

复杂网络节点相似性算法及其在癫痫病辅助诊断的应用^①

何 艳¹, 赵晓婷¹, 于云莉²

¹(贵州医科大学 生物与工程学院, 贵阳 550004)

²(贵州医科大学附属医院 神经内科, 贵阳 550004)

摘 要: 节点相似性分析是链路预测和社团挖掘中的重要部分. 引入 CN(Common Neighbor, 共同邻居)算法、RA(Resource Allocation, 资源分配)算法、AA(Adamic-Adar)算法、Sorenson 算法等四种节点相似性算法作用于真实网络以及仿真网络(即小世界网络和无标度网络)网络, 计算 AUC(Area Under the Curve, 曲线下面积)曲线从而比较算法的预测准确性, 结果表明 RA 算法的预测准确性优于其他三种算法. 随后将四种算法用于分析 8 例全身性癫痫患者脑电数据功能连接网络, 结果发现 RA 算法预测准确性最佳, 通过 RA 算法能确定最大节点相似度组成的节点簇, 为量化大脑功能状态提供客观指标, 未来可以将该方法用于临床辅助诊断.

关键词: 复杂网络; 节点相似性; 链路预测; 癫痫

Node Similarity Algorithm on Complex Network and Its Application in Epilepsy Auxary Diagnosis

HE Yan¹, ZHAO Xiao-Ting¹, YU Yun-Li²

¹(School of Biology & Engineering, Guizhou Medical University, Guiyang 550004, China)

²(Department of Neurology, Affiliated Hospital of Guizhou Medical University, Guiyang 550004, China)

Abstract: The investigation of node similarity is an important component in link prediction and community detection. In this paper, four kinds of algorithms including common neighbor (CN), resource allocation (RA), Adamic-Adar (AA) and Sorenson are introduced into various kinds of real networks and two kinds of simulation networks comprised of small world network and scale free network. The Area Under the Curve (AUC) is computed to compare their predictive accuracy. It's found that RA performs much better than the other three kinds of algorithms. Then four algorithms are adopted in functional connectivity networks that characterize electroencephalograph (EEG) recordings from eight patients with generalized epilepsy. It's demonstrated that RA performs best from the point of prediction accuracy. According to RA technique, clusters could be determined from nodes that own maximum similarity which provides an objective index for quantifying brain condition, and this might be applied for clinical auxiliary diagnosis in the future.

Key words: complex network; node similarity; link prediction; epilepsy

大脑是一个复杂系统, 其信息传递和认知功能实现依赖于大脑中的神经电活动. 已知大脑电活动其时空分辨率随记录手段而变化, 但总体呈现出高维度和非平稳的特点. 多通道电极记录能尽可能地捕捉大脑电活动的空间分布特征. 在癫痫疾病中, 患者癫痫发作时往往伴随着神经集群的阵发性放电异常. 随着癫痫发作次数的增加, 癫痫患者大脑认知功能也会受到

损伤, 可能表现出学习记忆能力减退或衰弱. 目前国内约 1000 万癫痫患者, 一般假设癫痫患者大脑回路特性发生改变, 进而引发功能异常, 临床上通过长程脑电检测和颅内脑电记录信号对癫痫发作相关的病灶区域进行定位. 由于癫痫疾病病理机制异常复杂, 多种神经递质和神经受体参与神经电活动, 典型的癫痫发作如何对神经回路发生影响目前还没有确切的结论; 此

① 基金项目: 国家自然科学基金(81460206); 贵州医科大学博士启动基金(J2014[003])

收稿时间: 2016-04-18; 收到修改稿时间: 2016-06-01 [doi:10.15888/j.cnki.csa.005554]

外,已知大脑静息态功能网络对维持大脑反应临界性从而对外界刺激产生快速响应具有重要作用,通过研究大脑静息态功能网络连接特征从而对大脑状态进行客观评价将有利于临床辅助诊断与疗效估计。

复杂网络是一种简单高效的描述方法,网络节点为记录电极位点,网络连接边为不同大脑区域的相互作用估计,基于复杂网络的大脑神经电活动分析方法能从系统角度量化大脑功能状态。

复杂网络是刻画复杂系统的有效方法,从1998年Watts和Strogatz提出小世界网络的概念以后,复杂网络受到学者的广泛关注^[1],比如通讯网络、互联网、交通运输网等,还有科学家合作网、微信朋友圈等社会网络。其中团簇发现以及链路预测^[2,3]是复杂网络研究的重要组成部分。链路预测融合网络节点及网络结构信息,通过对复杂网络设置不同的特征估计,如随机分块模型、相似性预测等预测网络中的缺失边并识别错误链接,有助于准确高效地检测出癫痫患者大脑复杂网络的异常通路。节点相似性于1998年由Lin提出,可以通过节点属性的相似性对网络链路进行预测,如果两节点的相似度越大,则说明这两个节点可能存在某种链接的可能性^[4,5]也就越大。Lin通过节点的属性定义节点间相似性并以此来对链路进行预测。Fouss、Pirrotte提出基于随机游走的协同推荐^[6],Liben-Nowell、Kleinberg等人在对科学家网络进行分析,发现网络结构同时对预测有着一定影响,并提出新的相似性定义^[7]及基于网络拓扑的预测方法优于随机的预测方法。局部路径指数、全局随机游走指数、协同过滤预测方法均取得了较好的链路预测效果。2008年,Newman在《自然》上的论文中又提出了基于网络层次结构的预测方法^[8]。节点和边是构成复杂网络的两个重要因素,网络结构特征和网络性质是复杂网络研究的重要组成部分^[9-13]。节点度的大小表示此节点在网络中的地位的重要性,两节点之间的联系可以用节点间的最短距离描述。

在传统计算节点相似性算法的过程中主要考虑节点的共同邻居和节点的度,通过这些算法对两节点相似度的计算,判断这两个节点之间存在链接的可能性大小。在基于节点相似性的预测算法中,如果两节点的相似值越大,则被认为两节点存在链接的可能性就越大。比如,现在大多数人都在用的社交工具QQ,如果两个陌生人之间存在的共同好友越多,那么他们就

被认为相似,则他(她)们成为朋友的几率也越大。基于节点属性(主要为共同邻居及共同邻居的节点度)的节点相似性方法计算复杂度小,公式简单,速度快,如CN算法;AA算法又被称为频率加权共同邻居算法,是将不寻常的特征赋予更多权重,即稀有特征蕴含更多信息,该算法假设一个朋友较少的人可能更倾向于将他的一对朋友相互介绍;Sorensen指标除了考虑共同邻居的规模,同时假设低节点度的共同邻居节点有更高连接可能性;RA算法来自网络中的资源配置理论^[13],节点对之间可以通过共同邻居传递信息,假设每个传递者均有一个资源单元且在所有邻居中平均分配,节点对之间传递并被接收的资源数量被定义为节点对之间的相似性,与此同时RA算法考虑所有排序结果,并引入次临近邻居信息度量节点节点相似性。不同的预测算法的表现与网络结构信息相关,本文将引入CN、RA、AA、Sorensen等四种相似性算法作用于真实复杂网络、仿真网络及癫痫患者脑电复杂网络,为揭示癫痫患者大脑功能连接异常通路提供技术支持。

1 节点相似性算法描述

CN算法是基于局部信息中最简单的相似性算法,即为共同邻居算法(Common Neighbors)的英文缩写。表示的是如果两个节点的共同邻居数越多,则这两个节点越相似。共同邻居CN算法定义为:对于网络中的节点 m ,定义 m 的邻居集合为 $\varphi(m)$,那么两个节点 x 和 y 的节点相似性 S_{xy} 就定义为它们的共同邻居数量,公式即为:

$$S_{xy} = \varphi(x) \cap \varphi(y) \quad (1)$$

以下举一个简单的例子来理解CN算法,网络结构如图1所示。

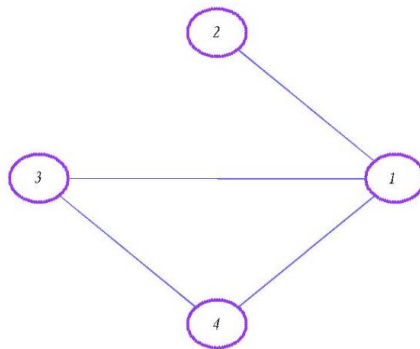


图1 网络结构示意图

图 1 中总共包含 4 个节点, 根据网络的连接情况, 我们可以得到节点 2 和节点 4 的共同邻居只有 1 个, 那就是节点 1, 所以得到节点 2 和节点 4 的相似性为 1, 即为 $S_{\{2,4\}}=1$.

与 CN 算法不同, RA 算法考虑的是共同邻居节点的度, 而不是共同邻居. 由于受到资源分配的启发, 周涛等人提出了资源分配这一相似性指标^[13]. 网络中没有直接相连的两个节点, 从节点 v_x 到节点 v_y , 这一传递过程中作为媒介的就是共同邻居, 通过共同邻居这一“中间人”, 可以从 v_x 中传递一些资源到 v_y . 在算法中, 每个媒介都有一个单位的资源并且将其平均分配传给它的邻居.

其节点相似性定义为:

$$S_{xy} = \sum_{z \in \varphi(x) \cap \varphi(y)} \frac{1}{k(z)} \quad (2)$$

上述公式中, $\varphi(x) \cap \varphi(y)$ 表示的是节点 v_x 与节点 v_y 的共同邻居的集合, $k(z)$ 代表两个节点共同邻居的节点的度. 如图 2 所示, 如果要用 RA 算法计算节点的相似性时, 我们需要考虑的是节点相似度的影响. 由图可知节点 3 和节点 5 的共同邻居为节点 1 和节点 2, 节点 1 的度为 4, 节点 2 的度为 3, 可以得到节点 3 和节点 5 的相似度为 $S_{\{3,5\}} = \frac{1}{4} + \frac{1}{3} = \frac{7}{12} \approx 0.583$. 同理可得, 节点 1 为节点 3 和节点 4 的共同邻居, 那么节点 3 与节点 4 的相似度则为 $S_{\{3,4\}} = \frac{1}{4}$, 两个相似性值相比较, 节点 3 与节点 5 的相似值明显比节点 3 和节点 4 的相似值要大, 也即说明节点 3 和节点 5 更有可能产生连接.

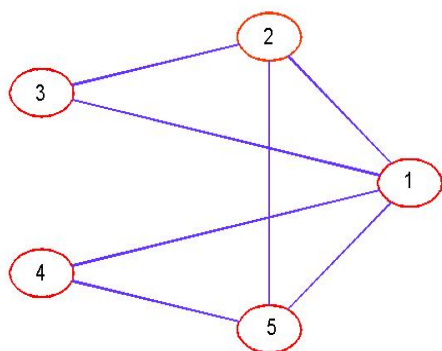


图 2 示例网络图

在节点相似性的算法中, 考虑节点共同邻居的节

点的度的影响, 著名的有 Adamic-Adar 算法, 即是 AA 算法, 它的主要思想为度小的节点的贡献度大于度大的节点的贡献度. 比如在微博中关注度非常高的人, 他们要么是某些领域的专家, 要么是明星, 因此关注他们的那些人很有可能并没有什么相同的兴趣爱好, 共同点; 相反, 如果是一位粉丝很少的人, 当有两个人同时关注了他, 那则说明这两个人很有可能具有相同的爱好, 共同的志趣. 在网络中则表现为这两个节点存在某种链接的可能性很大.

对于 AA 算法的定义为: 对于网络中节点 m , 定义 m 的度为 $k(z) = \varphi(z)$, 该算法是在共同邻居算法的基础上赋与其权重, 节点 x 和 y 的节点相似性定义为:

$$S_{xy} = \sum_{z \in \varphi(x) \cap \varphi(y)} \frac{1}{\log k(z)} \quad (3)$$

用 AA 算法计算图 2 所示网络节点相似性, 同样计算节点 3 与节点 5 的相似值, 结果如下:

$$S_{\{3,5\}} = \sum_{z \in \varphi(x) \cap \varphi(y)} \frac{1}{\log k(z)} = \frac{1}{\log 4} + \frac{1}{\log 3} \approx 3.757$$

AA 算法和 RA 算法都是赋予共同邻居节点权重, 而它们的最大的区别在于, 赋予共同邻居节点权重的方式不同, AA 算法是以 $\frac{1}{\log k(z)}$ 的形式递减, RA 算法

以 $\frac{1}{k(z)}$ 形式递减.

Sorenson 算法与前三种节点相似性算法相比, 它不仅受共同邻居的影响, 同时也受到节点的度的影响. 此算法一般用于生态学数据的研究. 其相似性的值的计算方法为共同邻居数目的两倍与这两个节点的度之和的比值, 定义为:

$$S_{xy} = \frac{2 \times \varphi(x) \cap \varphi(y)}{k(x) + k(y)} \quad (4)$$

式中, $k(x)$ 、 $k(y)$ 表示的是两个节点的度, $\varphi(x) \cap \varphi(y)$ 表示的是两节点的共同邻居. 同样列举一个实例, 加深我们的理解. 根据图 2, 得到节点 3 与节点 5 的相似值为:

$$S_{\{3,5\}} = \frac{2 \times \varphi(x) \cap \varphi(y)}{k(x) + k(y)} = \frac{2 \times 2}{2 + 3} = \frac{4}{5}$$

2 节点相似性分析结果

2.1 数据预处理

首先将网络随机划分成训练集和测试集. 即将一

个完整的网络划分成两个部分, 其中一部分作为训练集, 剩余部分为测试集, 比较 CN、RA、AA、Sorenson 四种算法的预测效果, 通过测试集得到预测能力的结果. 针对一个无向网络 $G(V, E)$, 给定一种预测算法并运用这种算法为每对没有连边的节点赋予一个值, 所赋予的这个值与这两个节点的连接概率正相关. 然后将所有没有连接的节点的对按照值从大到小进行排列, 排在前面的说明两个节点相互连接的可能性越大. 若给出一个完整网络, 此网络包含有 13 个节点, 19 条边, 该网络的全集为 78 条边, 说明有 59 条边是不存在的. 随机选出这 19 条边中的 4 条作为测试集, 剩下的 15 条边作为训练集. 给出某种预测算法, 此算法赋予 63 条未知边一个值(包括 4 条测试边跟 59 条不存在的边), 然后将这 63 条边的分数值按照从大到小的顺序进行排序, 如果能使 4 条测试边尽可能多的排列在前面, 则说明算法的预测准确性越高.

引入 AUC 作为衡量预测准确性的指标, AUC 可以理解成在测试集中随机选取一条边的分数值比随机从不存在的边选取一条边的分数值要高的概率. 其中, 测试边与不存在的边的集合为未知边. 每次从测试边中随机抽取一条边, 然后从不存在的边中随机抽取一条边, 比较两个边的分数值, 如果测试边的分数值大于不存在的边的分数值, 那么就加 1 分, 要是这两个分数值相等就加 0.5 分. AUC 定义如下:

$$AUC = \frac{n' + 0.5n''}{n} \quad (5)$$

式中, n 表示的是独立比较的次数; n' 表示的是测试边的分数大于不存在的边的次数, n'' 表示的是测试边的分数与不存在的边的分数相等的次数. AUC 值越大, 预测的准确性就越高.

2.2 真实网络分析结果

四个真实网络分别为线虫代谢网络(Metabolic), 科学家合作网(NetScience)、爵士音乐家合作网(Jazz)以及美国航空网络(USAir), 其基本统计特征如表 1 所示. 分别取不同的训练集比例, 运用 CN、RA、AA、Sorenson 四种相似性算法, 计算出每个训练集比例对应的 AUC 值, 绘制曲线图.

表 1 真实网络基本统计特征

网络	边数	节点数
NetScience	2742	1536
Metabolic	4596	453

USAir	2126	1532
Jazz	5484	198

由于训练集的取值范围受到网络的限制, 不同网络训练集的取值范围不同. 通过对线虫的新陈代谢网络数据的处理, 得到 CN、RA、AA、Sorenson 四种相似性算法在线虫的代谢网络中 AUC 曲线图 3 所示.

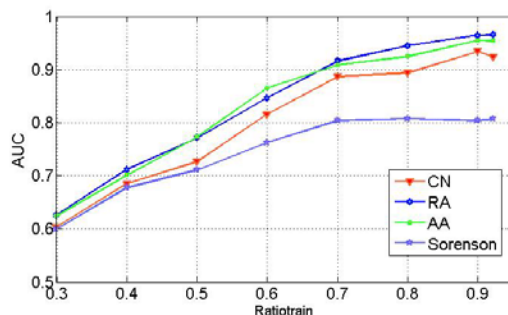


图 3 线虫代谢网络链路预测的 AUC 曲线图

由此可以看出, 在线虫的代谢网络中, CN、RA、AA、Sorenson 四种相似性算法的预测效果还是存在比较明显的差异的, AUC 的值越大说明预测准确性越高, RA 算法优于其他算法.

利用科学家合作网(NetScience)、爵士音乐家合作网(Jazz)、美国航空网络(USAir)三个真实网络检验上述四种算法, 三个网络的性质基本都不一样, 其中, 科学家合作网是一个无向加权网络, 而爵士音乐家合作网是一个无向无权的网络. 依据 AUC 曲线图的变化趋势, 随着训练集样本比例的上升, 结果发现当曲线趋于平缓时, RA 算法预测精确度明显高于其他三种算法.

2.3 仿真网络分析结果

由于国际标准 10-20 脑电信号采集系统一般为 21 通道左右, 网络中节点和连边数目较小, 因此引入两个仿真网络, 其中一个为 WS 小世界网络、另外一个为 BA 无标度网络, 节点的数目分别为 20 和 30, 根据 AUC 曲线图分析, CN、AA 算法的预测精度基本相同, 最佳的仍然是 RA 算法. 不管是在复杂的真实网络中还是在仿真的只有 30 个节点的 BA 网络(如图 4 所示)中, RA 算法都比其他三种算法具有优势.

2.4 癫痫患者大脑功能连接网络分析结果

癫痫是一种常见神经系统疾病, 临床认为大脑神经元异常放电导致癫痫发作. 给患者生活带来极大不便, 严重的甚至影响到患者的心理健康, 多通道脑电

信号是临床癫痫疾病筛查和诊断的重要手段. 本文中的脑电数据来自于 8 例全身性癫痫患者, 其男/女比例为 5/3, 年龄为 8.50 ± 4.01 , 脑电采集设备为 Nicolet 长程视频脑电采集系统. 原始脑电信号记录时间为 3 小时至 24 小时不等, 其不同频率波段脑电信号组成的复杂网络提取来自于其相位锁定值^[14]. 以 α (8-13Hz) 波段脑电信号构成的复杂网络为研究对象, 结果发现 RA 算法优于其他 3 种算法, 如图 5 所示. 在算法检验中, 由于对癫痫患者大脑功能性连接矩阵进行了稀疏化处理^[15], 网络节点及连接边数目较少, 因此测试集比例需要在 70% 及以上.

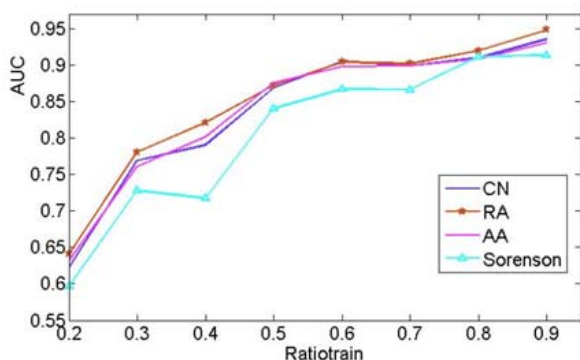


图 4 BA 网络 AUC 曲线图

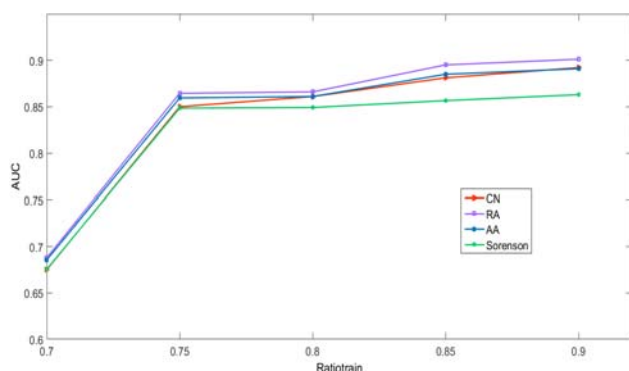


图 5 癫痫患者脑电网络 AUC 曲线图

随后利用 RA 算法分析癫痫患者脑电特征, 图 6 显示了某一例癫痫患者脑电信号 α 波段组成复杂网络的相似度矩阵. 该网络总共含有 19 个节点(1 至 19 分别对应于 Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2), 图中矩阵的行与列分别表示节点 V1 到节点 V19, 矩阵里的每一个元素代表两节点的相似度. 若相似值越大则说明两个节点越有可能产生链接. 然而, 每个节点都有一个最大相似度

节点, 而这个最大相似度节点对此节点的影响力也是最大的. 将 9 例癫痫患者脑电信号中每个记录节点的最大相似度节点找出来(由于部分患者脑电信号仅采集了 18 通道信号, 因此以 18 为网络节点个数), 结果如表 2 和表 3 所示, 其中列出了每个节点的最大相似度节点以及对应的相似值. 图 7 显示了基于 RA 算法的某癫痫患者脑电信号 α 波段节点簇示意图.

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16	v17	v18
1	0	0.34300	0.27546	0.31633	0.25645	0.24045	0.35419	0.2743	0.3233	0.23552	0.23977	0.27237	0.18864	0.07535	0.06845	0.15255	0.29696	0.07549
2	0.34300	0	0.43963	0.35982	0.37598	0.29345	0.38678	0.34082	0.35662	0.32864	0.33977	0.37369	0.30869	0.06521	0.08205	0.11453	0.32955	0.06699
3	0.27546	0.43963	0	0.37925	0.37314	0.30885	0.34312	0.33321	0.32018	0.31599	0.3546	0.34684	0.33642	0.07923	0.07194	0.10879	0.28735	0.08113
4	0.31633	0.35982	0.37925	0	0.32174	0.32895	0.34466	0.29723	0.31909	0.27573	0.29826	0.33812	0.25802	0.07618	0.06891	0.14472	0.27559	0.07803
5	0.25645	0.37598	0.37314	0.32174	0	0.24545	0.2776	0.2694	0.28944	0.25997	0.28968	0.34158	0.27976	0.07337	0.08937	0.18613	0.25989	0.07518
6	0.24045	0.29345	0.30885	0.34312	0.24545	0	0.28772	0.27468	0.28739	0.26764	0.24168	0.26453	0.20586	0.07079	0.0659	0.18881	0.27657	0.0776
7	0.35419	0.38678	0.34312	0.34466	0.2776	0.28772	0	0.28099	0.34478	0.28373	0.28908	0.30557	0.23978	0.07241	0.06821	0.19276	0.31421	0.0743
8	0.2743	0.34082	0.33321	0.29723	0.2694	0.27468	0.28099	0	0.26054	0.29124	0.26879	0.24442	0.20091	0.07154	0.07097	0.12003	0.29228	0.07556
9	0.3233	0.35662	0.32018	0.31909	0.28944	0.26739	0.34478	0.26054	0	0.26132	0.25622	0.38223	0.23531	0.06008	0.05566	0.17118	0.29433	0.06205
10	0.23552	0.32864	0.31639	0.27573	0.25997	0.26374	0.28373	0.29124	0.26132	0	0.29449	0.25995	0.23692	0.04636	0.04076	0.14179	0.26488	0.05031
11	0.23977	0.33977	0.3546	0.29825	0.28908	0.24188	0.28908	0.26879	0.25622	0.29449	0	0.21943	0.26584	0.0592	0.05782	0.23836	0.2336	0.06114
12	0.27237	0.37369	0.34684	0.32812	0.34158	0.26543	0.30657	0.26462	0.30223	0.25995	0.24943	0	0.25798	0.06441	0.05943	0.17633	0.27627	0.06988
13	0.18864	0.30869	0.33642	0.25802	0.27976	0.20586	0.23978	0.20091	0.23531	0.23992	0.26584	0.25798	0	0.06237	0.05826	0.15524	0.26627	0.06443
14	0.07535	0.06521	0.07923	0.07618	0.07337	0.07079	0.07241	0.0754	0.06008	0.04336	0.0592	0.0641	0.06237	0	0.38826	0.17656	0.05997	0.04012
15	0.06845	0.08205	0.0744	0.06891	0.06937	0.0659	0.06821	0.07097	0.05566	0.06076	0.05782	0.05943	0.05826	0.38826	0	0.17216	0.29978	0.39151
16	0.26565	0.31453	0.30879	0.31477	0.26613	0.26891	0.29276	0.23003	0.27118	0.24179	0.23836	0.21633	0.21524	0.07666	0.07216	0	0.25339	0.07823
17	0.29696	0.32955	0.28735	0.27559	0.25989	0.27057	0.31421	0.29228	0.29433	0.26488	0.2336	0.27627	0.20427	0.05397	0.03978	0.25339	0	0.04585
18	0.07549	0.06699	0.08113	0.07803	0.07518	0.0726	0.0743	0.07556	0.06205	0.05031	0.06114	0.04998	0.04443	0.40124	0.39151	0.37823	0.04585	0

图 6 某一例癫痫患者脑电信号 α 波段相似度矩阵

结果发现, 在 α 波段组成的复杂网络中, 样本 1 形成 4 个节点簇(如图 7 所示), 分别为 13(10) (其中 13 代表节点位点, 10 代表节点位点连接的节点个数), 10(6), 9; 样本 8 则形成 6 个簇, 分别为 11(2), 13(3), 17(5), 15(5), 9(2), 1; 而样本 2 至样本 7 则分别形成 4 至 8 个不同的簇. 在 δ (0.5-4Hz) 波段组成的复杂网络, 样本 8 形成 13, 3, 14(4), 11(2), 18(2), 15(2), 1(2), 6, 9, 17, 8 等 11 个簇, 而样本 2 则形成 18(13) 13(3) 10 15 等 4 个子簇, 其余样本分别形成 5 至 9 个子簇.

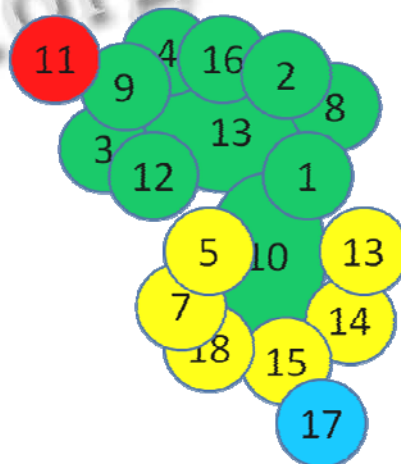
图 7 样本 1 中 α 波段基于 RA 算法的节点簇示意图

表 2 癫痫患者脑电 alpha 波段所有节点对应的最大相似度节点

样本	最大相似度节点(其中原始节点依次为节点 1 至节点 18)																	
1	13	13	13	13	10	13	10	13	13	13	9	13	10	10	10	13	15	10
2	17	5	12	11	18	16	11	16	16	12	7	11	18	11	11	9	1	11
3	7	7	7	7	7	16	1	7	7	7	7	7	7	7	17	7	7	7
4	18	18	18	18	18	18	10	18	17	18	18	15	16	10	10	12	18	10
5	4	4	4	7	4	4	4	4	4	2	4	3	4	18	18	3	4	15
6	4	3	2	1	4	1	4	1	7	2	2	1	6	18	18	4	4	15
7	4	8	4	3	4	3	4	3	7	3	4	3	3	15	14	3	3	14
8	11	13	17	15	17	15	15	13	13	9	1	15	9	17	17	17	15	11

表 3 癫痫患者脑电 delta 波段所有节点对应的最大相似度节点

样本	最大相似度节点(其中原始节点依次为节点 1 至节点 18)																	
1	18	12	17	18	1	13	14	16	13	12	16	1	14	13	12	1	16	1
2	18	13	18	18	10	13	18	18	18	18	18	18	18	18	18	18	15	13
3	16	16	16	16	16	16	11	16	1	16	16	16	16	16	16	1	15	10
4	4	15	4	16	11	3	4	4	12	4	4	4	4	4	4	4	4	13
5	3	3	7	7	4	3	3	3	3	3	3	3	3	18	18	3	3	15
6	2	8	4	7	4	3	4	2	3	4	4	6	6	18	18	5	2	15
7	5	5	4	3	2	2	8	2	2	2	2	8	1	15	14	2	2	14
8	13	3	14	11	18	15	14	14	18	15	17	8	1	1	6	14	11	9

3 结语

本文以复杂网络的基本特征出发,研究其节点的相似性,基于节点的相似性算法对癫痫患者复杂网络链路特征进行分析.自从 1998 年 Watts 和 Strogatz 提出小世界网络的概念后^[16],有许多研究学者们开始利用小世界这一模型展开对大脑网络的研究.此外,张方风等人^[17]也利用网络建构分析方法对手指活动的大脑功能连接特征进行了分析,结果表明该网络具有小世界的特性.基于复杂网络的研究有助于量化分析大脑特征.

本文以复杂网络节点的相似性为基本特征,引入 CN、RA、AA、Sorenson 四种节点相似性算法,分别在线虫的新陈代谢网络(Metabolic)、科学家合作网(NetScience)、爵士音乐家合作网(Jazz)、美国航空网络(USAir)四个真实系统中实现了预测实验,根据 AUC 的值评判预测的准确性,发现 RA 算法的预测效果优于其他算法;同时在仿真的小世界与无标度网络中也进行了预测实验,结果也发现 RA 算法预测准确性比较高.随后利用 RA 算法分析 8 例全身性癫痫患者脑电信号,发现该方法能较好刻画由节点相似度组成的节点对子簇;对同一样本而言,该方法也能刻画不同频率波段(delta, alpha)组成的复杂网络其节点相似度组

成的子簇.结果初步表明基于节点相似度的分析方法能量化并可视化癫痫患者脑电信号功能连接特征.

不同节点相似度算法的预测能力可以显示网络潜在的结构信息,网络结构特征和节点属性将影响网络中不同链路预测的算法表现.本文中的四种算法都是基于局部信息的节点相似性度量技术,相对基于全局信息度量指标,这些算法复杂度低,计算速度快,但预测精度较低. CN 算法公式简单,复杂度低,但容易被节点度异质性影响. Sorenson 算法同时考虑共同邻居个数和节点对的节点度,仅从网络结构和节点属性出发,难以与物理意义关联. RA 和 AA 算法均抑制高节点度共同邻居的作用力,当节点度小时二者表现类似,当节点度大时则 RA 算法优于 AA 算法. RA 描述的是网络传输能力和连接性的非线性关系,引入了次邻近节点信息描述节点相似性,有效消除了由局部相似性度量引发的退化态,这可能是 RA 算法优于其他算法的潜在原因(补充节点相似性算法 CN、RA、AA、Sorenson 展开全面的对比,包括算法描述、特点、算法复杂度等).

癫痫病是一种严重危害患者生命健康的恶性疾病,且原发性癫痫患者大多为儿童,由于其病理机制复杂,多种神经递质和生理因素都可能影响疾病进程,临床

诊断和治疗主要依赖于医生的经验, 缺乏准确的量化指标. 基于节点相似性的癫痫患者脑网络分析不仅有效量化其大脑特征, 还有助于分析和预测癫痫患者大脑中的异常连接及相关脑区病例变化. 未来可以引入大规模样本患者及对照组数据, 在本文节点相似性及其相似值的基础上提炼关键指标, 为临床辅助诊断与病人脑功能状态量化估计提供技术支持.

参考文献

- 1 Du F, Xuan Q, Wu TJ. Empirical analysis of attention behaviors in online social networks. *International Journal of Modern Physics C*, 2010, 21(7): 955–971.
- 2 吕琳媛. 复杂网络链路预测. *电子科技大学学报*, 2010, 39(5): 651–661.
- 3 Lu L, Zhou T. Link prediction in complex networks: A survey. *Physic A: Statistical Mechanics and its Applications*, 2011, 390(6): 1150–1170.
- 4 Lv LY, Zhou T. Link prediction in weighted networks: The role of weak ties. *EPL*, 2010, 89(18001): 1–6.
- 5 Getoor L, Diehl CP. Link mining: A survey. *ACM SIGKDD Explorations Newsletter*, 2005, 7(2): 3–12.
- 6 Yamada T, Bork P. Evolution of biomolecular networks - lessons from metabolic and protein interactions. *Nature*, 2009, 10(11): 791–803.
- 7 Liben-Nowell D, Kleinberg J. The link prediction problem for social networks. *Proc. of the 12th International Conference on Information and Knowledge Management (CIKM)*. 2003. 556–559.
- 8 Clauset A, Moore C, Newman MEJ. Hierarchical structure and the prediction of missing links in networks. *Nature*, 2008, 453(7191): 98–101.
- 9 Wagner A, Fell DA. The small world inside large metabolic networks. *Proc. of the Royal Society of London. Series B: Biological Sciences*, 2001, 268(1478): 1803–1810.
- 10 Maslov S, Sneppen K. Specificity and stability in topology of protein networks. *Science*, 2002, 296(5569): 910–913.
- 11 Latora V, Marchiori M. Efficient behavior of small-world networks. *Physical Review Letters*, 2001, 87(19): 198701.
- 12 Milgram S. The small world problem. *Psychology Today*, 1967, 2(1): 60–67.
- 13 Zhou T, Lu L, Zhang YC. Predicting missing links via local information. *The European Physical Journal B-Condensed Matter and Complex Systems*, 2009, 71(4): 623–630.
- 14 Yan H, et al. Frequency dependent network flexibility analysis in epileptic brain based on phase locking value and resilience test. 2014 10th International Conference on Natural Computation(ICNC). Xiamen. 2014. 25–29.
- 15 孙俊峰, 洪祥飞, 童善保. 复杂脑网络研究进展-结构、功能、计算与应用. *复杂系统与复杂性科学*, 2010, 4: 74–90.
- 16 Watts DJ, Strogatz SH. Collective dynamics of “small-world” networks. *Nature*, 1998, 393: 440–442.
- 17 张方凤, 陈春辉, 姜璐. 数字背诵过程的大脑功能网络. *中国医学物理学杂志*, 2006, 23(6): 419–422.