

基于层次社团划分算法的名老中医治疗肺癌用药规律研究*

李 昱 杨 涛 胡孔法

(南京中医药大学人工智能与信息技术学院 南京 210023)

〔摘要〕 介绍国内外中医药领域应用复杂网络相关研究现状, 提出构建中药带权网络并利用层次社团划分算法挖掘名老中医治疗肺癌的用药层次规律, 阐述层次社团划分算法具体应用方法并对实验结果进行分析。

〔关键词〕 层次社团划分; 肺癌; 用药规律; 中医诊疗经验

〔中图分类号〕 R-058 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2022.09.007

Study on the Medication Rule of Famous Veteran Physicians of Traditional Chinese Medicine in Treating Lung Cancer Based on Hierarchical Community Partition Algorithm LI Yu, YANG Tao, HU Kongfa, *Artificial Intelligence and Information Technology Institution, Nanjing University of Chinese Medicine, Nanjing 210023, China*

〔Abstract〕 The paper introduces the research status of the application of complex network in the field of Traditional Chinese Medicine (TCM) at home and abroad, proposes to construct weighted network of Chinese medicine and use hierarchical community division algorithm to mine the hierarchical rule of drug use in the treatment of lung cancer by famous veteran physicians of TCM, expounds the application method of hierarchical community division algorithm, and analyzes the experimental results.

〔Keywords〕 hierarchical community division; lung cancer; medication rule; diagnosis and treatment experience of Traditional Chinese Medicine (TCM)

1 引言

肺癌是我国发病率、死亡率最高的恶性肿瘤^[1]。肿瘤治疗的主要方式为西医手术及放化疗, 对缩小瘤体的效率高, 但是患者常在放化疗过程中

出现免疫功能抑制、血象下降的现象^[2]。中医在改善肺癌患者生活质量、恢复体质、延长生存期等方面具有优势。如何挖掘和传承名老中医治疗肺癌的经验成为中医传承与发展的重要课题。传统中医经验传承主要通过医案整理、方药总结等途径, 耗费大量人力^[3]。随着信息技术的发展和普及, 中医医

〔修回日期〕 2021-11-05

〔作者简介〕 李昱, 本科生, 发表论文 2 篇; 通信作者: 杨涛, 副教授。

〔基金项目〕 国家自然科学基金委员会面上项目“基于知识图谱的现代名老中医诊治肺癌用药规律及其机制研究”(项目编号: 82074580); 教育部产学研合作协同育人项目“新工科背景下医学信息创新人才培养模式研究”(项目编号: 202101224003); 教育部产学研合作协同育人项目“医药人工智能联合创新实践教学基地建设”(项目编号: 201902183039); 江苏高校哲学社会科学研究一般项目“基于医案数据挖掘的名老中医临床经验数字化传承模式构建及方法研究”(项目编号: 2021SJA0333)。

案数量急剧增长,传统的经验传承模式已经不能满足需求,迫切需要利用现代化技术和方法进行传承。中医药蕴含丰富的非线性复杂关系,复杂网络是探寻这一复杂关系的有效方法之一,将复杂系统内部元素和关系用节点和边表示,可帮助发现复杂系统中各因素间交互规律的网络模型^[4]。社团结构是复杂网络中的重要特性,通过构造肺癌用药的复杂网络提取出其中的社团结构,进而总结名老中医临床处方的配伍经验,对于深入了解肺癌的诊治规律具有重要意义。

2 国内外研究现状

复杂网络及社团发现在中医药发展过程中发挥重要作用,虽然中医药领域应用复杂网络的研究尚处于初级阶段,但越来越多的学者开始尝试将中医药学与复杂网络相结合^[5]。周雪忠、刘保延和王映辉等^[6]利用复方配伍的无尺度复杂网络现象,对名老中医的处方配伍经验及面向特定病症的配伍经验进行挖掘分析,从而发现核心药物、药对等;过红玲、顾浩和王耘等^[7]依据中医基础理论以及临床用药规律,揭示中药功效之间的关系;赵君霞^[8]从复杂网络角度对中医学规律进行分析,通过对丰富的中医临床数据构建药物配伍网络,并利用社团分析方法进行研究,尝试对中医基础理论的复杂网络特性进行揭示;Gao Z、Li S和Shang Q等^[9]利用复杂网络分析稳定期冠心病患者的中医证型演变与心血管复发的关联性,结果显示该病患者以气虚证、血瘀证、毒症为主要证型。Guo F F、Zhang W和Su J等^[10]从中药成分-网络靶标层面构建中药治疗疾病的网络模型,设计网络扰动定量评价算法来评估药物对于网络的干预程度。上述研究从不同角度进行了中医复杂网络研究及应用,对中医药传承进行了有益探索。中医用药具有层次性和结构性,在复杂网络中挖掘层次结构能够更精准地把握复杂网络的结构和组成^[11]。层次社团挖掘方法分为两种:一种是自上而下的分裂式,即不断将1个大社团分裂为多个小社团;另一种是自下而上的凝聚式,即不断选取相似的小社团合并为大社团。GN算法是经典

的分裂式社团发现算法。但该算法采用全连接组网的方式,计算复杂度较高,冗余连接增加,影响后续社团发现效果^[12]。2004年基于模块度 Q 的Fast Newman算法被提出。通过不断计算模块度增量,沿着使模块度增加最多或减少最少的方向依次合并两个社团,时间复杂度明显优于GN算法^[13]。鉴于此,本文以Fast Newman算法为基础,提出层次社团划分算法,并将其应用到肺癌用药挖掘中,用以揭示肺癌用药社团结构以及社团内部的层次性。

3 层次社团划分算法

3.1 中药带权网络构建

首先,对方剂集进行预处理,清洗冗余信息,如药物剂量和药物用法,得到互不重复的处方集合。遍历处方集,将处方集转换为以药物总数为行维度、处方总数为列维度的0-1矩阵。在矩阵中的同一行,1代表该味中药出现在这条处方中,0代表未出现。其次,构造以中药为节点、两两中药出现频次为权重的邻接关系矩阵。再次,遍历中药关系矩阵,将矩阵中药对出现的频次与阈值进行对比,如果大于阈值则将两节点构成的边输入边集,并将药对出现频次作为边的权值。最后,输出带权边集与节点集形成复杂网络。

3.2 中药带权网络划分

3.2.1 中药社团划分 以Fast Newman算法为基础,采用自底而上的中药节点合并,通过最大化模块度 Q 来实现中药社团划分。根据贪心算法的思想,不断沿着使 Q 增加最大或减小最少的方向依次合并两个中药社团。 Q 函数表示为:

$$Q = \sum_i (e_{ii} - a_i^2) = Tre - \|e^2\| \quad (1)$$

其中 $\|e^2\|$ 表示矩阵 e^2 的模,即矩阵中各元素之和。 $a_i = \sum_j e_{ij}$ 表示与社团 i 中节点相连的边占所有边的比例。

3.2.2 中药带权网络划分过程 第1步,将构成复杂网络的 n 个中药节点视作 n 个初始社团,定义变量 a_i 及 e_{ij} 如下:

$$a_i = k_i/2m \quad (2)$$

$$e_{ij} = \begin{cases} \frac{1}{2m}, & \text{节点 } i \text{ 与节点 } j \text{ 间有边相连。} \\ 0, & \text{其他。} \end{cases} \quad (3)$$

式中 k_i 为节点 i 的度, m 为网络中的总边数。第 2 步, 依次合并有连边的两个社团, 计算合并前后社团的模块度增量, 沿着使 Q 增加最大或减小最少的方向对社团进行合并。模块度增量公式如下:

$$\Delta Q = e_{ij} + e_{ji} - 2 a_i a_j = 2(e_{ij} - a_i a_j) \quad (4)$$

合并社团后, 更新相应变量。第 3 步, 重复执行步骤 2, 直至整个网络合并成 1 个大社团, 完成社团划分过程。

3.2.3 中药社团层次结构提取 社团依据模块度增量依次顺序合并直至整个网络合并为 1 个大社团。每层合并 2 个社团, 层级越低合并越优先。通过树状图对上述社团合并过程进行展示, 可以明显观察到社团内部层次结构, 即顺序合并至同一个中药组合的药物可以认为归属于同一社团。对于划分出来的社团采取迭代划分, 观察社团内部分裂结果, 直至社团合并成链状结构或模块度无明显增益时停止划分。根据社团合并顺序和社团模块度提取中药组合的层次结构信息, 确定最终划分类别并采用环状树状图进行展示。

4 实验结果及分析

4.1 数据来源及预处理

4.1.1 数据来源 本研究的实验数据来源于南京中医药大学名老中医传承中心。检索近 5 年有关肺癌的病案数据, 根据数据纳排标准最终得到 167 例肺癌患者的 796 诊次医案。其中最小 2 诊次, 最多 33 诊次, 平均 4.67 诊次。

4.1.2 预处理 对于医案中的中药数据进行预处理, 包括药名规范化、异常值去除、药物用法去除等。其中医案纳入采取以下标准: 依据《原发性肺癌诊疗规范》, 确诊为原发性肺癌; 医案包含患者基本信息、舌象、脉象、临床表现等内容。医案排

除采取以下标准: 医案中临床表现、舌象、脉象存在缺失; 患有严重的心、肝、肾等疾病, 影响肺癌四诊判断; 非原发性肺癌患者。

4.2 实验结果

4.2.1 肺癌高频病机 医案中出现的病机为痰瘀阻肺、肺气亏虚、肾虚痰阻等, 见表 1。

表 1 肺癌高频病机 (频数 >5)

证型	频数
痰瘀阻肺	745 (93.59%)
肺气亏虚	24 (3.01%)
肾虚痰阻	11 (1.38%)
痰瘀阻滞	9 (1.13%)
肺失宣降	8 (1.00%)
气血两虚	8 (1.00%)
痰瘀水阻肺	6 (0.75%)

4.2.2 核心药物频次 方剂中共含有 190 味中药, 其中山慈菇等 20 味中药出现频次最高, 见表 2。

表 2 核心药物频次 (排名前 20 位)

药物名称	频次	药物名称	频次
山慈菇	743 (93.34%)	广郁金	599 (75.25%)
茯苓	691 (86.80%)	天龙	594 (74.62%)
法半夏	685 (86.05%)	白花蛇舌草	568 (71.36%)
川贝粉	683 (85.80%)	太子参	548 (68.84%)
苦杏仁	680 (85.43%)	醋五味子	514 (64.57%)
桔梗	674 (84.67%)	三七粉	502 (63.07%)
姜厚朴	645 (81.03%)	炙甘草	499 (62.69%)
酒地龙	626 (78.64%)	白术	482 (60.55%)
半枝莲	617 (77.51%)	薏苡仁	467 (58.67%)
麦冬	600 (75.38%)	猪苓	456 (57.29%)

4.2.3 药物共现频次 为中药两两出现的次数, 其中苦杏仁 + 桔梗等 20 对药物共现频次最高, 见表 3。

表 3 配伍网络对应的药物共现频次 (排名前 20 位)

药物名称	药物名称	共现频次	药物名称	药物名称	共现频次
苦杏仁	桔梗	672 (84.42%)	茯苓	苦杏仁	592 (74.37%)
山慈菇	茯苓	657 (82.54%)	茯苓	桔梗	586 (73.62%)
山慈菇	苦杏仁	638 (80.15%)	法半夏	苦杏仁	585 (73.49%)
山慈菇	法半夏	635 (79.77%)	川贝粉	茯苓	584 (73.37%)
山慈菇	川贝粉	635 (79.77%)	川贝粉	苦杏仁	581 (72.99%)
山慈菇	桔梗	630 (79.15%)	山慈菇	半枝莲	580 (72.86%)
山慈菇	姜厚朴	612 (76.88%)	法半夏	桔梗	579 (72.74%)
法半夏	川贝粉	600 (75.38%)	酒地龙	山慈菇	578 (72.61%)
天龙	酒地龙	594 (74.62%)	川贝粉	桔梗	575 (72.24%)
法半夏	茯苓	592 (74.37%)	山慈菇	麦冬	565 (70.98%)

4.2.4 中药网络 以 0.1 作为节点阈值筛选高频中药构建中药网络, 其中共有 51 味中药节点, 1 578 条药对连边。中药的节点度越大则网络中对应的节点半径越大, 见图 1。

白术、薏苡仁、木香等 6 味中药组成; 社团 C3.3 由山慈菇、半枝莲、白花蛇舌草等 11 味中药组成。

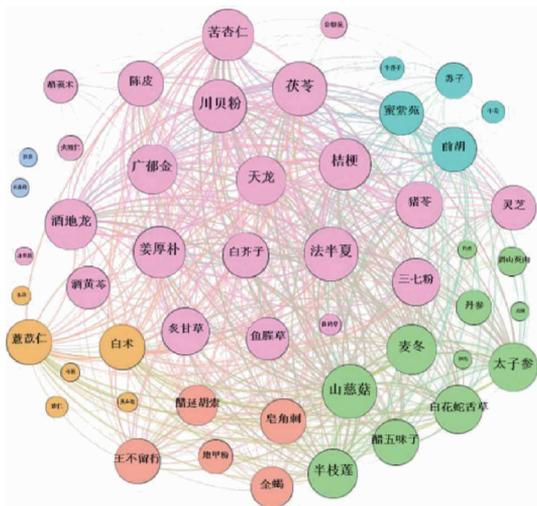


图 1 中药配伍复杂网络

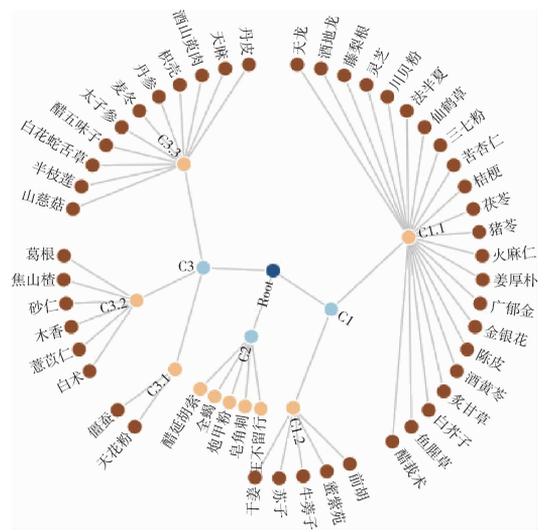


图 2 社团划分层次

4.2.5 社团划分 对中药带权网络进行社团划分, 采用树状图形式对社团合并过程进行展示。根据树状图、模块度变化情况, 结合中医专家建议, 最终划分 6 个社团并采用环形树状图进行展示, 见图 2。药物被划分成两个层次, 第 1 层有 C1、C2、C3 这 3 个社团, 其中 C1 又划分成 C1.1 和 C1.2, C3 划分成 C3.1、C3.2 和 C3.3。社团 C1.1 由天龙、酒地龙、藤梨根等 22 味中药组成; 社团 C1.2 由前胡、蜜紫苑、牛蒡子等 5 味中药组成; 社团 C2 由王不留行、皂角刺、炮甲粉等 5 味中药组成; 社团 C3.1 由天花粉、僵蚕 2 味中药组成; 社团 C3.2 由

4.3 分析与讨论

4.3.1 概述 本文通过筛选高频中药构建加权网络, 采用模块度衡量社团划分指标, 结合中药合并过程的树状图结构确定社团划分数目。算法共划分出两层中药社团, 第 1 层 3 个中药社团 (C1、C2、C3), 在第 1 层社团的基础上进行进一步划分, 得到第 2 层 5 个中药社团 (C1.1、C1.2、C3.1、C3.2、C3.3)。相较于传统方法, 本文提出的方法较好地呈现了社团之间的层次性和结构性, 可以从不同层次 (粒度) 分析名老中医诊疗经验。

4.3.2 药物配伍网络以及药物出现频次分析 可发现核心药物主要集中于清热解毒散瘀类 (山慈

菇、酒地龙、半枝莲、白花蛇舌草等)、利水渗湿类(茯苓、薏苡仁、猪苓等)、燥湿化痰类(法半夏、川贝粉、白术等)、益气生津类(太子参、醋五味子等)、养阴润肺类(麦冬、炙甘草等)、宽中理气类(姜厚朴等)。

4.3.3 社团药物组成分析 社团 C1.1 的药物较多,以天龙、酒地龙、醋莪术、广郁金活血化瘀,法半夏、陈皮、川贝粉、白芥子、鱼腥草、苦杏仁、桔梗等药理气止咳化痰,酒黄芩、金银花、藤梨根等药清热解毒燥湿,猪苓、茯苓、火麻仁、姜厚朴利湿通腑,仙鹤草、三七粉止血活血,再辅以灵芝补虚,炙甘草调和诸药。社团 C1.2 以前胡、蜜紫苑、苏子、牛蒡子止咳化痰、降气平喘,干姜温中散寒。社团 C2 以王不留行、皂角刺、全蝎、醋延胡索、炮甲粉活血止痛、攻毒散结。社团 C3.1 以天花粉、僵蚕滋阴润燥、化痰散结。社团 C3.2 以白术、薏苡仁、木香、砂仁、焦山楂、葛根健脾和胃、生津止渴。社团 C3.3 以山慈菇、半枝莲、白花蛇舌草、丹参活血化瘀、清热解毒,以五味子、麦冬、酒山萸肉、太子参、丹皮气阴双补,天麻祛风通络,枳壳破气消积。从上述功效分析来看,肺癌治疗以化痰祛瘀为主,也可兼有清热解毒、滋阴润燥、行气止痛、温中健脾等。这一结果与原始数据中病机分布一致,痰瘀阻肺出现频数最高,频率达到 93.59%,其他依次为肺气亏虚(3.01%)、肾虚痰阻(1.38%)、痰瘀阻滞(1.13%)、肺失宣降(1.00%)、气血两虚(1.00%)、痰瘀水阻肺(0.75%)。

4.3.4 肺癌病机分析 依据上述社团划分结果结合临床专家意见进行合理推测(以药测证),肺癌的病机较为复杂,临床以痰瘀互结为主,兼有痰热、气虚、阴虚、气滞、中寒等。肺癌的实性病理因素主要包括痰、瘀、气郁、癌毒等,其中以痰、瘀最为常见;虚性病理因素中气虚、阴虚最为常见^[14]。临床治疗应审证求因、圆机活法、灵活用药。

5 结语

肺癌是危害人类生命健康的重大疾病,具有发病率高、死亡率高等特点,中医药在肺癌治疗和术

后调理方面具有优势。发掘和传承名老中医诊治肺癌的经验对肺癌防治意义重大。本文借助层次社团划分算法对名老中医诊治肺癌医案进行分析,成功挖掘出核心的中药配伍社团及其层次结构,较好地反映了名老中医诊治肺癌的用药特点和规律,对名老中医诊疗经验的数字化传承具有重要意义。

参考文献

- 1 孙可欣,郑荣寿,张思维,等. 2015 年中国分地区恶性肿瘤发病和死亡分析 [J]. 中国肿瘤, 2019, 28 (1): 1-11.
- 2 容志航. 北京市名老中医治疗肺癌的经验总结与临床研究 [D]. 北京: 北京中医药大学, 2013.
- 3 陈曦,李宜放. 基于医案解构探讨名老中医经验数据挖掘现状 [J]. 中华中医药杂志, 2019, 34 (6): 2608-2611.
- 4 Newman M. Networks: An Introduction [M]. New York: Oxford University Press, 2010.
- 5 严蓓华,杨铭,陈佳蕾,等. 复杂网络在中医药方面的研究和应用 [J]. 中国实验方剂学杂志, 2012, 18 (7): 276-280.
- 6 周雪忠,刘保延,王映辉,等. 复方药物配伍的复杂网络方法研究 [J]. 中国中医药信息杂志, 2008, 15 (11): 98-100.
- 7 过红玲,顾浩,王耘,等. 中药功效概念网络构建及其应用 [J]. 世界科学技术(中医药现代化), 2013, 15 (4): 638-642.
- 8 赵君霞. 复杂网络在中医临床知识发现中的应用研究 [D]. 北京: 北京交通大学, 2009.
- 9 Gao Z, Li S, Shang Q, et al. Complex Networks Approach for Analyzing the Correlation of Traditional Chinese Medicine Syndrome Evolution and Cardiovascular Events in Patients with Stable Coronary Heart Disease [J]. Evidence-Based Complementary and Alternative Medicine, 2015 (2015): 824850-824856.
- 10 Guo F F, Zhang W, Su J, et al. Prediction of Drug Positioning for Quan-Du-Zhong Capsules Against Hypertensive Nephropathy Based on the Robustness of Disease Network [J]. Front Pharmacol, 2019 (10): 49.
- 11 余成进,赵姝,陈洁,等. 复杂网络中的层次结构挖掘 [J]. 南京大学学报(自然科学), 2016, 52 (5): 861-870.
- 12 陶妍心,杨涛,谢佳东,等. 基于关联网络和层次聚类的中药社团发现算法的设计与应用 [J]. 世界科学技术-中医药现代化, 2020, 22 (6): 1962-1968.
- 13 Newman M E J. Fast Algorithm for Detecting Community Structure in Networks [EB/OL]. [2021-08-16]. <http://www.doc88.com/p-0307203007226.html>.
- 14 王珊珊. 周仲瑛教授从痰瘀热毒辨治肺癌的临床经验及益肺解毒汤的抗肿瘤实验研究 [D]. 南京: 南京中医药大学, 2017.