imposition of time-delay vectors, effectively equivalent to considering the brain configuration space

基于持续同调的脑网络构建和分析流程



由时间序列到点云

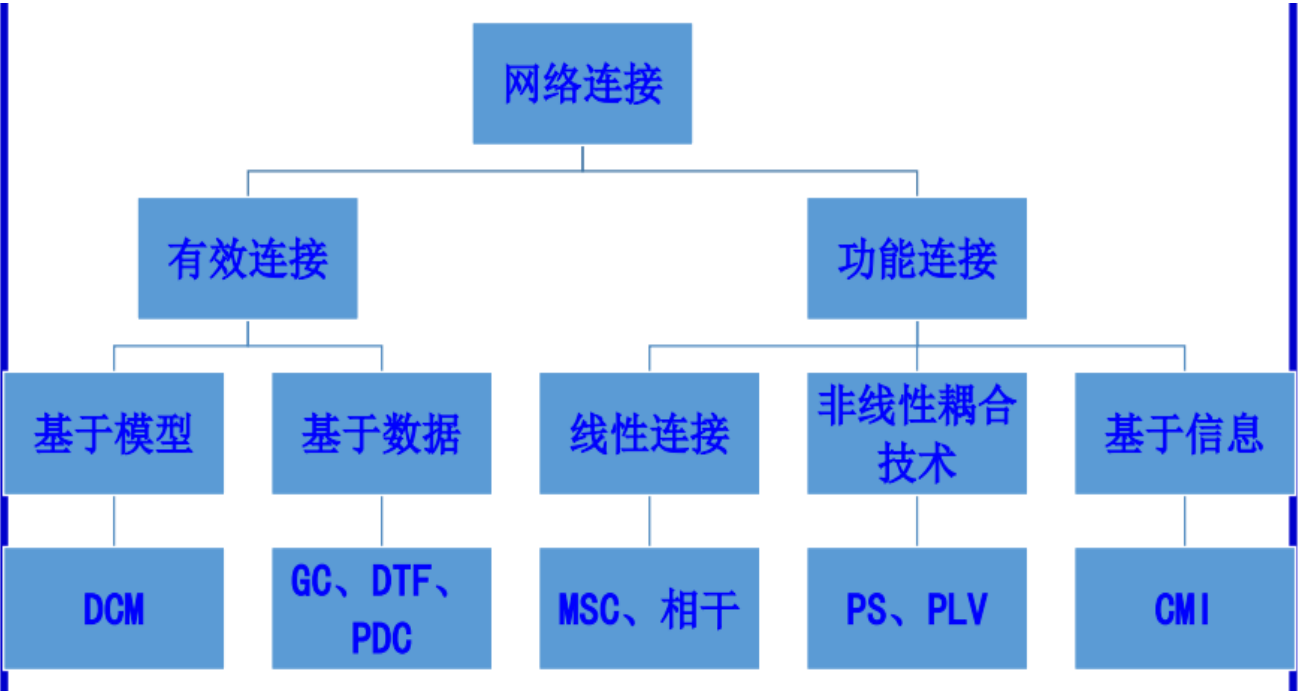


时间序列信号到3D点云的转换过程

. Time delay embedding uses signal samples as the coordinates of state space points. Time window determines the number of points, phase delay determines the amount of time window shift, and embedding dimension corresponds to the state space dimension.

- Time window : 决定点的数量
- phase delay : 决定时间窗口的移动量
- embedding dimension: 状态空间的维度

构造脑网络邻接矩阵



脑功能网络连接

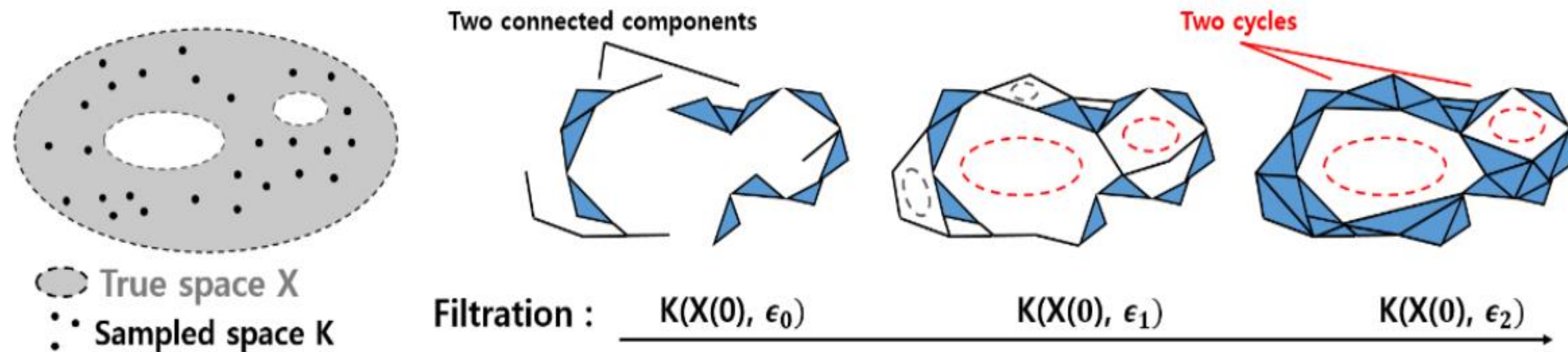
相位同步性: $PLV(t) = \frac{1}{N} \left| \sum_{n=1}^N \exp \left(j(\Delta\varphi_n(t)) \right) \right|$

相位滞后指数: $PLI = |\langle \text{sign}[\Delta\varphi(t)] \rangle|$

皮尔逊相关性: $PCC = \frac{COV(x,y)}{\sigma_x \sigma_y}$

欧式距离: $\rho = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$

由 3D 点云过滤嵌套复形



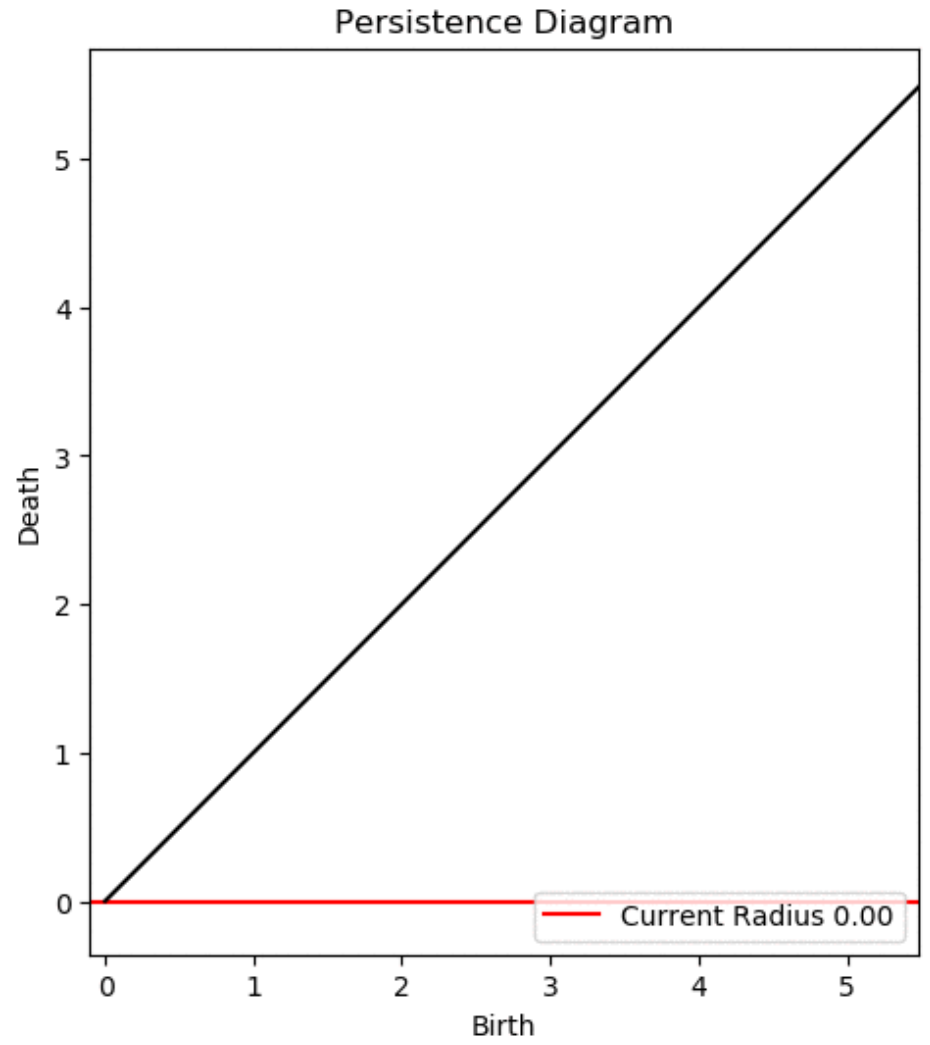
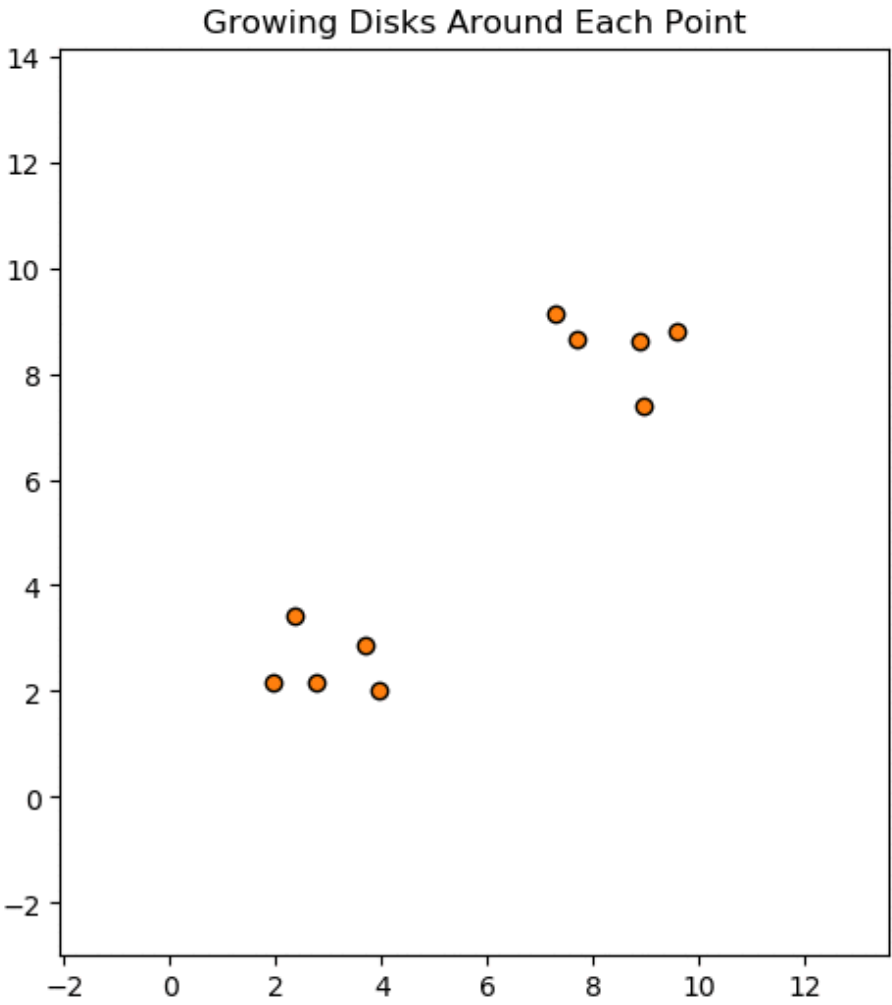
由点云过滤嵌套复形的过程

关键：

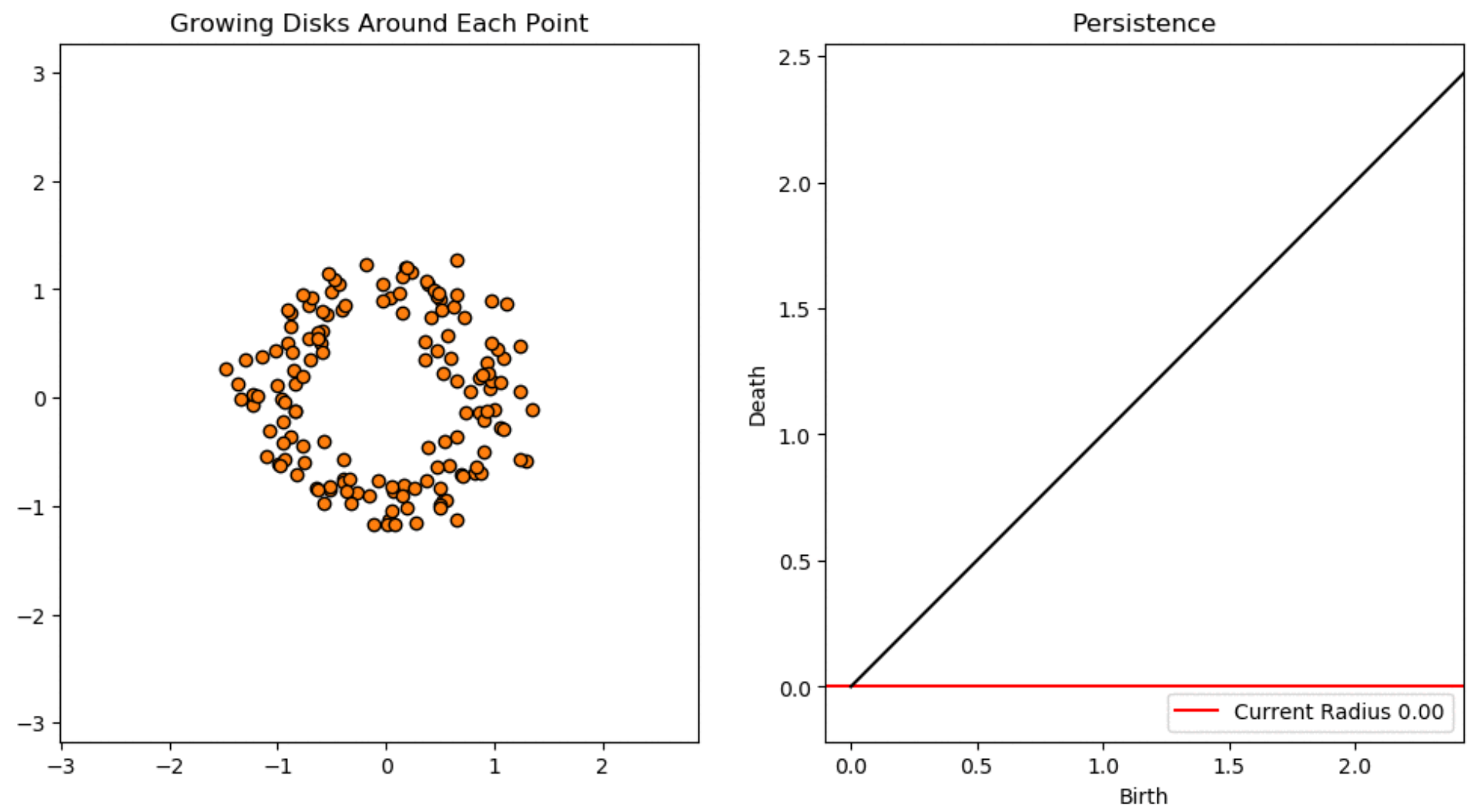
- (1) 过滤阈值 ϵ 的选择
- (2) 过滤算法的选择

● 现有疾病检测中 ϵ 用欧式距离或者皮尔逊相关系数。

过滤过程演示



过滤过程演示



过滤算法的选择

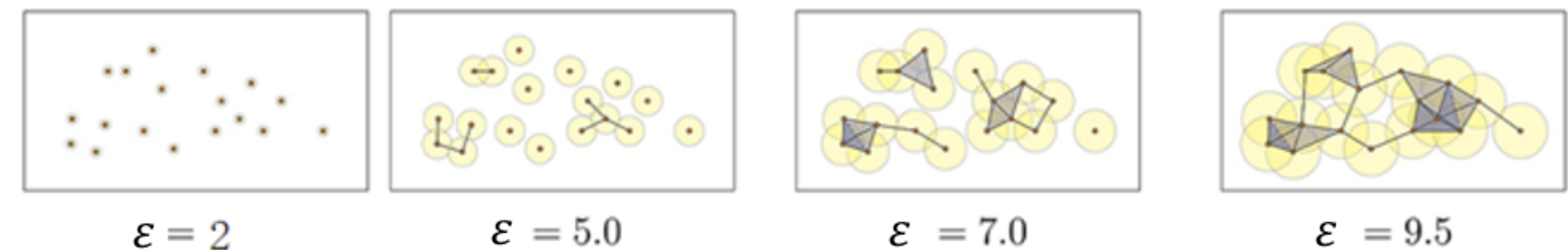
- Čech complex算法
- Vietoris–Rips complex算法
- Alpha complex算法
- Witness complex算法

Table 1 We summarize several types of complexes that are used for PH

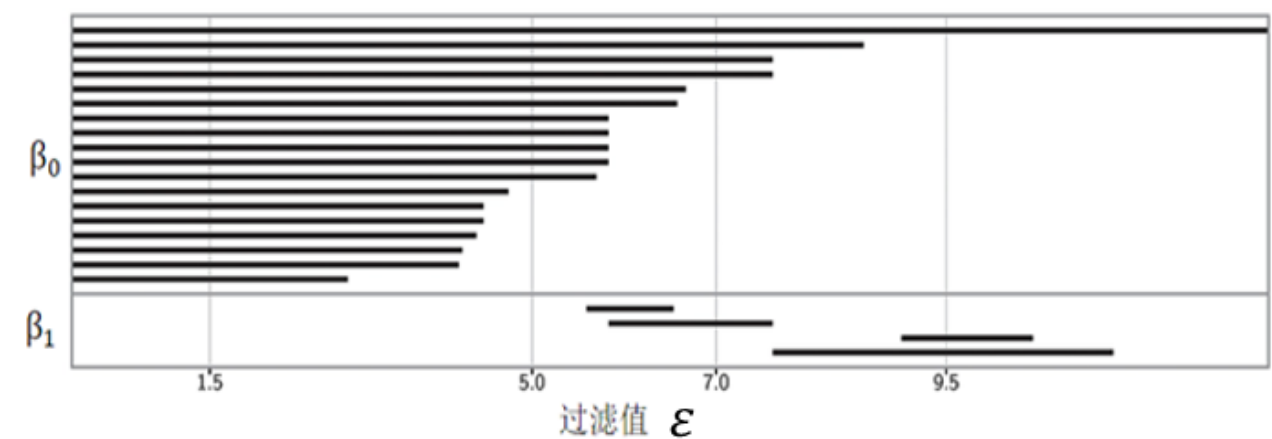
Complex K	Size of K	Theoretical guarantee
Čech	$2^{\mathcal{O}(N)}$	Nerve theorem
Vietoris–Rips (VR)	$2^{\mathcal{O}(N)}$	Approximates Čech complex
Alpha	$N^{\mathcal{O}(\lceil d/2 \rceil)}$ (N points in $\mathbb{R}^d$)	Nerve theorem
Witness	$2^{\mathcal{O}(I)}$	For curves and surfaces in Euclidean space
Graph-induced complex	$2^{\mathcal{O}(Q)}$	Approximates VR complex
Sparsified Čech	$\mathcal{O}(N)$	Approximates Čech complex
Sparsified VR	$\mathcal{O}(N)$	Approximates VR complex

Čech complex算法是经典的过滤算法，但是它的缺点是嵌套复形过程中需要检查大量的分岔，Vietoris–Rips complex算法是近似于Čech complex算法且规避了这一问题的一种具有高效计算性能的一种过滤算法，非常适合应用于复杂脑网络复形构造。

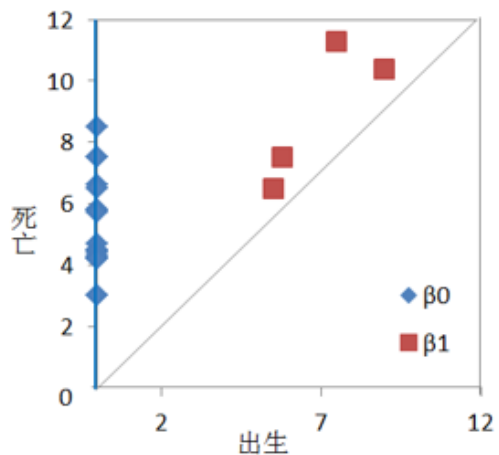
拓扑特征可视化



(a) 嵌套复形



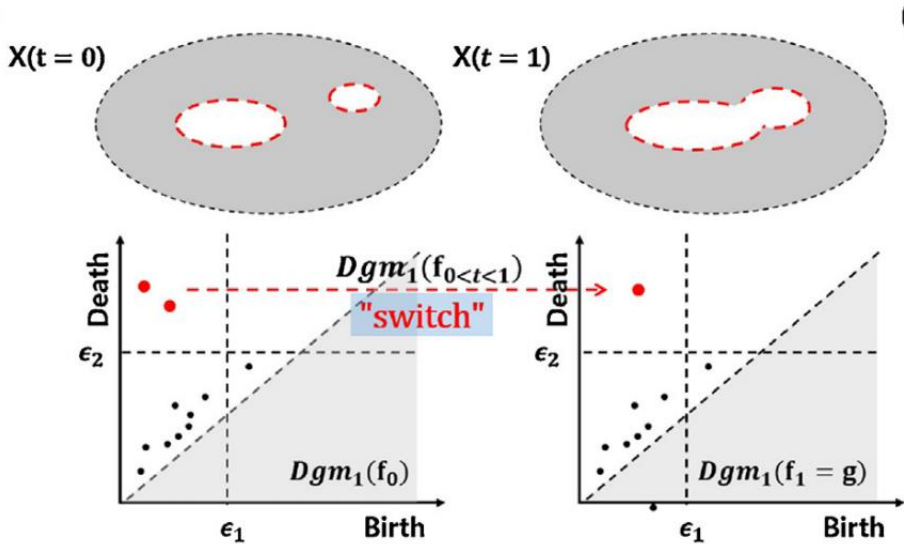
(b) 条形码



(c) 持久性图

持续同调下的特征演变

- 在条形码中，横坐标表示过滤阈值 ϵ ，用 $[\epsilon_{birth}, \epsilon_{death}]$ 的长度表示条形码的长度，条形码长度大的表示持续拓扑特征，条形码长度很短或者只是一个点的表示噪声。
- 每一维度下PD中每个VR复形的生成阈值 (bi) 与消失阈值 (di)，得到点集 $\{(bi, di, dim)\}$ ， dim 表示维度，由此可得到对应的PD。
- 当且仅当两个网络的PD相等时，两个持续性模块之间才存在同构。



持续拓扑特征的统计和解释

- (1) 研究随机单纯复形的拓扑性质，也就是在研究PH时，将随机单纯复形视为空模型，实验数据与空模型进行比较。
- (2) 研究持续图在特定度量空间的属性。使用稳定的度量空间来度量持久图之间的距离衡量。其优势是对数据集的一个小扰动只会在该指标之前的持久性关系图中造成一个小的变动，可有效的衡量持续拓扑特征的稳定性，并具有丰富的持续性同调数据包用于拓扑稳定性分析。
- (3) 研究持续性图的“特征”，即将持久性图的空间映射到适合统计分析和机器学习技术的空间（例如Banach空间）。该方法包括使用代数函数、持久性图像和核化技术，还有使用几何函数空间的Persistence Landscapes，适用于网络的动态拓扑特征。

持续拓扑特征稳定性分析

两个度量标准：Bottleneck距离和Wasserstein距离

Bottleneck距离度量两个持续图对应匹配点之间的最大距离，也就是说只输出最大离群值之间的距离，优势是它可以捕获持续图大的变化。简单来说，Bottleneck距离捕捉匹配点极端情况的距离最大值，Wasserstein距离侧重于全部对应匹配点的平均表现。

Wasserstein距离度量的是两个图对应匹配点之间的总距离，因此可以提供持久性图之间相似性的总体量化。有时是它对持续图小的变化比较敏感。

Bottleneck距离捕捉匹配点极端情况的距离最大值，Wasserstein距离侧重于全部对应匹配点的平均表现。

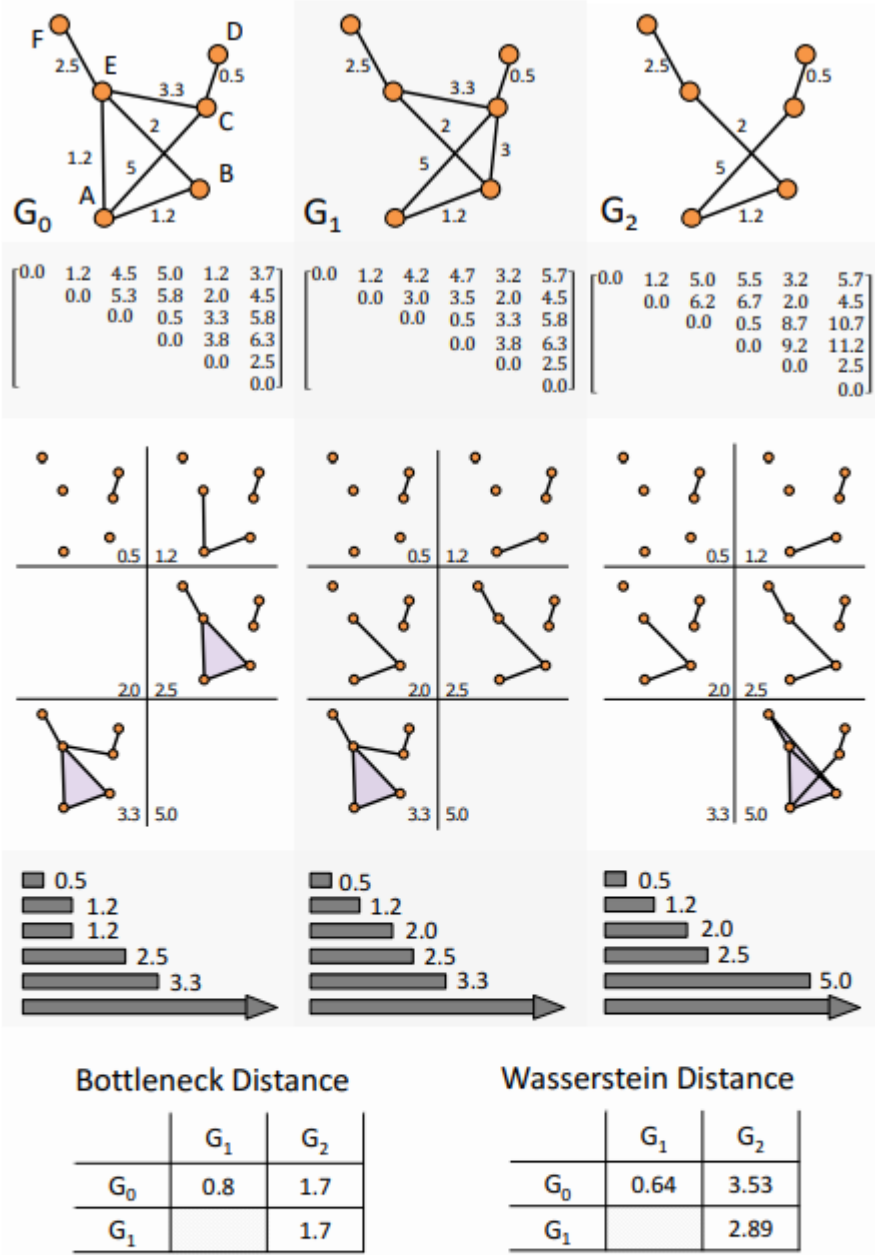
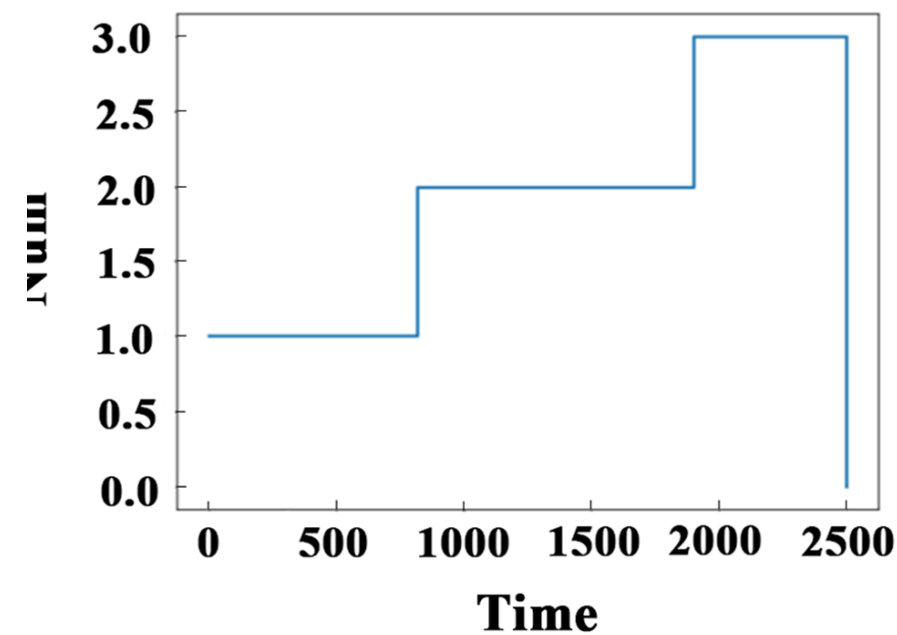
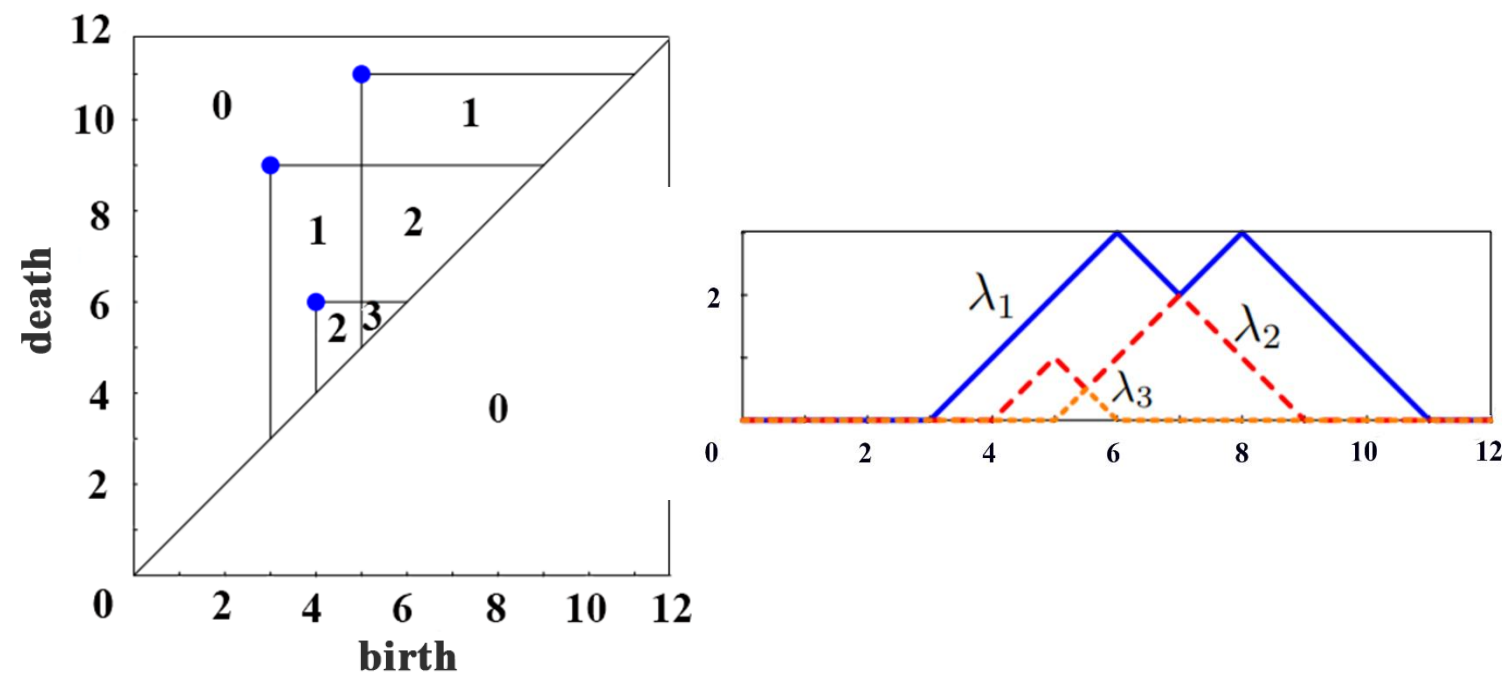


Figure 6: From left to right, 1st row: three weighted graph instances G_0 , G_1 and G_2 representing a time varying graph. 2nd row: each graph instance is embedded into a metric space, represented by a shortest-path distance matrix. 3rd row shows the filtrations in which topologically significant events occur, resulting in persistence barcodes in the 4th row. 5th row: the persistence diagrams are compared pairwise using bottleneck and Wasserstein distance.

持续图的多维度特征转换

PH 提取的拓扑特征（贝蒂数）无法直接用于机器分类算法输入，为此引入持续景观、贝蒂曲线、热核、持续熵等新的特征，这些特征本质上是将PD向量化，可直接应用于机器学习分类算法，且具有极高的可解释性与可视化性。



Heat kernel representation of diagram 0 in homology dimension 0



(1) 持续景观 (Persistence landscape, PL)

持续景观 (Persistence landscape, PL)

是持续图在Branch空间或Hilbert空间的映射。

(1) PD中的点 (b_i, d_i) 转换为 $(\frac{b_i+d_i}{2}, \frac{d_i-b_i}{2})$

(2) 由旋转持续图的特征创建的一组函数来定义持续景观，计算方法如公

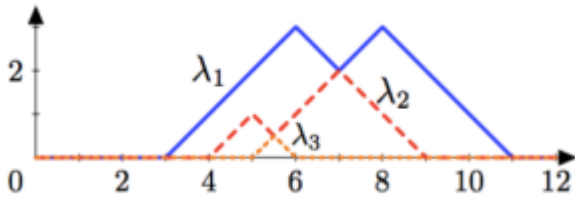
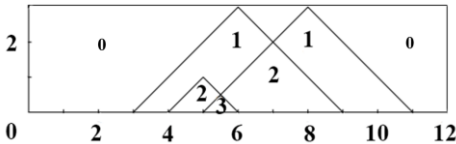
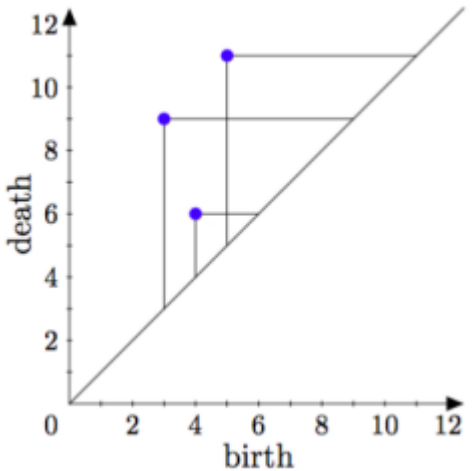
式下：

$$A_i(t) = \begin{cases} t - b_i, & t \in [b_i, \frac{b_i + d_i}{2}] \\ d_i - t, & t \in (\frac{b_i + d_i}{2}, d_i] \\ 0, & \text{其他} \end{cases} \quad (1)$$

$$\alpha_i(t) = \max(A_i(t), 0) \quad (2)$$

$$\gamma_k(t) = k \max\{\{\alpha_i(t)\}_{i \in I}\} \quad (3)$$

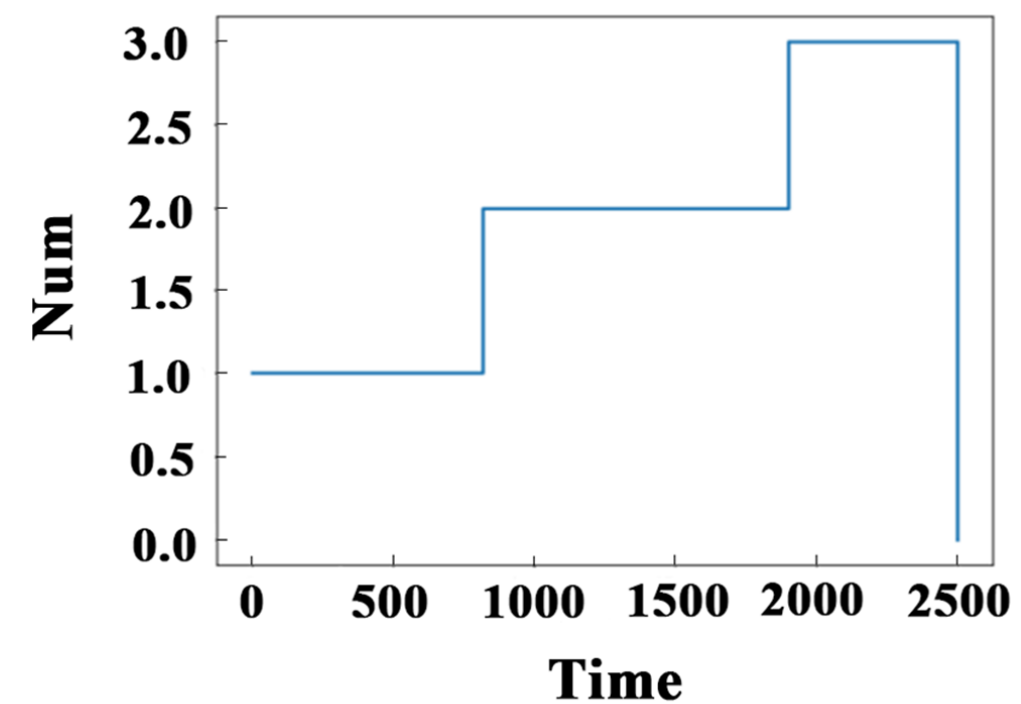
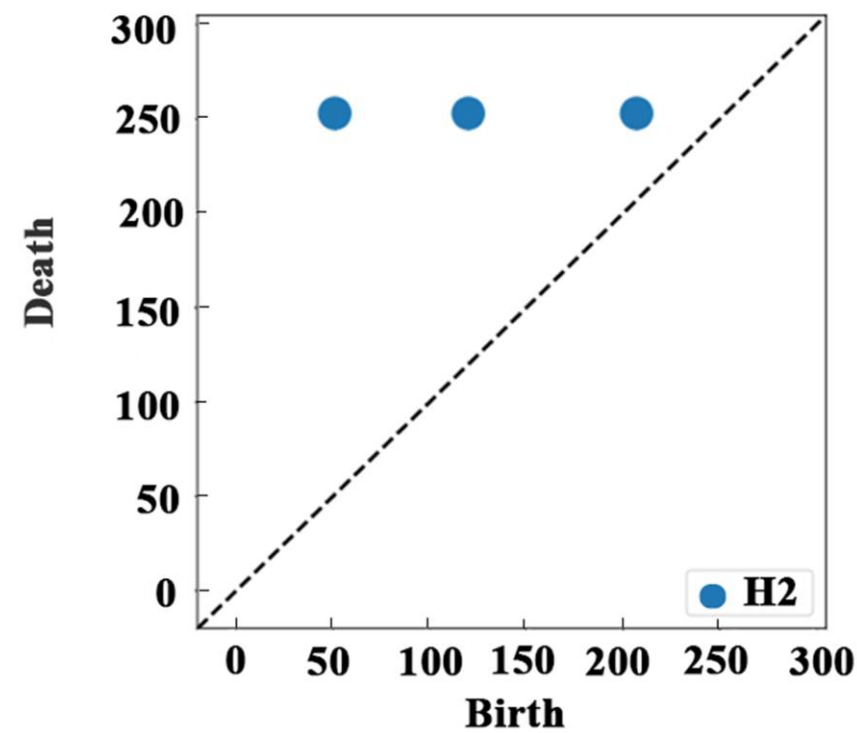
式中 $A_i(t)$ 是PD中的对角集合, $\alpha_i(t)$ 是每一同源维度 k 取 $A_i(t)$ 的最大值, 则 $\gamma_k(t)$ 为第 k 个同源维度的持续景观。



PD到PL转换过程

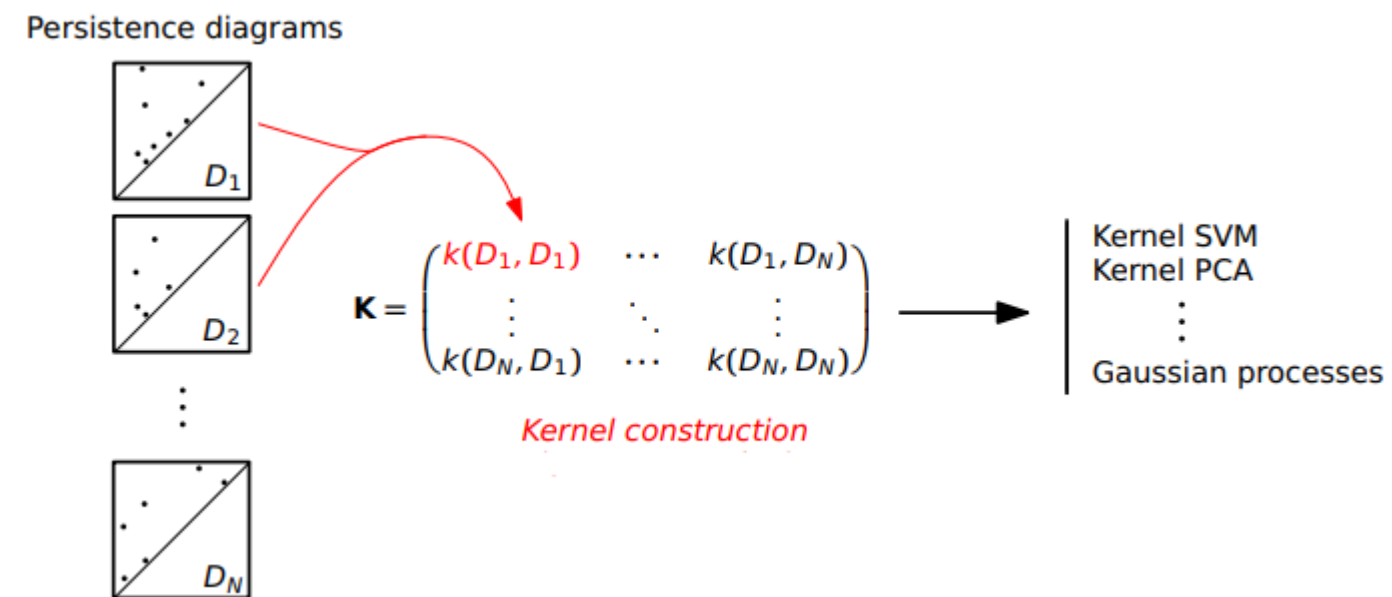
(2) 贝蒂曲线(Betti Curve, BC)

贝蒂曲线(Betti Curve, BC) 是由 VR 过滤中每个 ε 下每一维的拓扑特征个数组成。



(3) 热核 (Heat kernels, HK)

- 是一种多尺度的具有高斯核的 PD 的卷积^[14]。它将不同同源维度的 $PD = \{(d_i - b_i)\}_{i \in I}$ 作为 $Dirac-Delta$ 函数的和从指定的过滤参数中对 PD 进行均匀采样, 使得 PD 矩阵化, 然后使用高斯核对其进行卷积计算。对 PD 对角线的反射图也作同样的处理, 并计算两次卷积的差异, 结果为一个多通道的光栅图像即热图 (Heatmap, HM)。



(4) 持续熵(Persistence Entropy , PE)

是根据熵的定义对持续图进行熵的求解，计算如公式如下：

$$E(PD) = -\sum_{i \in I} p_i \log(p_i)$$

其中 $p_i = \frac{(d_i - b_i)}{L_D}$, $L_{PD} = \sum_{i \in I} (d_i - b_i)$ 。

持续熵通过与拓扑结构的强关联性，能全局地概括拓扑结构的信息。

应用1：精神分裂症显著性持续拓扑特征提取及分类

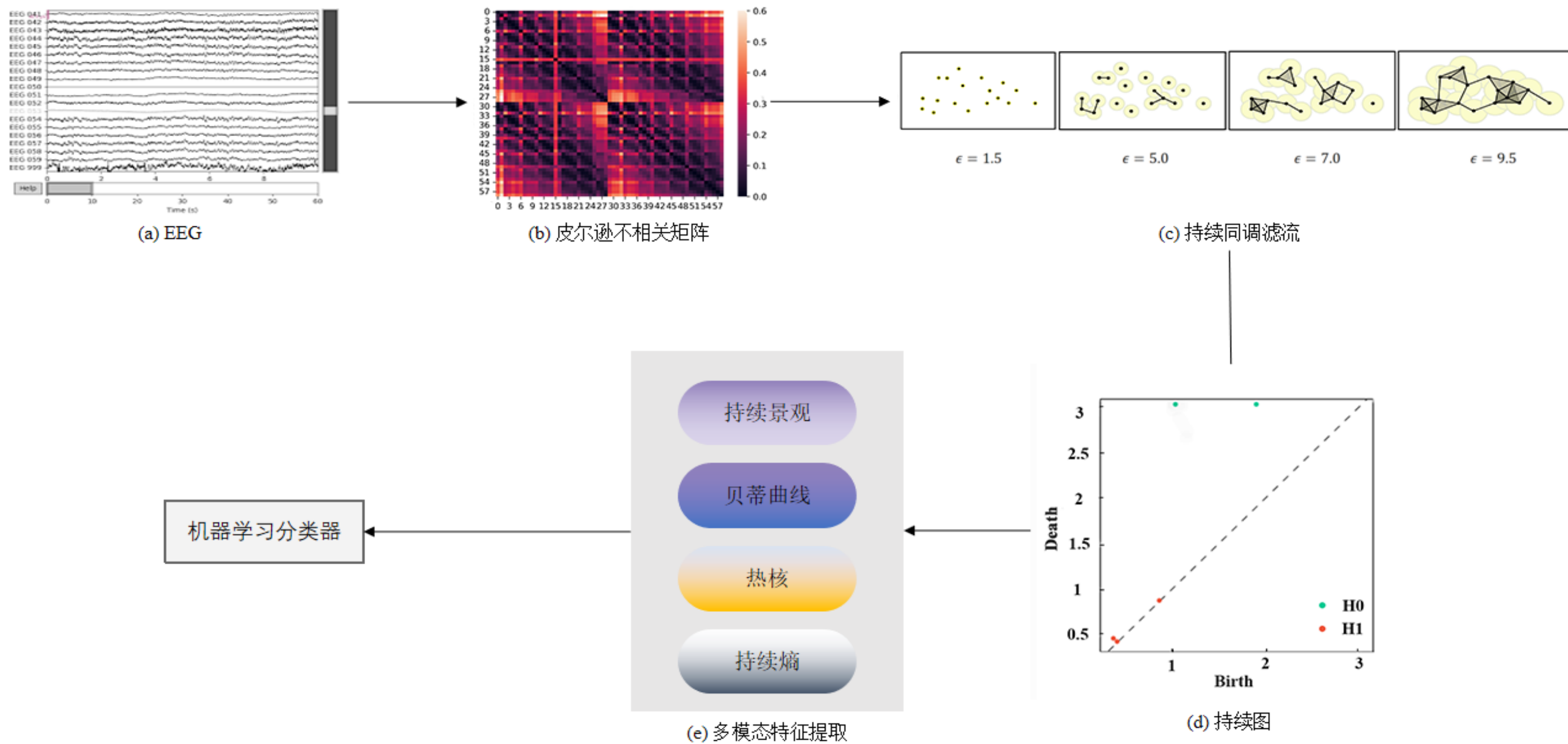
数据集包括 103例首发精分患者和92例健康被试的静息态脑电数据，两组被试在年龄、性别和受教育程度上均不存在统计学差异，如表1所示。数据采集设备使用NeuroScan公司的64导电极帽，采样频率为1000 Hz，阻抗不超过5 kΩ，左、右侧乳突连接参考电极，在左眼的上、下方放置电极记录垂直眼电，在右眼眼眶边缘放置电极记录水平眼电。

表 1 两组被试人口统计学和临床数据统计表

Table 1 Demographic and clinical data of the two groups of

特征	subjects		统计值
	精分病人 (n = 103)	健康被试 (n = 92)	
平均年龄 (years)	30.55±8.002	30.55±7.287	$F_{1,195} < 1$
受教育时 间(years)	14.17±2.494	14.79±2.557	$F_{1,195} = 3.022,$ $p = 0.084$
性别 (男/女)	51/52	48/44	$\chi_1^2 = 0.182,$ $p = 0.670$
PANSS 分数	74.31±10.87		
阳性分数	20.53±4.72		
阴性分数	17.18±5.63		

基于持续同调的脑网络分析模型



基于持续图的特征提取

LandScape("n_bins": 100,"n_layers":1,"p": 1) fea 1 2

LandScape("n_bins": 100,"n_layers":2,"p": 1) fea 3 4

LandScape("n_bins": 100,"n_layers":2,"p": 1) fea 5 6

LandScape("n_bins": 100,"n_layers":2,"p": 2) fea 7 8

BettiCurve("n_bins": 100,"p": 1) fea:9 10

BettiCurve("n_bins": 100,"p": 2) fea:11 12

HeatKernel("n_bins": 100,"p": 1, "sigma": 1.6) fea: 13 14

HeatKernel("n_bins": 100,"p": 1, "sigma": 3.2) fea: 15 16

HeatKernel("n_bins": 100,"p": 2, "sigma": 1.6) fea: 17 18

HeatKernel("n_bins": 100,"p": 2, "sigma": 3.2) fea: 19 20

PersistenceEntropy(nan_fill_value=4) fea: 21 22

每个EEG最后获得[1,22]的特征 X_fea

n_bins 采样点数目 sigma 高斯核的标准偏差 nan_fill_value 对空值的填充值 p 矩阵范数 (1范数 2范数)

机器学习分类

1 对X_fea添加一维特征：频段 共5个频段 用1-5进行表示获得[1,23]X_fea

2 总共有[980,22]数据X_data 划分X_train[900,23] X_test[85,23]

3 使用四种机器学习模型进行训练比较

SVM: kernel = 'linear' stop_tol = $1e^{-3}$

LR: penalty='l2' stop_tol = 0.0001

RandomForest: n_estimators= 10 criterion = gini max_depth=default

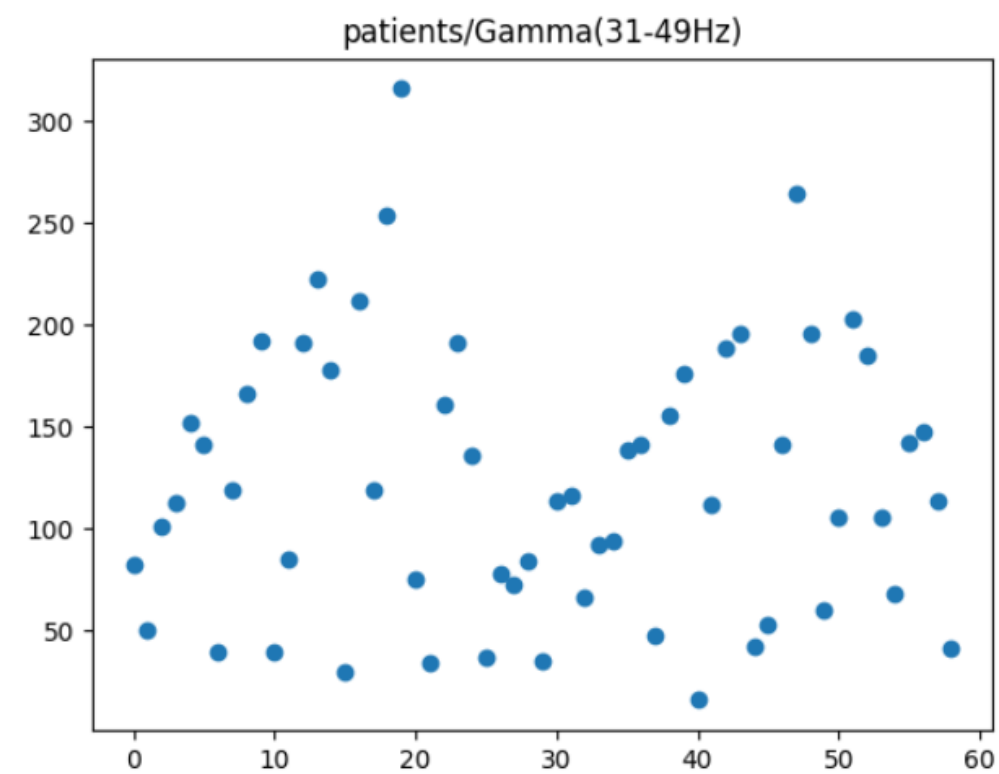
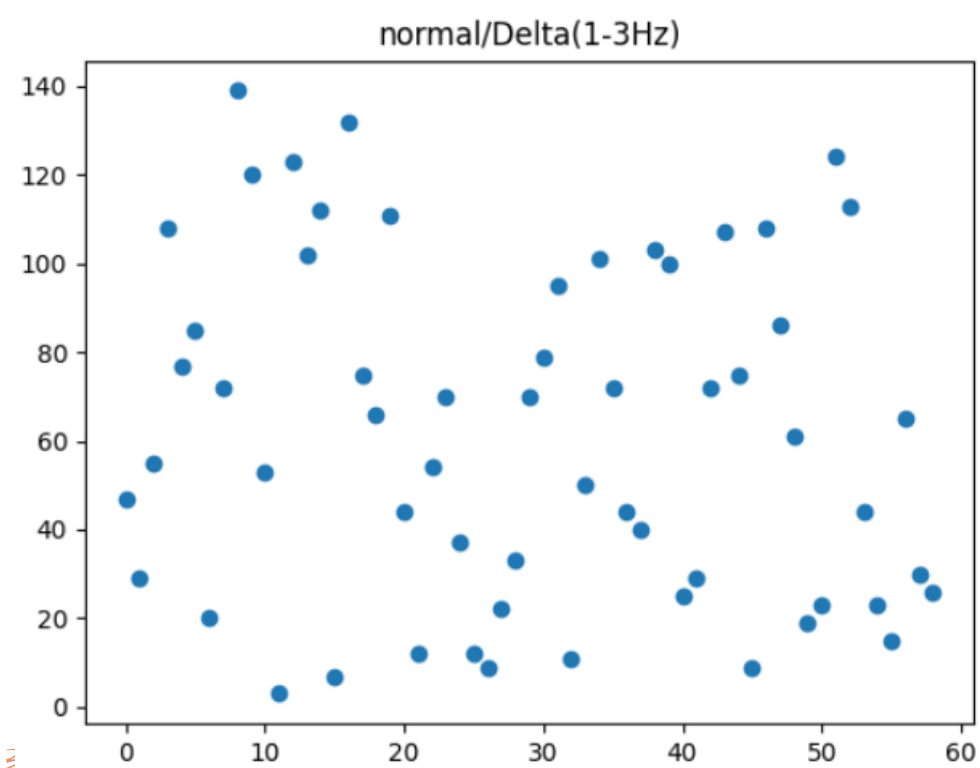
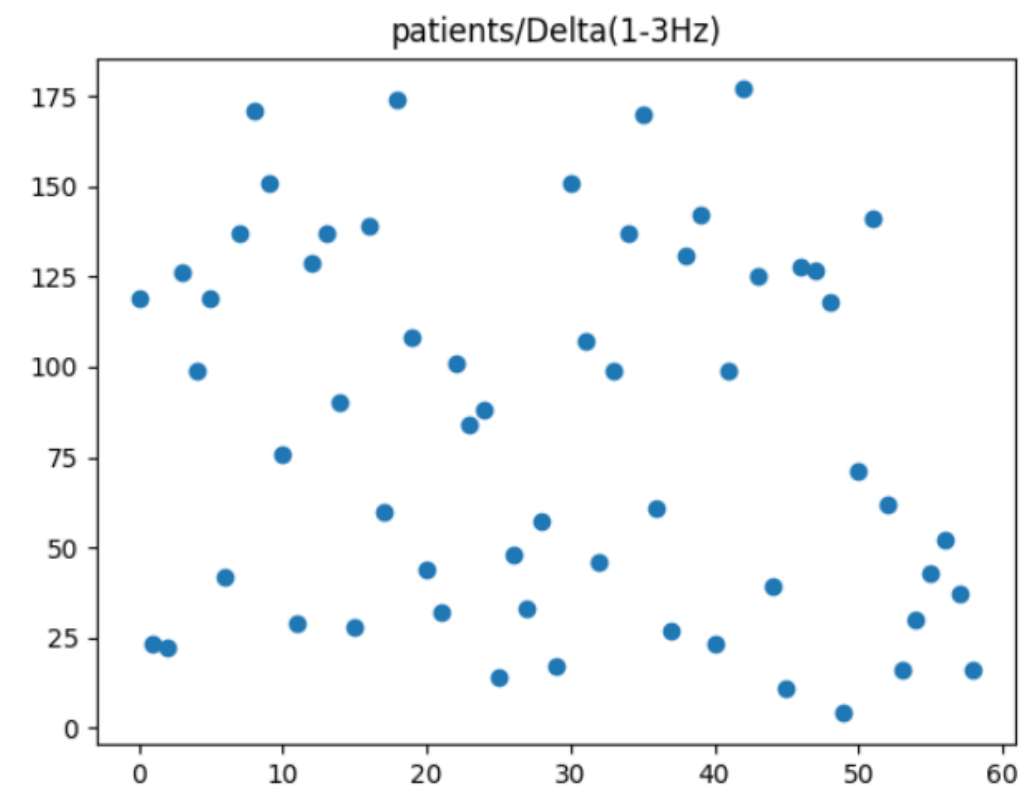
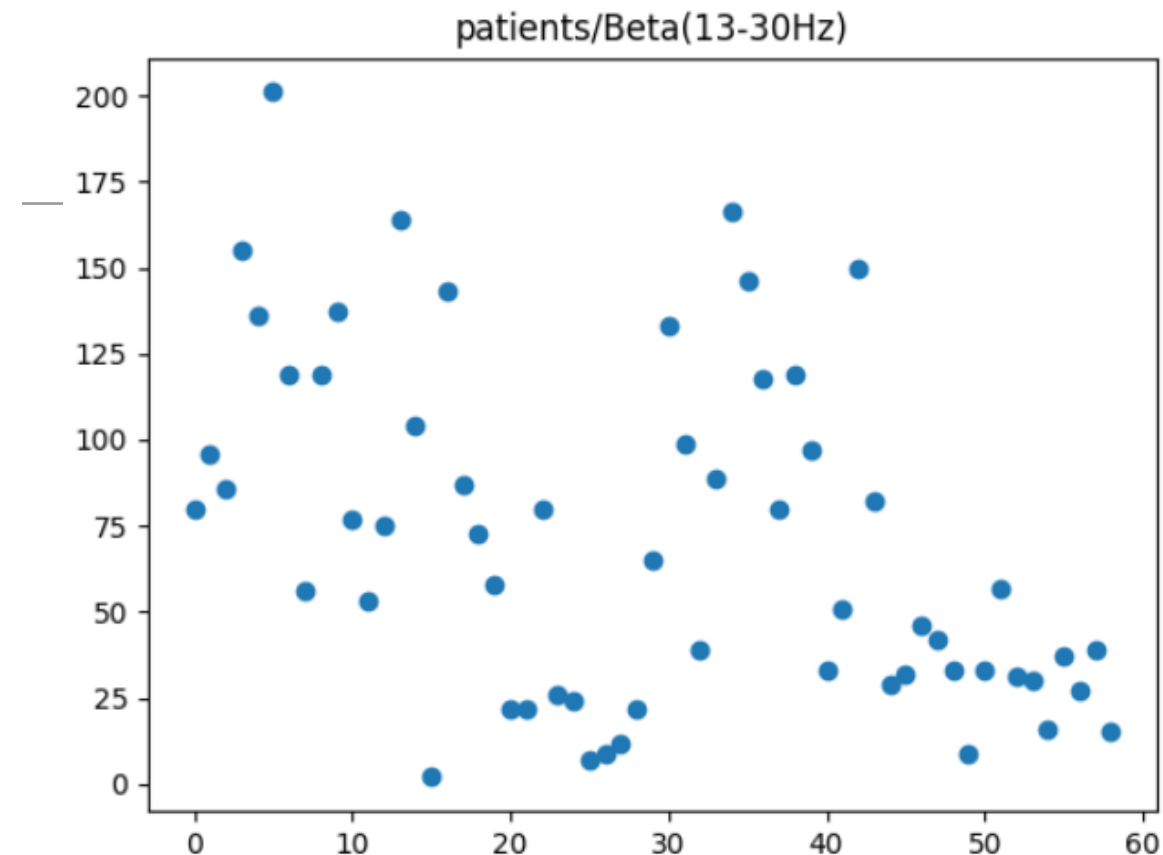
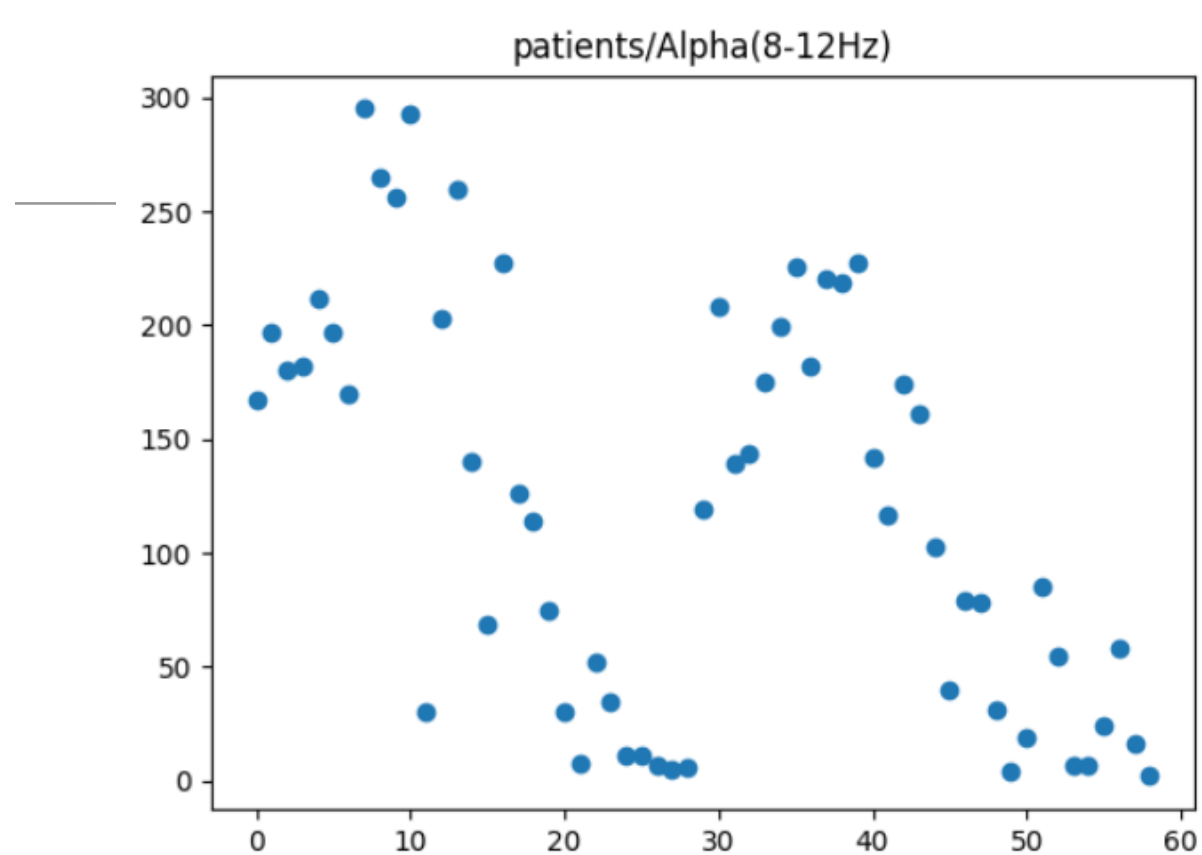
LigthGBM: num_leaves=60 learning_rate=0.05 min_data_in_leaf = 30 round = 50

机器学习分类结果

特征属性/分类器	accuracy	percision	recall	F1-score
图属性（全频段）	0.721	0.723	0.714	0.718
贝蒂曲线	0.766	0.745	0.756	0.775
持续熵	0.822	0.812	0.815	0.756
持续景观	0.866	0.875	0.866	0.852
热图	0.837	0.833	0.812	0.836
贝蒂曲线+持续熵+持续景观 +热图	0.902	0.905	0.911	0.898

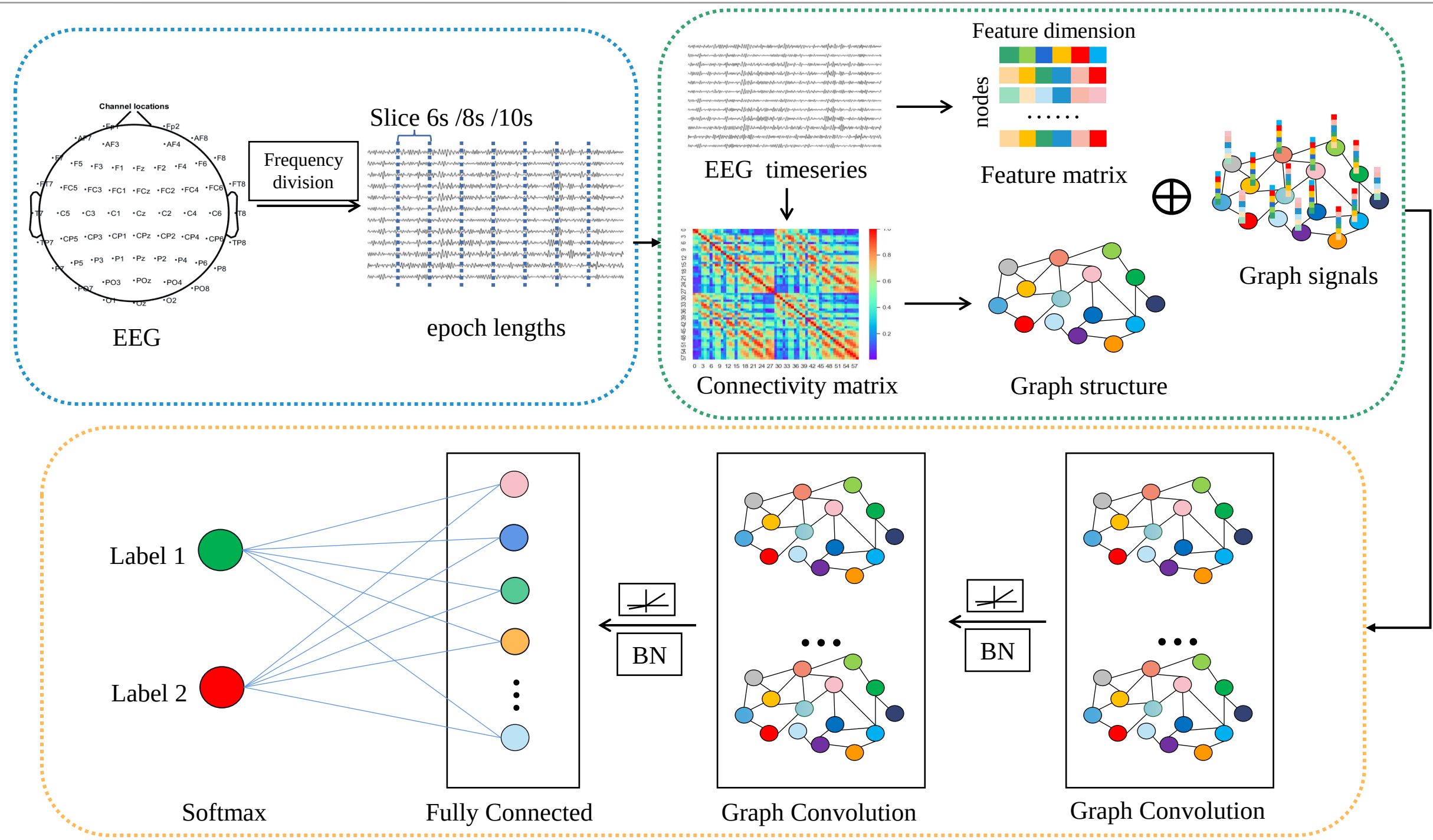
特征重要性排序

Feature importance			
0	16	42	BettiCurve "p": 1
1	1	18	BettiCurve "p": 1
2	2	36	BettiCurve "p": 2
3	3	31	BettiCurve "p": 2
4	4	19	HeatKernel "p": 1, "sigma": 1.6
5	5	41	HeatKernel "p": 1, "sigma": 1.6
6	6	30	HeatKernel "p": 1, "sigma": 3.2
7	7	32	HeatKernel "p": 1, "sigma": 3.2
8	8	68	HeatKernel "p": 2, "sigma": 1.6
9	9	29	HeatKernel "p": 2, "sigma": 1.6
10	10	20	HeatKernel "p": 2, "sigma": 3.2
11	11	15	HeatKernel "p": 2, "sigma": 3.2
12	12	13	PersistenceEntropy
13	13	13	PersistenceEntropy
14	14	42	Frequency_band



更多“人工智能&复杂系统”

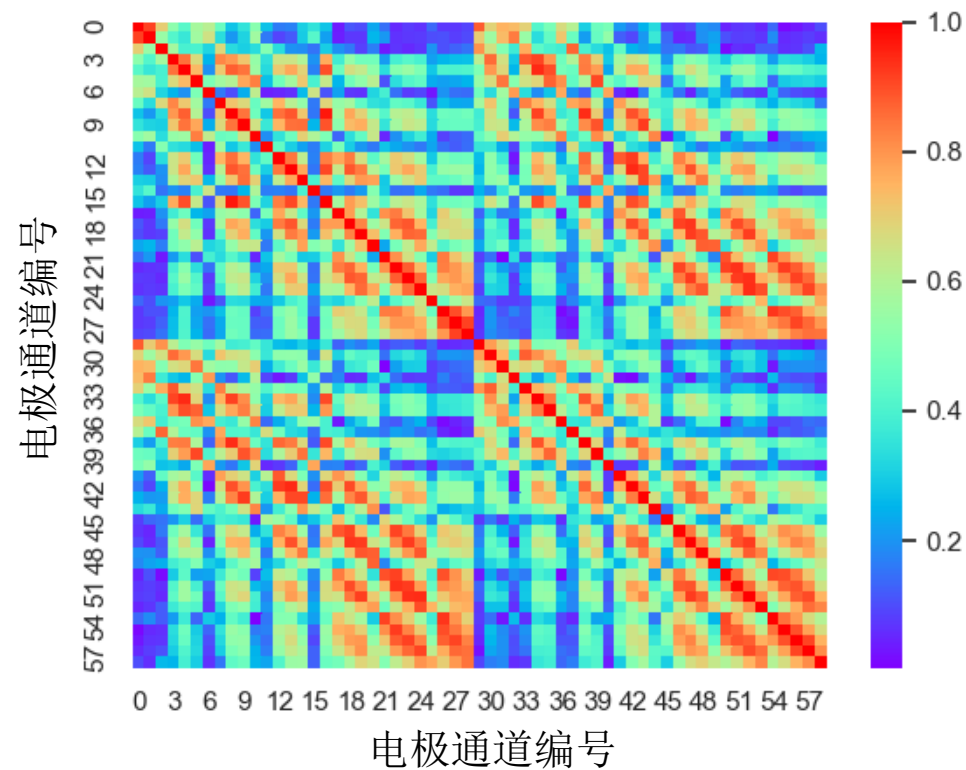
应用2：基于图卷积的精神分裂症分类预测模型



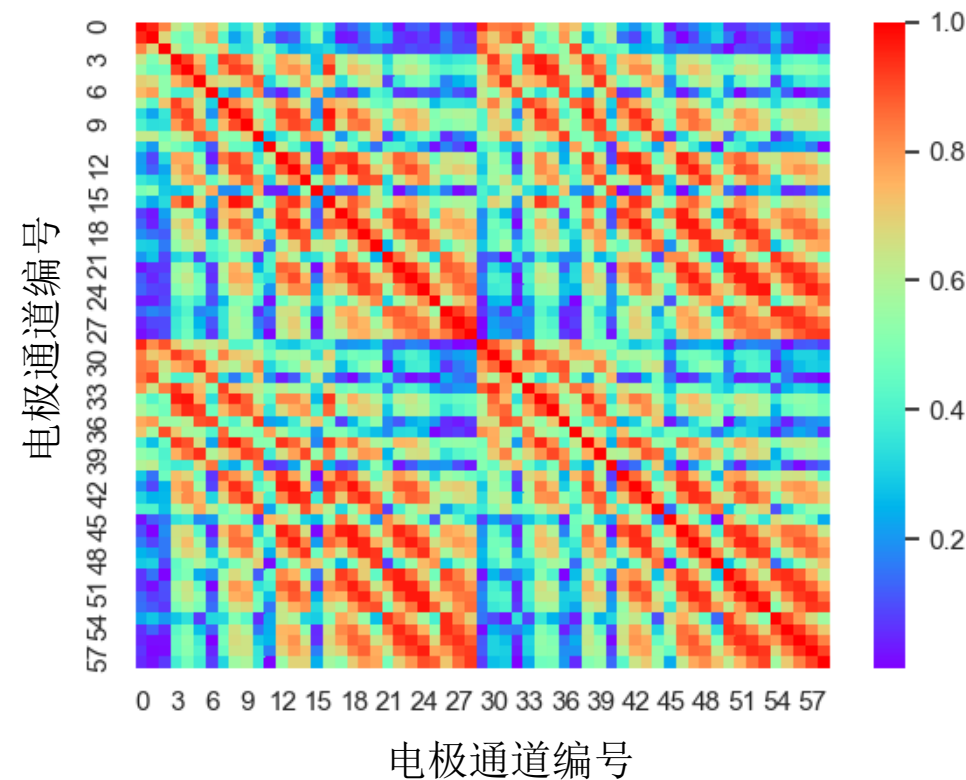
构建图信号

(1) 构建图结构

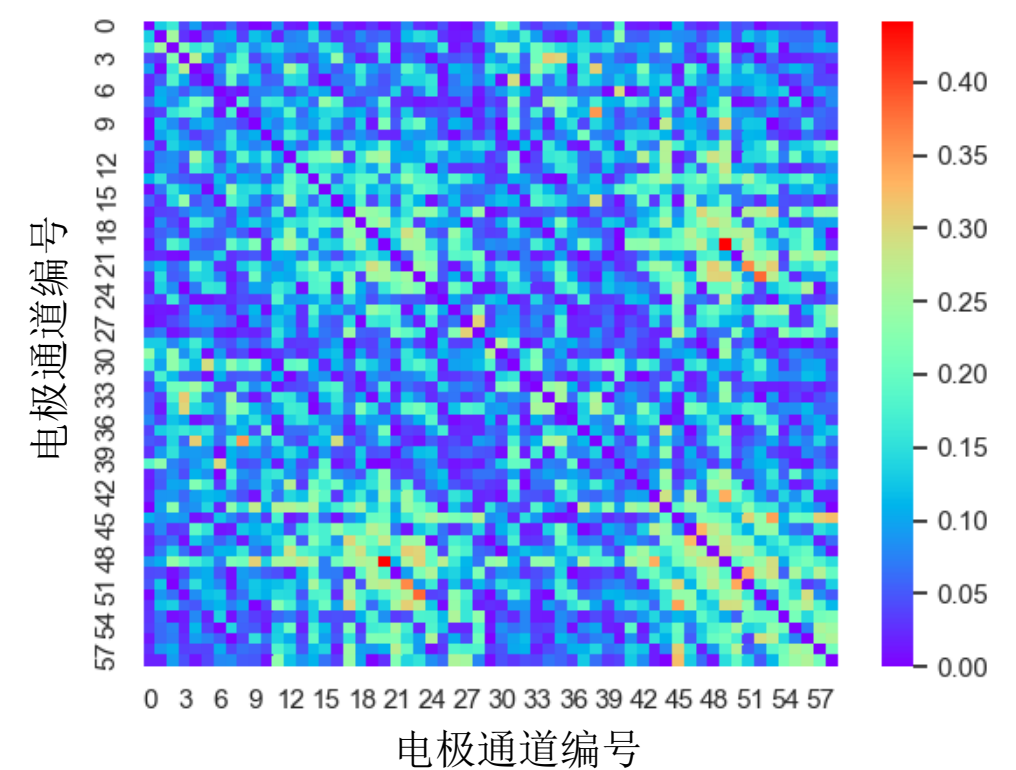
对于每个被试5个频段、3个分段长度下分别采用3种功能连接指标（PLV、 $|PCC|$ 、PLI）构建脑网络，大小为 59×59 ，取值范围为 $[0, 1]$ ，构建的脑功能网络作为脑图的邻接矩阵形成图结构。



(a) PLV关联矩阵



(b) PCC关联矩阵



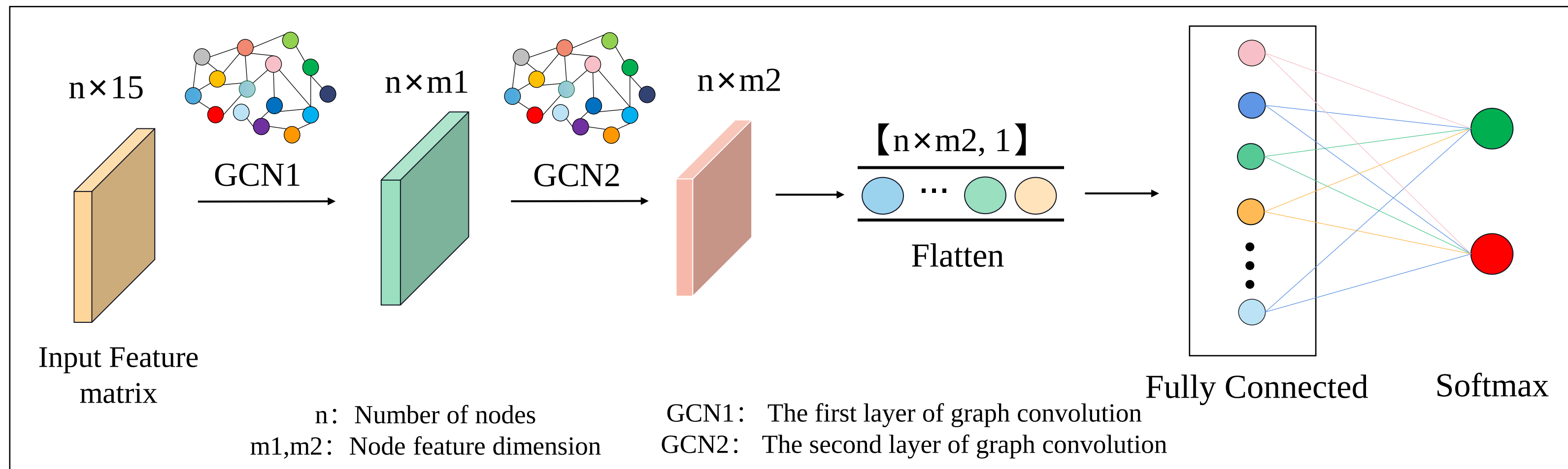
(c) PLI关联矩阵

(2) 节点特征

EEG信号节点特征采用时、频域和脑网络局部特征表示。对于时域特征，计算平均绝对值(MAV)、方差(Variance)、均方根(RMS)、差值绝对标准差(DASDV)、峰值(Peak)、偏斜度(Skewness)和峰度(Kurtosis)、最小值(MIN)、最大值(MAX)、标准差(STD)，对于频域特征，实验计算了平均功率谱密度，对于脑网络局部特征，实验计算了度、介数、聚类系数和局部效率，得到节点的特征矩阵为 59×15 。

对于每个样本，构建的图结构（邻接矩阵）和节点特征矩阵一起作为图信号输入图卷积层进行卷积运算。

图卷积和分类预测



图卷积公式

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right)$$

其中， $\tilde{A} = A + I$ ，其中 I 为单位矩阵， $\tilde{D}$ 为度矩阵， $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ 。 l 为卷积的层数， $H^{(l)}$ 为输入的节点特征矩阵， $H^{(l)} \in R^{N \times d}$ ，其中 N 为脑网络中节点的个数， d 为每个节点的特征维数。 $W^{(l)}$ 为权重矩阵， $W^{(l)} \in R^{d \times F}$ 。 $H^{(l+1)}$ 为输出的特征矩阵， $H^{(l+1)} \in R^{N \times F}$ 。

实验设置

实验采用十折交叉验证，将样本数据随机分为10份，其中1份作为测试集，其余9份作为训练集，生成的训练和测试集中的被试互不相交，最后将交叉验证的分类结果进行平均。

结果及分析

表1 EEG不同分段长度、频段和功能连接指标下模型平均准确率的比较

分段长度	功能连接指标	Delta	Theta	Alpha	Beta	Gamma
6	PLV	0.8380	0.9001	0.7748	0.7711	0.6731
	PCC	0.8513	0.8966	0.7995	0.7490	0.6734
	PLI	0.6418	0.6490	0.5796	0.6132	0.6271
8	PLV	0.8379	0.8948	0.7519	0.7225	0.6603
	PCC	0.8391	0.8802	0.7625	0.7209	0.6536
	PLI	0.6500	0.6456	0.5883	0.5996	0.6262
10	PLV	0.8280	0.8909	0.6734	0.7163	0.6361
	PCC	0.7991	0.8704	0.7159	0.6954	0.6216
	PLI	0.6450	0.6456	0.5751	0.6067	0.6280

将患者组和正常组的脑网络局部特征AUC值进行K-S检验

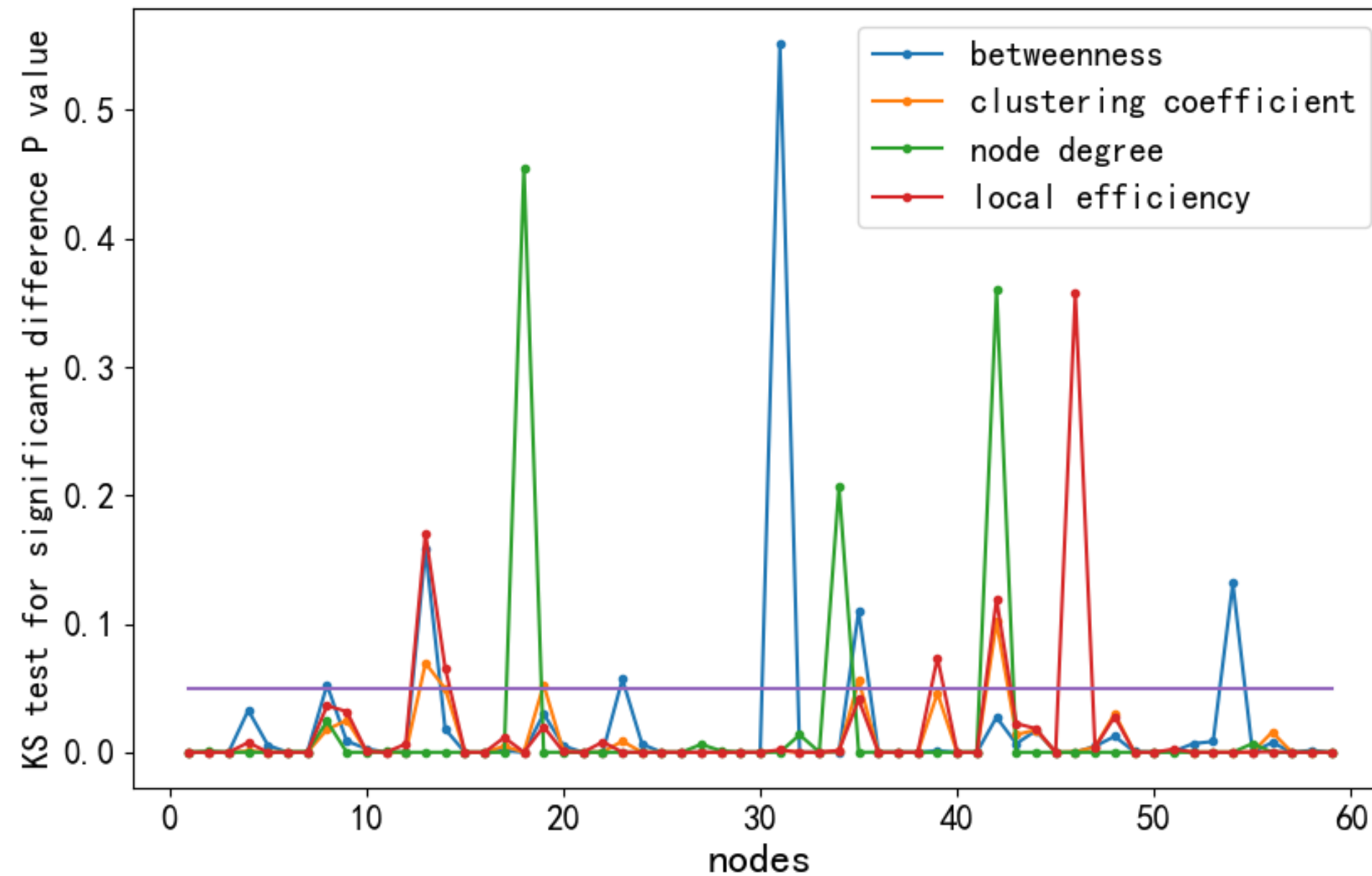


图1 不同节点局部特征差异值

表2 不同特征矩阵对图卷积网络的影响

特征矩阵	Accuracy	Precision	Recall	F1-Score
A	0.8368	0.8166	0.9328	0.8708
B	0.8858	0.8669	0.9526	0.9077
C	0.9106	0.9107	0.9407	0.9255
D	0.9262	0.9278	0.9486	0.9381

说明：矩阵A是仅使用时域特征构成的特征矩阵。矩阵B表示仅使用频域特征构成的特征矩阵。矩阵C表示仅使用脑网络局部特征构成的特征矩阵。矩阵D是本文采用的包含时、频域和脑网络局部特征的特征矩阵。

结论：使用时、频域特征和脑网络局部特征相结合时模型的性能最佳，**分类准确率为92.62%，精确率为92.78%，召回率为94.86%，F1-Score为93.81%**，这是因为与其他三种特征矩阵相比，矩阵D集合了多个方面的特征信息，经过图卷积运算可以充分学习时、频、拓扑特征信息。

表3 同类方法结果比较

方法	Accuracy	Precision	Recall	F1-Score	AUC
SVM	0.8127	0.8249	0.8381	0.8236	0.8841
CNN-2	0.8716	0.8649	0.9091	0.8805	0.9273
CNN-5	0.8606	0.8660	0.8807	0.8679	0.9232
EFC-GCN	0.9001	0.8818	0.9445	0.9085	0.9484

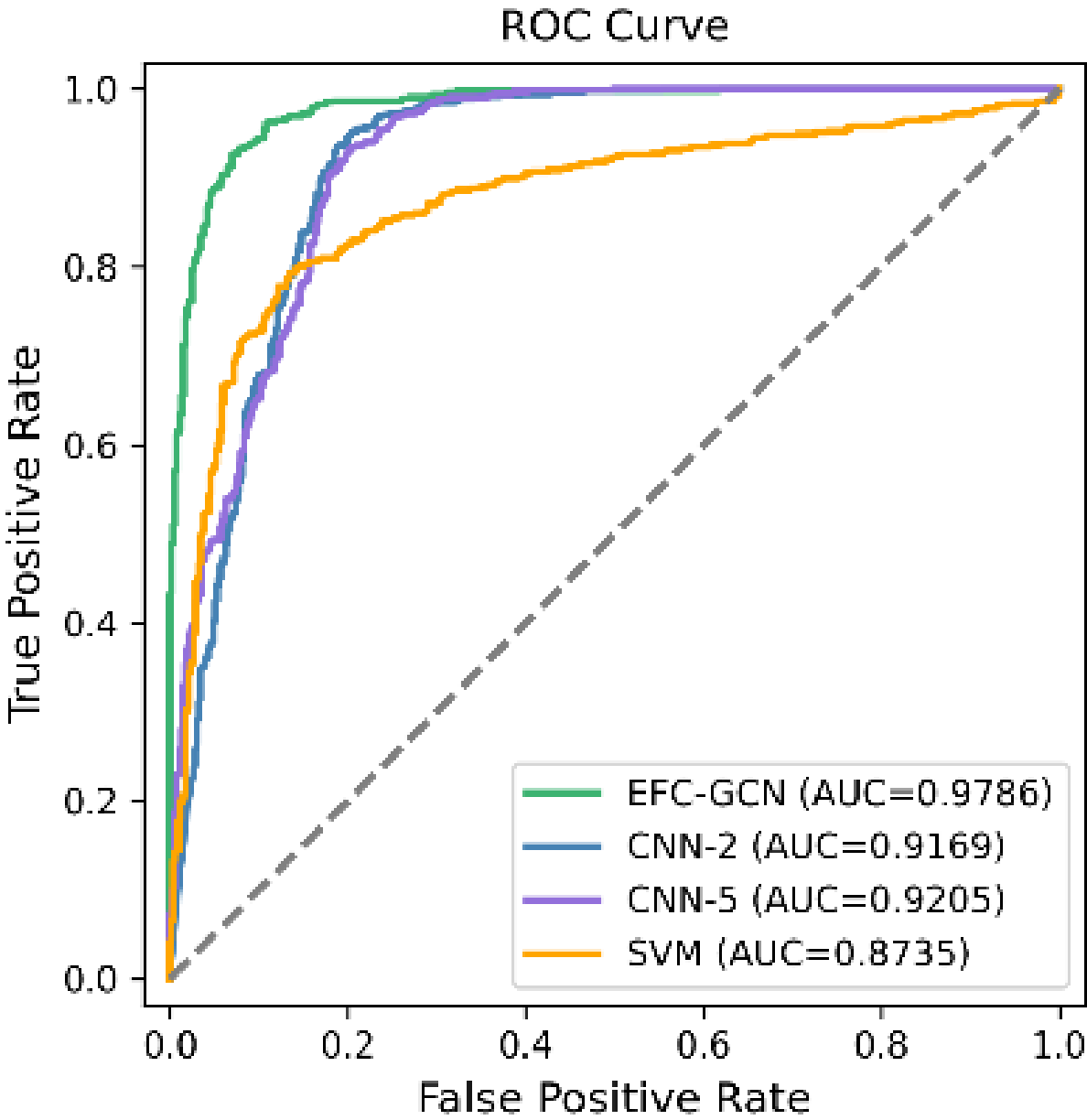


图5 各方法的ROC曲线

确定有助于分类的最显著区域

我们使用梯度加权类激活映射（Gradient-weighted Class Activation Mapping , Grad-CAM）来识别有助于GCN分类的最显著脑区。

Grad-CAM

假设 F_k 为最后一个卷积层的第 k 个特征图， y^c 为类别 c 的预测值（softmax层之前的值），对 y^c 进行反向传播，得到 F_k 的权重值为：

$$\alpha_k^c = \frac{1}{N} \sum_{n=1}^N \frac{\partial y^c}{\partial F_{k,n}}$$

其中 N 为节点的个数。计算类别 c 对于第 k 个通道的权重值 α_k^c ，节点 n 对于类别 c 的激活值可以计算为：

$$L_{Grad-CAM}^c[n] = ReLU \left(\sum_K \alpha_k^c F_{k,n} \right)$$

统计分析

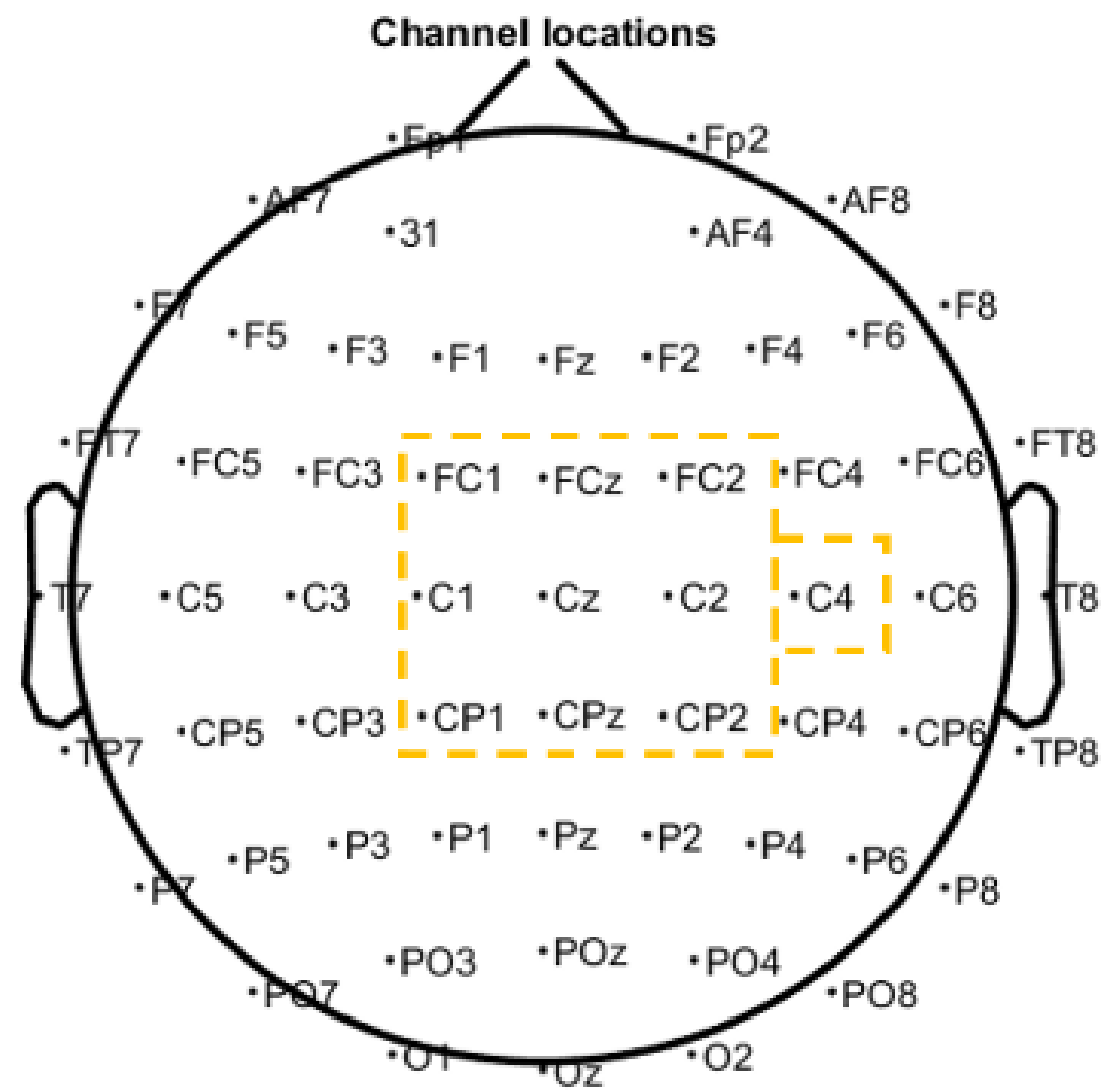


图2 前10个显著性节点

显著区域的网络拓扑分析

为了研究临床测量值与识别出的有助于分类的显著区域的拓扑特征之间的关系，将前10个显著节点的局部特征（度、介数和局部效率）与阳性分（PLBF）、阴性分（NLBF）、整体分（GLBF）和PANSS分进行皮尔逊相关性分析，若 $P < 0.05$ ，则表明两者之间存在显著性相关。

表 5 对 EFC-GCN 以及拓扑分析结果有贡献的前 10 个显著区域

Rank	Brain Region	P value of correlation between topological metrics and symptoms											
		Degree	Efficiency				Betweenness						
			PLBF	NLBF	GLBF	PANSS1	PLBF	NLBF	GLBF	PANSS1	PLBF	NLBF	GLBF
GCN													
1	C2	0.92231	0.43549	0.25754	0.40694	0.51847	0.64944	0.08858	0.20196	0.70272	0.41276	0.16987	0.3203
2	C1	0.64426	0.40957	0.19714	0.30167	0.68519	0.64918	0.0669	0.21214	0.76045	0.31052	0.09276	0.30043
3	Cz	0.51514	0.42979	0.16024	0.2624	0.85837	0.55047	0.09316	0.25695	0.68672	0.26788	0.18896	0.42714
4	FC2	0.65122	0.70171	0.21727	0.32955	0.4498	0.84087	0.04854	0.14363	0.50824	0.82701	0.04277	0.14711
5	CP1	0.57352	0.53941	0.07071	0.19072	0.79454	0.86619	0.23114	0.48831	0.47336	0.53083	0.80481	0.8979
6	CPz	0.73422	0.64193	0.11554	0.29405	0.76971	0.81317	0.27457	0.48373	0.23794	0.65648	0.91653	0.93359
7	FC1	0.79026	0.50643	0.22446	0.29752	0.28734	0.56336	0.03461	0.14966	0.18687	0.30585	0.02969	0.1818
8	CP2	0.50938	0.50962	0.13131	0.20715	0.75814	0.91199	0.15189	0.37896	0.30497	0.68365	0.42924	0.76965
9	FCz	0.56555	0.58602	0.20848	0.28949	0.51978	0.59016	0.04235	0.16756	0.56913	0.3879	0.04268	0.20707
10	C4	0.6004	0.67879	0.15123	0.30636	0.60249	0.97638	0.09778	0.17088	0.89717	0.99164	0.21685	0.29102

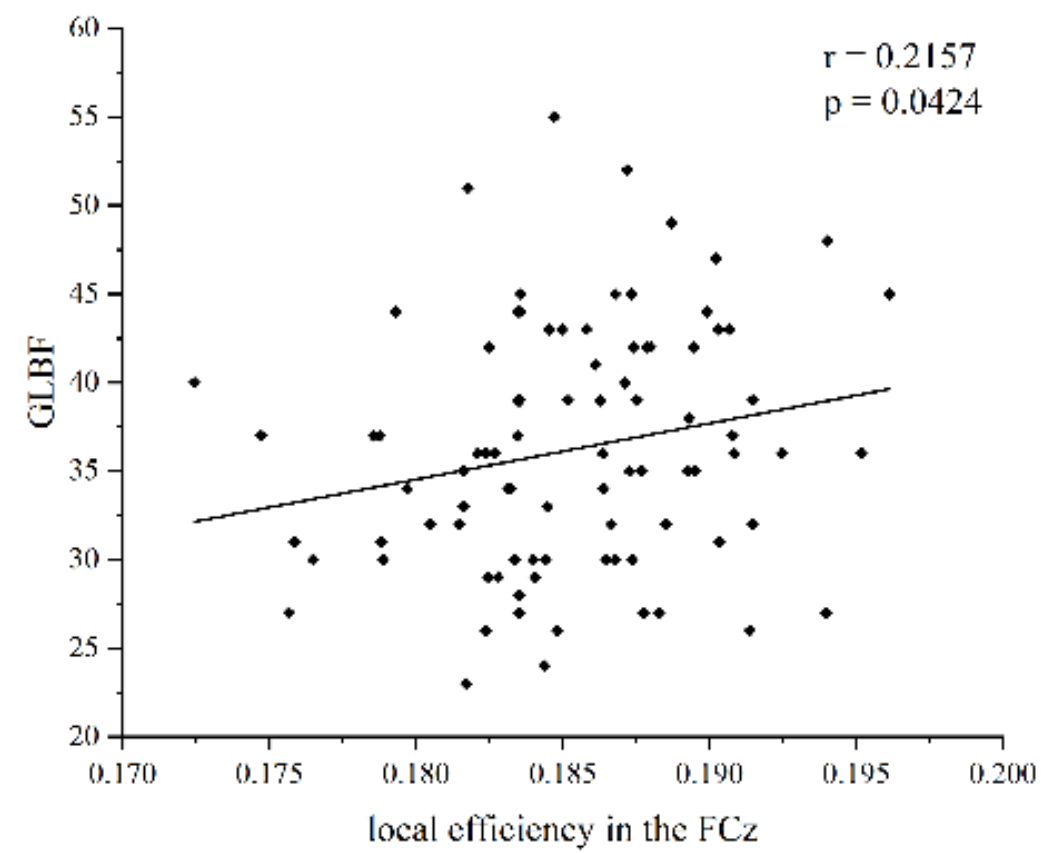
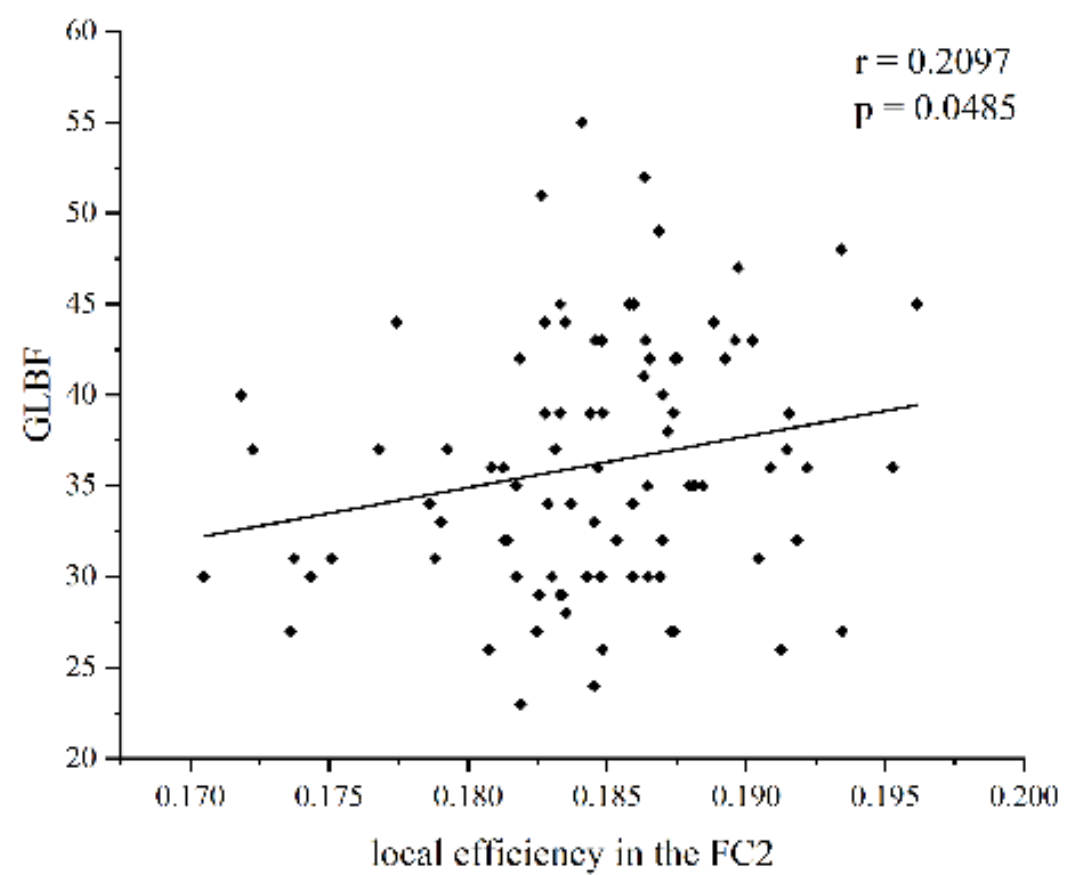


图3 网络拓扑特征与整体分的相关性

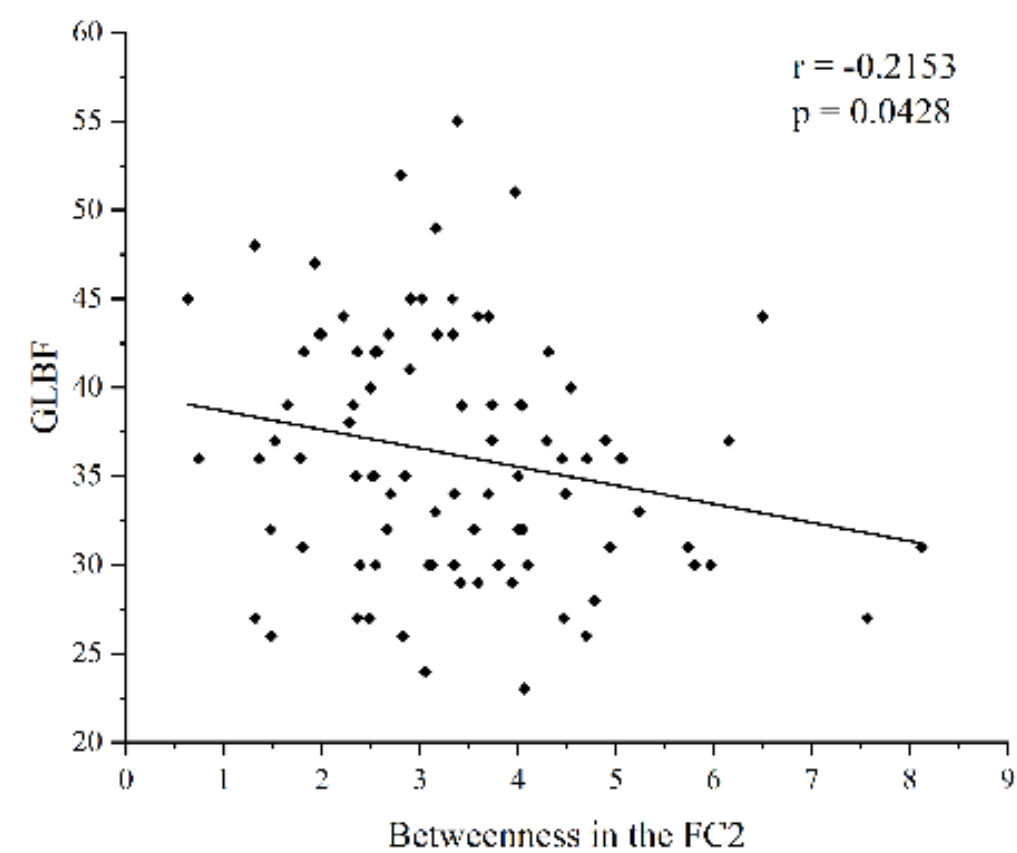
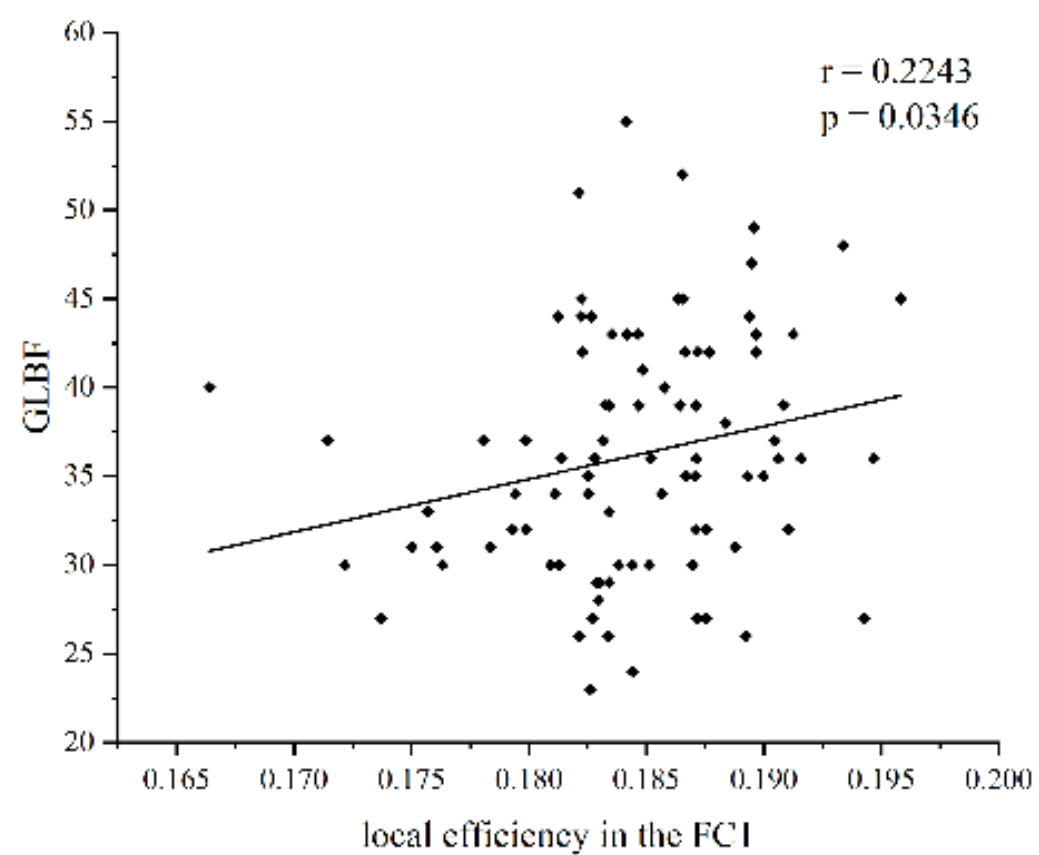


图3 网络拓扑特征与整体分的相关性

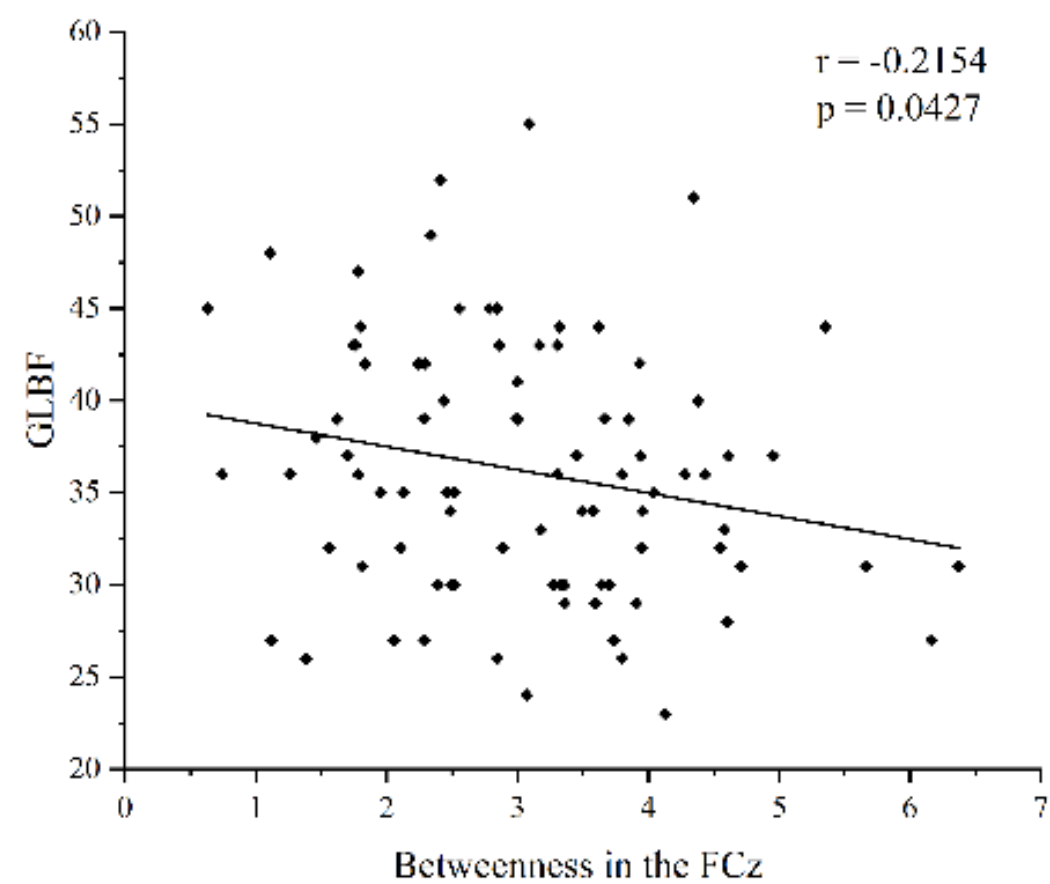
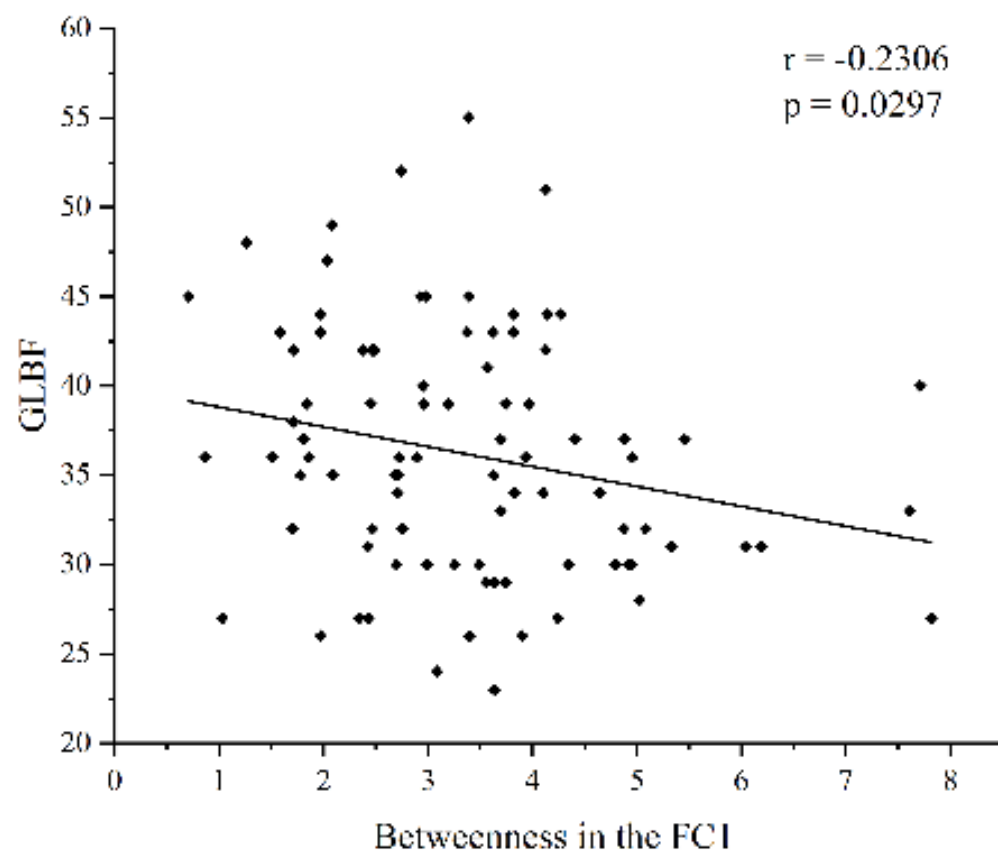
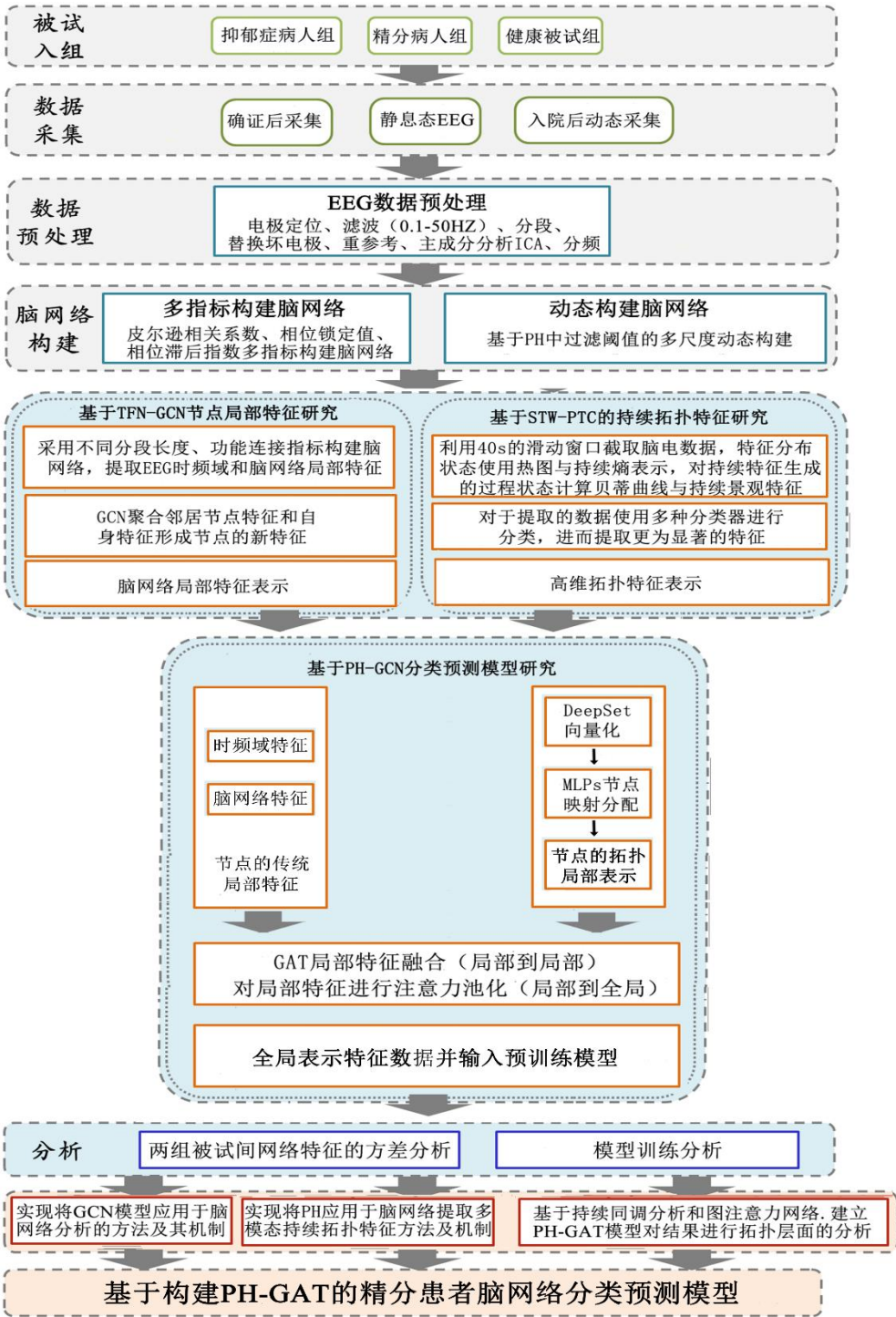


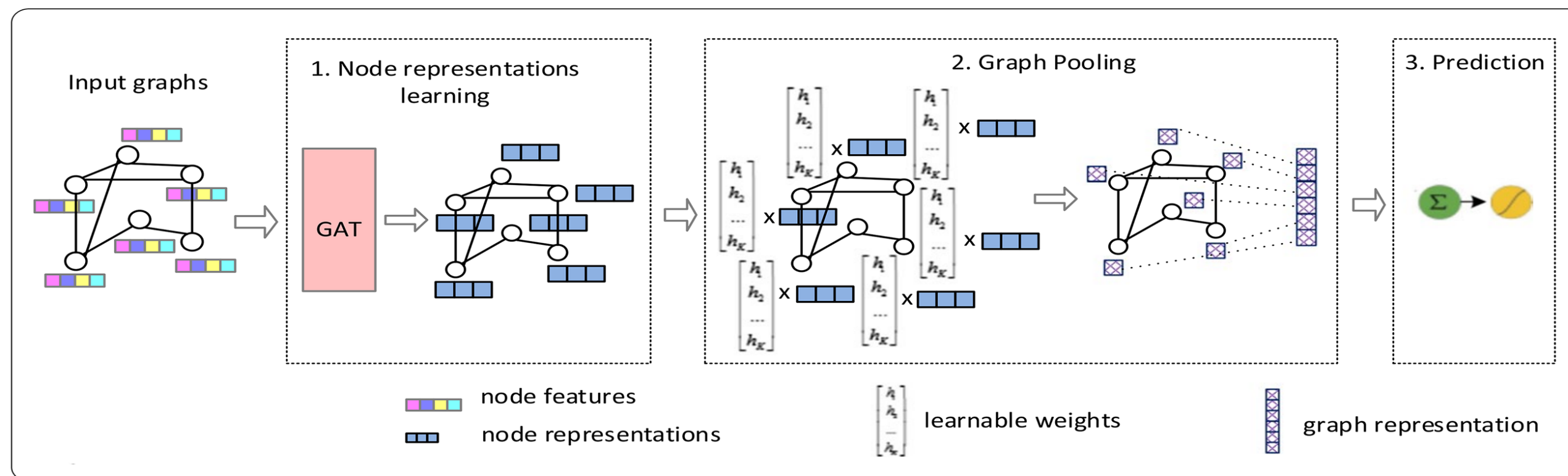
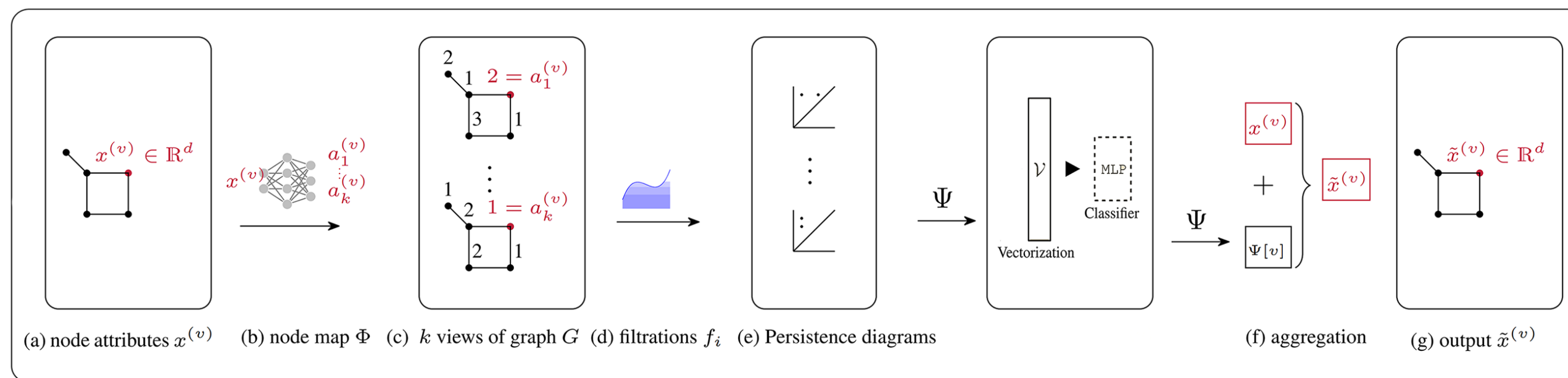
图3 网络拓扑特征与整体分的相关性

应用3： 基于PH-GAT的精神分裂症分类预测模型



更多“人工智能&复杂系统”的知识, 请关注官网: pattern.swarmu.org

PH-GAT模型框架



持续同调集成到GAT模型

采用拓扑数据分析工具来提取隐藏在脑网络复杂结构中的基本信息。模型中的PH模块以具有 V 个顶点 E 条边的图和 d 维节点特征矩阵 $x^v \in R^d$ 作为输入, $v \in V$ 。使用一系列的 k 顶点过滤函数 $f_i: R^d \rightarrow R$, 且 $i=1, \dots, k$ 。每个过滤函数都可以关注图的不同属性。过滤的图像是有限的并产生一组节点值 $a_i^1 < \dots < a_i^n$, 图 G 根据 $\emptyset = G_i^{(0)} \in G_i^{(0)} \in \dots \in G_i^{(n)} = G$, $G_i^{(j)} = (V_i^{(j)}, E_i^{(j)})$, $V_i^{(i)} := \{v \in V \mid f_i(x^{(v)}) \leq a_i^{(j)}\}$, 且 $E_i^{(i)} := \{v, w \in E \mid \max\{f_i(x^{(v)}), f_i(x^{(w)})\} \leq a_i^{(j)}\}$ 进行节点值过滤, 然后形成持续图, 将 k 个持续图分别向量化, 采用MLP映射到每一个节点, 形成节点的拓扑特征, 将形成的节点拓扑特征与基于图论的脑网络局部特征拼接作为新的特征矩阵输入到模型进行学习。

4.2 训练GAT模型

在图学习阶段，提出了一个新的基于持续同调的图注意力网络模型PH-GAT，使用图注意力层来学习节点表示，并使用一个可学习权重的注意力池层来获得用于脑功能网络分类的图表示，最后通过全连接层进行分类预测。

基于点任务的GAT模型

把每一份样本数据（样本数据是780份，一共195人，每个人利用滑动窗口切割成四份独立的数据）都当作一个节点构建邻接矩阵，特征矩阵是780*22的矩阵。最后的分类准确率是了75%。

存在的问题：

- 1、由于是深度学习的模型，数据量过小，导致模型训练时准确率有所抖动。
- 2、由于数据特征只用了高阶拓扑特征，GAT模型只在高阶拓扑特征中增加权重，之后会把传统的脑网络特征一起加入训练GAT模型，效果会更好。
- 3、目前只做了点任务的分类，但是针对数据属性而言，基于图任务的分类会更加适合，之后会考虑基于图任务来对数据样本分类。

联系我们

- 公众号：集智俱乐部
- 公众号：集智学园
- 添加读书会负责人为好友，参与线上讨论



集智学园



集智俱乐部



读书会负责人