

# 多节点规模下的抑郁症功能脑网络分类研究

程 忱 郭 浩 陈俊杰

(太原理工大学计算机科学与技术学院 太原 030024)

**摘 要** 脑网络作为复杂网络分析方法在神经影像领域的应用已得到广泛的认可。研究发现脑网络中的节点规模对网络的拓扑属性会产生重要的影响。利用静息态功能影像数据,在 5 种不同的节点规模下分别完成抑郁症患者和正常对照的脑网络构建,比较了网络拓扑属性的变化,并选择了 4 种不同的分类算法进行脑疾病分类研究。结果表明,网络节点数量不仅对拓扑属性产生了影响,而且对分类模型的构建也有直接作用。支持向量机(RBF 核函数)模型在节点规模为 250 时,分类效果最好,平均正确率为 83.18%。该研究结果在抑郁症的临床诊断中具有重要的应用价值,为基于脑网络的机器学习分类研究在网络节点规模的选择上提供了重要的参考依据。

**关键词** 脑网络,拓扑属性,节点规模,机器学习,抑郁症

**中图分类号** TP18 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.7.048

## Classification of Multi-scale Functional Brain Network in Depression

CHENG Chen GUO Hao CHEN Jun-jie

(College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan 030024, China)

**Abstract** As a complex network analysis method, brain network has been widely accepted in the field of neuroimaging. According to the research, the scale of nodes in the brain has a major impact on the network topological properties. This paper used the resting state functional imaging data to construct brain networks for patients and normal controls respectively under five different node scales and compared variances of the network topological properties, and then selected four different algorithms to do the classification. The results show that the node scale can not only affect the topological properties, but also has a direct effect on the construction of classification model. Support vector machine (RBF kernel function) model shows the best classification results when the node scale is 250, the average accuracy is 83.18%. The research results have an important application value in the clinical diagnosis of depression, and provide a significant reference basis on the network nodes' selection based on machine learning of brain network.

**Keywords** Brain network, Topological properties, Node scale, Machine learning, Depression

脑网络作为复杂网络,在神经影像领域的重要应用已经得到了越来越多研究人员的关注。有研究将复杂网络的基本原理<sup>[1]</sup>以及统计物理学<sup>[2]</sup>等方法应用于拓扑属性的分析来探索网络的基本属性及节点间潜在的拓扑关系。目前,脑网络在脑疾病的应用研究中已经得到了不少令人惊喜的结论,如:精神分裂症<sup>[3]</sup>、阿尔茨海默症<sup>[4]</sup>、癫痫<sup>[5]</sup>、多动症<sup>[6]</sup>、中风<sup>[7]</sup>以及抑郁症<sup>[8]</sup>。文献<sup>[9]</sup>将网络拓扑属性作为分类特征进行机器学习研究,构建辅助诊断模。从上述已有的研究中可以看出,各种神经疾病与脑网络的拓扑属性的异常变化有着密切的联系,对于脑网络的分析应用,其在临床中具有很强的价值,为神经疾病的诊断带来了新的视角。

但到目前为止,仍然有很多问题困扰着我们,其中一个重要的问题便是缺乏一种可以被广泛接受的节点定义方法。不同的节点定义将直接影响全脑网络功能特化的解释。同时,网络节点的规模将对网络的拓扑属性产生不可忽视的影响。在目前的研究中,节点定义方法主要利用先验的解剖模板来

进行脑区的划分,如 AAL<sup>[10]</sup>、ANIMAL<sup>[11]</sup>模板等。这一方法简单有效,但却存在一定的缺陷:1)对于脑区的分割而言,目前尚没有一个可以被广泛接受的黄金标准。在一些特定的区域中,特别是在皮层上,仍然没有明确的边界可以进行区域的分割。2)对于利用解剖模板实现的脑区划分,其所划分的脑区大小的差异很大,从数十个体素到上千个体素。以 AAL 模板为例,双侧额中回共包含了近 10000 个体素,而双侧杏仁核却只包含了 400 个体素。脑区体素数量的巨大差异对脑网络的构建以及拓扑属性,特别是脑区间的关联计算有着巨大的影响<sup>[12]</sup>。3)对于划分体积过大的脑区,其很有可能包含了来自不同功能亚区的信号,这将会导致对于特定脑区的结论变得很难解释。

不同尺度的节点规模对功能连接的计算产生极大影响的同时,对网络的拓扑属性仍然具有决定性的作用<sup>[12]</sup>。选择合理的节点规模是理解网络拓扑属性及相关结论的重要前提。需要注意的是,之前的研究针对的均为正常被试。而对于脑

到稿日期:2015-06-03 返修日期:2015-09-25 本文受国家自然科学基金项目(61170136,61373101,61472270,61402318),太原理工大学青年团队启动基金项目(2013T047)资助。

程 忱(1989—),女,博士生,主要研究方向为智能信息处理、脑信息学、脑影像学;郭 浩(1981—),男,博士,副教授,主要研究方向为智能信息处理、脑信息学;陈俊杰(1956—),男,教授,博士生导师,主要研究方向为数据库与智能信息处理,E-mail:chenjunjie\_tyut@sina.com(通信作者)。

疾病患者而言,其脑网络拓扑属性是否也会受到同样的影响?同时,在进行机器学习应用时,网络节点规模是否会对分类结果产生影响?这些问题据笔者所知,尚没有明确结论。

本文假设:在脑疾病状态下网络拓扑属性同样受到节点规模的影响;同时,不同的节点规模下所构建的分类模型的效果存在差异。为证明此假设,选择抑郁症作为疾病模型,计算了常见网络拓扑属性,包括最大连通组件、聚合系数、特征路径长度以及小世界属性等,并进行了定量分析。此外,利用4种机器学习方法,分别进行了5个节点规模下的分类模型构建,并比较分析了分类效果。本文为今后在进行脑网络指标分析特别是机器学习研究时,对网络节点规模的选择提供了重要的参考依据。

## 1 实验材料及方法

### 1.1 数据采集和预处理

研究中,共采集38名抑郁症患者及28名正常对照的磁共振数据。抑郁患者均为由山西医科大学诊断确定为首发未用药患者,记录患者的病程及严重程度,通过汉密尔顿量表确定。正常对照的筛选标准为没有精神病史(通过人格障碍定势访确定)。所有被试在数据采集前均签署了知情协议书(抑郁症患者由家属签署,正常对照由本人签署)。

实验数据的采集工作由山西医科大学放射科医师完成。数据采集过程中,被试要求闭眼并保持清醒。扫描参数具体为:33层,射频重复时间(TR)为2000ms,回波时间(TE)为30ms,皮层厚度4mm,视野范围(FOV)为192mm×192mm,翻转角90°。共采集248个时间点。

首先使用SPM8 (<http://www.fil.ion.ucl.ac.uk/spm>)对数据进行预处理。将采集到的被试的数据进行时间片校正和头动校正。其中有2例抑郁组和1例对照组数据由于头动大于3mm或转动大于3°而在校正过程中被弃除。之后,图像进行12维度的优化仿射变换,将其标准化到3mm体素的MNI标准空间中。最后进行低频滤波(0.06~0.11Hz)以降低低频漂移及高频的生物噪音。

### 1.2 功能脑网络构建

为了定义5种不同的节点规模,本文在AAL模板基础上提出了基于k-means聚类思想的节点定义方法。具体方法如下:首先,设置 $N$ 为预期的节点数。在研究中,尝试了250,500,1000,1500等4个节点数,加上原有AAL模板90节点的模板,共得到5个不同的节点规模模板。接下来,计算得到原有AAL模板中每个脑区占有所有脑区的体素比例 $V$ 。然后,根据预设的节点数 $N$ ,可计算得到AAL<sup>[10]</sup>原有脑区可细化的子区域的个数 $k=VN$ ,即脑区BR在 $N$ 的节点规模下应该细分为 $k$ 个子区域。此后,令脑区BR设置 $k$ 个随机种子体素 $S=s_1, s_2, s_3, \dots, s_k$ ,然后依次计算所有剩余体素与 $S_i$ 种子体素的距离。在此,课题组采用动态随机种子点的设置方法,即在计算距离之后,当前体素 $V$ 与距离最近的体素 $S_i$ 组合形成新的子区域,并且将 $V$ 和 $S_i$ 的物理中心设置为生成的新的种子体素。依次循环,直至脑区内所有体素均划分完成。至此,脑区BR便完成了 $k$ 的划分。当所有脑区完成划分后,即可得到预期节点规模 $N$ 下的脑区划分。

此后,在所有5个尺度下,计算每个节点的平均时间序

列,并利用多元线性回归去除头动、脑积液、白质对信号的影响。对于所得到的残差,利用皮尔逊线性相关算法对节点进行两两关联分析,得到 $N \times N$ 的相关矩阵 $R$ (见图1(a)), $N$ 为节点数量。然后根据预先设定的阈值 $\tau$ ,将每个相关矩阵转化为二值邻接矩阵 $A$ ,其中矩阵中的元素 $a_{ij}$ 的值为1或者0。当节点 $i$ 和节点 $j$ 的相关系数大于阈值 $\tau$ 时, $a_{ij}$ 的值等于1,否则 $a_{ij}$ 的值为0。

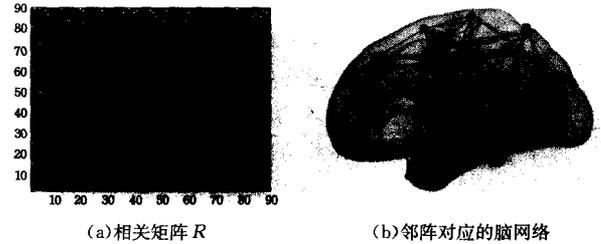


图1 38例正常对照在90节点下构建的功能脑网络

研究中,采用稀疏度 $S$ 来进行阈值设置。 $S$ 的定义为网络中实际存在的边数与网络中可能存在的最大边数的比值。由于先前的研究发现脑网络属于典型的低消耗网络,并表现出典型的小世界属性,因此利用小世界属性这一脑网络基本特性来进行阈值空间的设定,这在保证小世界属性的同时,最大程度地去掉伪连接。其具体规则为:

所有节点的平均度(节点的度是指所有连接到该节点的边的数量)大于 $2 \times \ln(N)$ 。此时 $N$ 为网络中的节点数量。

小世界标量 $\sigma > 1.1$ (小世界标量的定义见下节)。

由于研究中采用了5种不同的节点规模定义,其节点数各不相同,因此根据上述原则所得的阈值空间在5种节点规模下并不完全一致。为了在统一的空间内进行比较,以90个节点下的阈值空间 $S \in (8\%, 32\%)$ 为标准,并且在该阈值空间内,以步长为0.1的所有稀疏度构建出了所有被试的脑功能网络。

## 2 网络指标计算及分类模型构建

### 2.1 网络指标

大量的网络指标可以从不同层面刻画网络的拓扑属性<sup>[14]</sup>。研究中,选择了几种较为基础的指标:最大连通组件、特征路径长度、聚合系数以及小世界标量。

最大连通组件 $C$ 表征了网络的连通程度,通过最大组件个数与网络连接的比重来衡量网络连通的优劣。随着网络连通性的增加,网络中最大组件的个数趋于网络节点数。特征路径长度 $L_p$ 是网络中任意两点间最短路径长度的平均值,是表征网络整体路由效率的指标。网络的特征路径长度越低,意味着在任意两点间进行网络传递时所经过的路径越少,网络的全局效率越高<sup>[15]</sup>。特别地,有部分网络是不连通的,意味着在这个网络中会有两个节点间不存在连通路径。为了解决这个问题,研究中利用调和平均数来计算网络的特征路径长度。聚合系数 $C_p$ 定义为给定节点的直接邻居间实际存在的边数与其理论上可能存在的最大边数之间的比值,而网络中所有节点聚合系数的平均值则定义为网络的聚合系数。聚合系数反映了局部的连通性或者对于给定节点的邻居的社团性。网络的聚合系数越高,意味着网络局部的连通性越好,其局部效率越高。小世界标量 $\sigma$ 的计算涉及到两个指标,即

标准化特征路径长度  $\lambda$  以及标准化聚合系数  $\gamma$ 。这两个指标需要将网络的特征路径长度与聚合系数同对应的随机网络指标进行标准化计算 ( $\gamma = C_P / C_P^{rand}$ ,  $\lambda = L_P / L_P^{rand}$ )。典型地一个小世界网络需要满足条件  $\gamma > 1$  且  $\lambda \approx 1$ <sup>[16]</sup>, 因此小世界指标  $\sigma = \gamma / \lambda$  将会大于 1<sup>[17]</sup>。

## 2.2 分类模型构建

本文利用机器学习方法<sup>[18]</sup>来实现对疾病数据的建模及自动识别, 选取了 6 个网络属性作为特征, 包括最大连通组件、特征路径长度、聚合系数、标准化特征路径长度、标准化聚合系数以及小世界标量。选择 SVM(RBF 核函数和多项式核函数)、神经网络以及决策树 C4.5 等算法来构建不同的分类模型。利用交叉验证方法, 生成并评测模型。从样本中随机选择 70% 作为训练集, 剩余的 30% 作为测试集, 在每个阈值下分别进行 100 次, 然后计算正确率的算术平均值。利用交叉验证方法来对模型进行验证, 其已经广泛地应用于同类研究中(综述见文献[19])。

## 3 结果与讨论

在对网络指标进行组间比较时发现, 无论是正常组还是抑郁组, 不同节点规模下网络拓扑属性均表现出相似的变化规律, 包括最大连通组件、全局属性以及小世界标量等。这一结论说明, 节点规模对正常组和抑郁组的脑网络拓扑属性具有相同的影响。网络连通性的分析结果表明, 随着稀疏度的增加, 网络的连通性逐渐增强。在 30% 的稀疏度下, 所有 5 个节点规模的连通性均在 97% 以上。同时, 在相同的稀疏度下, 网络节点规模越大, 其网络连通性越好(见图 2(a))。而对网络全局属性的分析结果表明, 正常组和抑郁组表现趋势一致。随着稀疏度的增加, 网络聚合系数也随之增加, 意味着局部效率的增高。同时, 在同一稀疏度下, 节点规模越大, 其聚合系数越高, 意味着局部效率越高(见图 2(b))。而在特征路径长度中, 则表现出相反的结论。随着稀疏度的增加, 网络特征路径长度随之降低, 意味着全局效率的增高。同时, 在同一稀疏度下, 节点规模越大, 其特征路径长度越高, 意味着全局效率越高(见图 2(c))。通过分析发现, 同一稀疏度下节点规模越大, 其网络效率越高。随着稀疏度的增加, 不同节点规模的全局属性将趋于一致, 意味着其网络效率的差距逐步减少。特别地, 当稀疏度大于 25% 时, 除 90 节点模板外, 其他模板的特征路径长度基本重合。这一结果暗示着节点尺度规模对于网络指标的影响不是无限放大的, 随着网络中连接数量的增加, 其影响是逐步减弱的。同时, 在所选的阈值空间内, 相比抑郁组或者对照组, 其均表现出明显的小世界属性(见图 2(d))。具体表现为: 标准化的聚合系数大于 1, 且标准化特征路径长度近似于 1。上述结论与已有功能脑网络相关研究结论一致<sup>[20]</sup>。在相同的稀疏度下, 较大的尺度规模定义表现出更高的小世界标量。而随着稀疏度的增加, 其之间的差异逐步缩小。

本文选择了 4 种不同的分类模型: 支持向量机(RBF 核函数和多项式核函数)、神经网络以及决策树 C4.5。结果显示, 在所有 5 个节点规模下, 支持向量机(RBF 核函数)与神经网络模型均表现出较好效果。这一结果说明, 非线性分类器较线性分类器有一定优势, 这一点与已有结论不尽相同。

Cox<sup>[21]</sup>比较了线性分类器(SVM-线性和 LDA)算法和非线性分类器(SVM-多项式)在视觉刺激任务中的性能差异, 结果显示非线性分类器并没有提供更好的效果。这一结果说明, 与传统的利用体素 BOLD 作为特征的分类模型相比较, 以复杂网络指标作为特征的分类模型在性能上与前者有一定的差异。这表明不同的特征选择方法对模型性能的影响很大。当从脑网络角度进行分类模型构建时, 应该重新评价所选的模型构建算法。4 种分类器下的分类效果如图 3 所示。

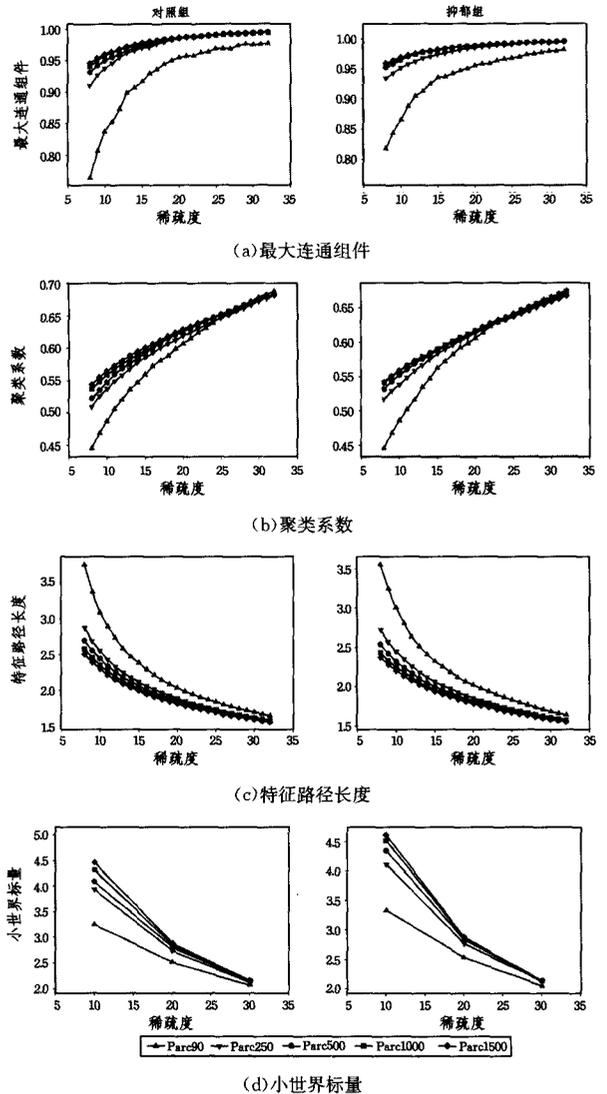


图 2 不同节点规模下网络指标的变化

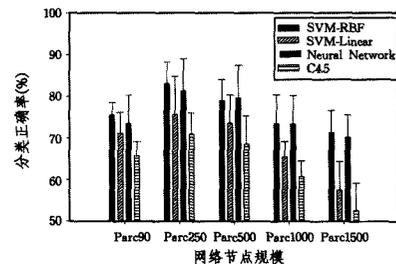


图 3 4 种分类器下的分类效果

此外分析发现, 在 5 个节点规模的横向比较中, 当节点数量为 250 个时, 分类正确率最高; 节点个数为 90 个时略低。而随着节点规模的增大, 分类正确率有明显的下降。这一

(下转第 284 页)

[5] Tomar V, Asnani H, Karandikar A, et al. Social network analysis of the Short Message Service[C]// 2010 National Conference on Communications. 2010: 1-5

[6] Ma Nan, Guan Jian-cheng, Zhao Yi. Bringing pagerank to the citation analysis [J]. Information Processing & Management, 2008, 44(2): 800-810

[7] Ding Ying, Yan E, Frazho A, et al. Pagerank for ranking authors in co-citation networks [J]. Journal of the American Society for

[8] Franceschet M. Pagerank: standing on the shoulders of giants [J]. Communications of the ACM, 2011, 54(6): 92-101

[9] Li Yong-li, Wu Chong, et al. A Tree-network Model for Mining Short Message Services Seed Users and Its Empirical Analysis [J]. Chinese Journal of Management Science, 2012(20): 49-54 (in Chinese)

李永立, 吴冲, 等. 基于树形网络分析的短信种子用户挖掘模型及其实证分析[J]. 中国管理科学, 2012(20): 49-54

(上接第 267 页)

象在所选择的 4 个分类模型中均有明显表现。这一结果暗示着网络节点的数量对分类模型的构建具有直接的影响。据分析,在大节点规模模板中,由于节点数量的增加,每个节点的体积(体素数量)减少,这将意味着所采集的功能影像数据更容易受到噪音信号的影响。而小节点规模模板能够更好地中和噪音,但节点体积的增大会造成脑区功能特化的模糊。所以,在节点规模的选择上,必须要考虑到二者的平衡,才能更好地表达脑网络基本拓扑结构。

**结束语** 本研究利用了不同节点规模定义的模板,对正常组及抑郁组进行了静息态功能脑网络拓扑属性的分析及对比,并进行了机器学习研究。结果发现,无论正常组还是抑郁组,不同节点规模下网络拓扑属性均表现出相似的变化规律。这一结论说明,节点规模对正常组和抑郁组脑网络拓扑属性具有相同的影响。同时,不同节点规模下的分类模型效果差异很大,当节点数目为 250 时,各分类器均表现出最好的效果。上述结论提示在进行脑网络指标分析特别是机器学习研究时,需要考虑到所采用的节点规模的影响。

### 参 考 文 献

[1] ERDdSP, Wi A. On random graphs [J]. I. Publ. Math. Debrecen, 1959, 6: 290-297

[2] Albert R, Barabási A L. Statistical mechanics of complex networks[J]. Reviews of Modern Physics, 2002, 26(1)

[3] Lynall M E, Bassett D S, Kerwin R, et al. Functional connectivity and brain networks in schizophrenia[J]. The Journal of Neuroscience, 2010, 30(28): 9477-9487

[4] Stam C. Use of magnetoencephalography (MEG) to study functional brain networks in neurodegenerative disorders[J]. Journal of the Neurological Sciences, 2010, 289(1): 128-134

[5] Horstmann M T, Bialonski S, Noenning N, et al. State dependent properties of epileptic brain networks; Comparative graph-theoretical analyses of simultaneously recorded EEG and MEG[J]. Clinical Neurophysiology, 2010, 121(2): 172-185

[6] Liang Wang, Zhu Chao-zhe, He Yong, et al. Altered small-world brain functional networks in children with attention-deficit/hyperactivity disorder[J]. Human Brain Mapping, 2009, 30(2): 638-649

[7] De Vico Fallani F, Laura A, Febo C, et al. Evaluation of the brain network organization from EEG signals; a preliminary evidence in stroke patient[J]. The Anatomical Record, 2009, 292(12): 2023-2031

[8] Guo H, C C, Cao Xiao-hua, et al. Resting-state functional connectivity abnormalities in first-onset unmedicated depression [J]. Neural Regen Res, 2014, 9(2): 153-163

[9] Guo H, Cao X H, Liu Z F, et al. Machine learning classifier using abnormal brain network topological metrics in major depressive disorder[J]. Neuroreport, 2012, 23(17): 1006-1011

[10] Tzourio-Mazoyer N, Landeau B, Papathanassiou D, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain[J]. Neuroimage, 2002, 15(1): 273-289

[11] Collins D L, Holmes C J, Deters T M, et al. Automatic 3-D model-based neuroanatomical segmentation[J]. Human Brain Mapping, 1995, 3(3): 190-208

[12] Salvador R, Martinez A, Pomaral-Clotet E, et al. A simple view of the brain through a frequency-specific functional connectivity measure[J]. Neuroimage, 2008, 39(1): 279-289

[13] Williams J B. A structured interview guide for the Hamilton Depression Rating Scale[J]. Archives of General Psychiatry, 1988, 45(8): 742-747

[14] Rubinov M, Sporns O. Complex network measures of brain connectivity: uses and interpretations[J]. Neuroimage, 2010, 52(3): 1059-1069

[15] Latora V, Marchiori M. Efficient behavior of small-world networks[J]. Physical Review Letters, 2001, 87(19): 198701

[16] Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks[J]. Nature, 1998, 393(6684): 440-442

[17] Humphries M D, Gurney K, Prescott T J. The brainstem reticular formation is a small-world, not scale-free, network[J]. Proceedings of the Royal Society B: Biological Sciences, 2006, 273(1585): 503-511

[18] Guo Li-li, Ding Shi-fei. Research Progress on Deep Learning[J]. Computer Science, 2015, 42(5): 28-33 (in Chinese)

郭丽丽, 丁世飞. 深度学习研究进展[J]. 计算机科学, 2015, 42(5): 28-33

[19] Pereira F, Mitchell T, Botvinick M. Machine learning classifiers and fMRI: a tutorial overview[J]. Neuroimage, 2008, 45(1 Suppl): S199-209

[20] Bullmore E, Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems[J]. Nature Reviews Neuroscience, 2009, 10(3): 186-198

[21] Cox D D, Savoy R L. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex[J]. Neuroimage, 2003, 19(2): 261-270