

基于 CNM-Centrality 算法的失眠症辨证论治中核心中药及配伍研究

王嫣然, 王明珠, 胡梦宇, 胡芳*

湖北中医药大学信息工程学院, 湖北 武汉 430065

摘要: 选取失眠症作为研究对象, 采用复杂网络节点中心性评估和聚类分析方法, 探索失眠症辨证论治中核心中药及配伍规律。首先, 通过构建失眠中药网络模型, 引入复杂网络节点中心性评估单指标算法, 挖掘中药网络的核心节点; 其次, 利用基于节点中心性的聚类算法 CNM-Centrality 对中药网络进行聚类划分, 准确地找到药物间的配伍规律。

关键词: 复杂网络分析; CNM-Centrality 算法; 失眠症; 核心中药; 中药配伍

中图分类号: R283.2 **文献标志码:** A **文章编号:** 0253-2670(2017)18-3897-04

DOI: 10.7501/j.issn.0253-2670.2017.18.033

Research on core Chinese materia medica and herb compatibility in syndrome differentiation and treatment of insomnia based on CNM-centrality algorithm

WANG Yan-ran, WANG Ming-zhu, HU Meng-yu, HU Fang

Information Engineering Institute, Hubei University of Chinese Medicine, Wuhan 430065, China

Abstract: Using node centrality evaluation algorithms and clustering algorithms in complex networks, this paper selects insomnia as study object, and explores the core Chinese materia medica (CMM) and CMM compatibility regularity of syndrome differentiation and treatment. Firstly, we construct the CMM network model of insomnia, and discover the core CMM nodes in this network using single-index evaluation algorithms. Then, the CMM network can be divided by the clustering algorithm CNM-centrality based on node centrality, and the compatibility regularity among CMM can be found accurately.

Key words: complex network analysis; CNM-centrality algorithm; insomnia; core Chinese materia medica; Chinese materia medica compatibility

中医学是从整体上把握疾病变化特点及规律的系统医学, 是在临床实践中产生并不断验证发展的学科, “从临床中来, 到临床中去” 是中医临床研究的基本模式。中医新技术、新方法、新方药的萌芽、发展、形成是“从临床中来” 阶段逐渐积累、提炼、升华的过程, 直至从个体应用到群体应用^[1]。在临床诊疗实践过程中, 积累的大量病历数据、典型医案数据和古今文献资料是十分重要的研究资源。大数据时代, 利用大规模临床数据开展科学研究, 从数据中归纳挖掘对临床有用或具备理论意义的知识, 进而形成创新知识和临床决策^[2]。

复杂网络研究是近年来科学研究中的热门问题, 在计算机科学、社会学、经济学、生命科学领域都得到了广泛的应用, 是一种研究网络节点潜在

关系的分析方法^[3]。中医临床诊疗是患者机体反映、医生思维决策和复杂干预手段的非线性互动过程, 具有复杂相关性、涌现性等特点^[4]。因此, 将复杂网络作为描述复杂系统的模型和工具, 可以应用到中医领域中, 对临床表征、症候、治则治法、药物之间关系等复杂问题进行探索。通过构建复杂网络模型, 将复杂关系简单化, 既能揭示其特征, 又能用数字说明各节点间关系的强度、紧密度等特性, 从而直观、定量、清晰地刻画出各种复杂网络关系, 以及某些节点的核心地位及作用。近年来, 复杂网络分析方法在医学领域中广泛使用, 尤其是在著名老中医处方经验的总结和继承研究方面比较多。聚类分析是研究数据集分类问题的一种多元统计方法, 是根据对象在某些属性上的相似性, 将其划分

收稿日期: 2017-02-28

基金项目: 2017 年度湖北省教育厅科研计划项目—基于优化聚类算法的中医失眠证候判别分析研究 (Q20172005)

作者简介: 王嫣然, 医学信息工程专业。E-mail: yranwwang@126.com

***通信作者** 胡芳, 副教授, 博士, 主要研究方向为医学信息学、复杂网络与复杂系统。E-mail: naomifang@126.com

成簇的过程。对该算法的研究有利于发现处方中药物的组成和配伍规律。

本文选取失眠症作为研究对象,采用复杂网络中聚类分析和中心性评估等方法,探讨失眠症辨证论治中的核心中药及配伍规律。首先构建中药网络模型,通过对其进行深入分析,挖掘出中药网络中的核心药物节点和聚类特征。引入复杂网络节点中心性相关算法,包括度中心性度量、介数中心性度量、紧密中心性度量^[5]和特征向量中心性度量^[6]等对中药节点的重要性进行客观分析。其中,度中心性反映了节点在网络中的影响力以及节点直接获取网络信息的能力;中介中心性反映了经过节点的最短路径的比例;紧密中心性反映了各节点与其他节点之间的距离远近^[5];特征向量中心性综合考虑节点的邻居节点的重要程度^[6]。其次,利用基于节点中心性的聚类算法 CNM-Centrality^[7]对中药网络模型进行聚类分析,找到中药配伍规律。

1 中药网络模型构建

本文以 PubMed、中国知网、万方、维普、中

国生物医药及中医药在线数据库为数据来源,收集失眠相关临床文献,对文献中治疗失眠症的复方进行分析挖掘并构建中药网络模型。将治疗失眠症的复方中药作为网络节点,节点的连边越多,说明该药物的节点度越高,节点度是考察该节点在网络中直接影响力的指标。2 种中药同时出现次数越多,频度越大,则 2 个节点之间的连接线越粗,即反映 2 个节点之间关系越密切。失眠症中药网络的构建过程见图 1。如“当归”与“生地黄”在同一处方中出现,则 2 者之间有一条连边;2 味中药在不同处方中同时出现频次越高,则边的权重越大,边的连线越粗,根据该组网规律构建失眠症中药网络模型。

2 聚类算法 CNM-Centrality 思想实现

2.1 算法思想

CNM-Centrality 算法首先利用 PageRank 算法^[8]识别出给定中药网络的核心（重要）中药节点，并按降序排序。将每个中药节点初始化为 1 个社团，计算出初始的模块度值。然后将每个中药节点合并

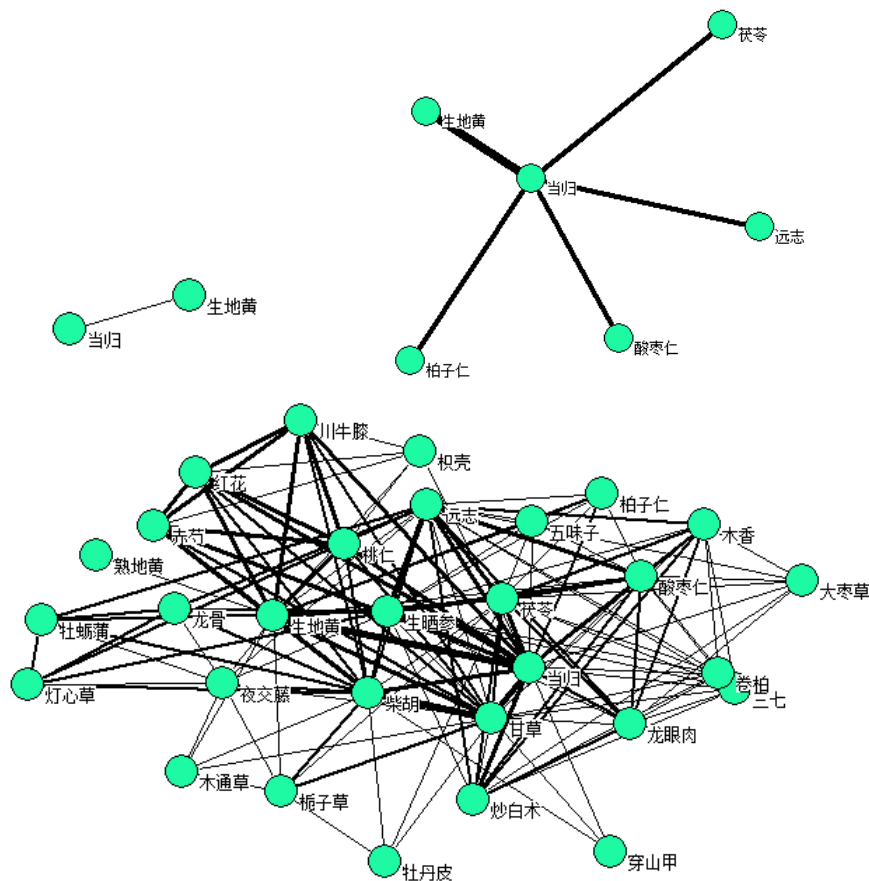


图 1 失眠症中药网络模型

Fig. 1 CMM network model of insomnia

生成的最大模块度值填入到最大值堆中，遍历最大值堆，选取出最大的模块度值并且挑选出对应的 2 个社团，若 2 个社团中未包含重要中心节点或只有 1 个社团包含有重要节点，则直接合并 2 个社团。当 2 个社团之中都有中心节点时，通过计算 SNN 相似度值^[9]判断这 2 个社团能否合并，若两者相似度够高，则合并 2 个社团，选取 PageRank 算法运行结果中排名靠前的节点为新社团的中心节点。若相似度不够高，则不合并，继续选取最大值堆中增长模块度值第 2 的 2 个社团进行合并尝试。合并成功后，更新模块度值矩阵。当所有的社团都不能继续合并后，停止迭代，输出最后的聚类结果。

2.2 算法步骤

2.2.1 步骤 1 构建一个无权、无向的中药网络矩阵 (G)， n 表示中药节点总数。

$$G = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix}$$

2.2.2 步骤 2 去除中药网络中的自旋节点， $diag(G)$ 为自旋节点集合。

$$G = G - diag(G)$$

$$diag(G) = \begin{pmatrix} x_{11} & 0 & \cdots & 0 \\ 0 & x_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_{nn} \end{pmatrix}$$

2.2.3 步骤 3 利用 PageRank 算法识别出中心节点。

2.2.4 步骤 4 基于节点中心性，运用 CNM-Centrality 算法进行社团划分，将 n 个中药节点初始化为 n 个中药社团，计算初始值 ΔQ_{ij} ，其中 $i = \{1, 2, \cdots, n\}$ ， $j = \{1, 2, \cdots, n\}$ ，若社团 n 和 j 中有边相连，则 $\Delta Q_{ij} = 1 - 2e - d_i d_j / (2e)^2$ ， e 代表网络中的边数， d_i 代表第 i 个节点的度。

$$e = \sum_{c=1}^n \sum_{d=1}^n G_{cd} / 2$$

$$d_i = \sum_{c=1}^n G_{ic}$$

2.2.5 步骤 5 计算矢量元素 a_i ，该矢量元素将会用于社团合并后更新最大值堆。

$$a_i = \frac{d_i}{2e}$$

2.2.6 步骤 6 构建大小为 n 的最大堆 H ，其中第 i 个元素为 $H(i) = \max(\Delta Q_{i1}, \Delta Q_{i2}, \cdots, \Delta Q_{in})$ ，挑选

出最大堆 H 中最大的 ΔQ_{ij} 。

2.2.7 步骤 7 判断社团和是否能合并，若两者的相似度大于 0.5，则说明 2 个社团可以合并，将 2 个社团中 PageRank 值排名更加靠前的 1 个中心节点选为新合并社团的中心节点。若相似度小于 0.5，则算法将不合并 2 个社团。挑选出最大堆 H 中第 2 大的 ΔQ_{ij} 。重复执行，直到没有社团能合并。

当有社团 i 和 j 合并后，在 ΔQ 的矩阵中只需更新第 j 行和第 j 列，并且移除第 i 行和第 i 列，以更新矩阵。

2.2.8 步骤 8 重复以上步骤，直接所有社团都不能两两合并，输出聚类结果。

3 中药网络仿真与分析

失眠症的中药网络由 48 个节点和 548 个边构成，每个节点代表 1 味中药，边表示 2 味药物在同一个药方中使用过。其中节点出现频次最高是当归 (120 次)，度值最高是当归 (62 次)、甘草 (62 次)。CNM-Centrality 算法对失眠症的中药网络划分结果如图 2 所示，通过测试得出，本网络被划分为 5 个聚类结果时，能够清晰地表现出节点的关系特征，其中相同的颜色表示同一类别的节点。根据度值和紧密中心性值对治疗失眠症的中药排序，排名前 15 名的中药见表 1。结合图 2 和表 1，甘草、当归、柴胡、茯苓、远志等节点中心性较高；甘草、夜交藤、天花粉、牡丹皮、泽泻、龙胆草、车前草、木通、白芍、栀子、柴胡、薄荷、穿山甲、知母、茵陈、地骨皮等中药划分到以甘草为核心节点的网络划分中，其配伍关系较为亲密；当归、茯苓、生晒参、远志、天冬、丹参、柏子仁、五味子、酸枣仁、炒白术、白芷、熟地黄、麦冬、大枣、木香、生地黄、三七、炙甘草、党参、龙眼肉、卷柏等中药划分到以当归为核心节点的网络划分中，其配伍关系较为亲密；桃仁、川牛膝、红花、赤芍、枳壳等中药划分到以桃仁、川牛膝、红花、赤芍为核心节点的网络划分中，其配伍关系较为亲密；黄连和肉桂自成 1 个社团网络，其配伍关系较为亲密，和其他药物之间配伍关系不密切。

4 结语

基于实际诊疗数据的复杂网络分析一直是近年来研究的热点，该研究能够揭示节点与节点之间的潜在关系，以及节点在整个网络中的地位和作用。本文通过构建治疗失眠症的中药网络模型，引入复杂网络节点中心性分析思想，对中药节点进行多角

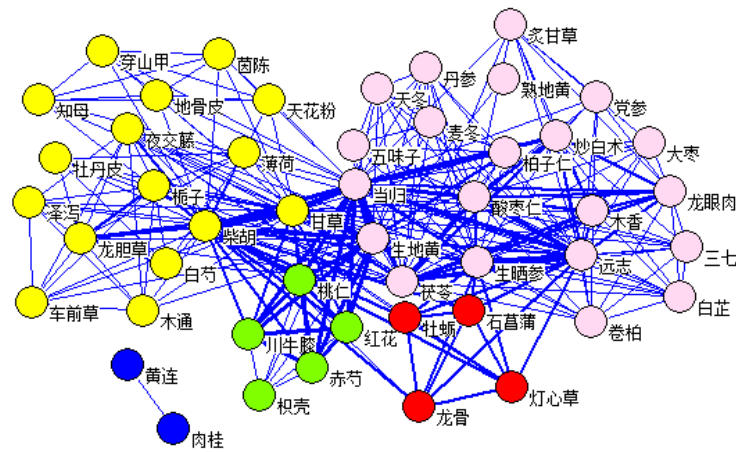


图 2 中药网络聚类分析示意图

Fig. 2 Schematic diagram of clustering analysis in CMM network

表 1 中药网络节点中心性分析

Table 1 Analysis of node centrality in CMM network

排行	药材	度值	紧密中心性值	中介中心性值	特征向量中心性值
1	甘草	62	0.730 932 2	0.133 571 144	0.311 598 888
2	当归	62	0.730 932 2	0.141 074 161	0.402 117 533
3	柴胡	56	0.695 564 5	0.124 888 601	0.301 289 021
4	茯苓	52	0.673 828 1	0.074 266 034	0.234 795 603
5	远志	52	0.673 828 1	0.066 016 621	0.263 452 594
6	生地黄	48	0.653 409 1	0.104 072 664	0.300 677 359
7	生晒参	46	0.643 656 7	0.042 417 635	0.215 654 322
8	酸枣仁	40	0.616 071 4	0.015 989 837	0.204 195 989
9	夜交藤	28	0.567 434 2	0.014 979 693	0.087 208 945
10	龙眼肉	28	0.560 064 9	0.002 472 034	0.141 255 694
11	木香	28	0.560 064 9	0.002 472 034	0.141 255 694
12	炒白术	28	0.560 064 9	0.002 472 034	0.155 277 155
13	龙胆草	24	0.552 884 6	0.005 235 679	0.085 694 434
14	栀子	24	0.552 884 6	0.005 235 679	0.085 694 434
15	柏子仁	20	0.539 062 5	0.000 000 000	0.097 982 785

度的评估计算,挖掘出中药网络的核心节点。同时,利用基于节点中心性的聚类算法 CNM-Centrality,该算法在网络划分上具有较高的精度和效率,利用该算法对中药网络进行聚类划分,计算结果能较为清晰地反映出中药之间的配伍关系。

参考文献

[1] 满姗姗,王立新. 基于数据挖掘及辨证论治探讨中药治疗肝病的用药特点 [J]. 中草药, 2016, 47(15): 2775-2778.

[2] 李泮霖,苏薇薇. 网络药理学在中药研究中的最新应用进展 [J]. 中草药, 2016, 47(16): 2938-2942.

[3] González M C, Barabási A L. Complex networks: From data to models [J]. *Nat Phys*, 2007, 3(4): 224-225.

[4] 侯 灿. 用系统方法探讨中医证的本质 [J]. 中国中西

医结合杂志, 2007, 27(5): 461-465.

[5] Freeman L C. Centrality in social networks conceptual clarification [J]. *Soc Networks*, 1979, 1(3): 215-239.

[6] Bonacich P. Some unique properties of eigenvector centrality [J]. *Soc Networks*, 2007, 29(4): 555-564.

[7] Hu F, Liu Y H. A new algorithm CNM-Centrality of detecting communities based on node centrality [J]. *Physica A*, 2016, 446: 138-151.

[8] Page L, Brin S, Motwani R, et al. The PageRank citation ranking: Bringing order to the web [R/OL]. Stanford InfoLab, 1999. <http://ilpubs.stanford.edu:8090/422>.

[9] Xu X, Yuruk N, Feng Z, et al. Scan: A structural clustering algorithm for networks [A] // Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining [C]. New York: ACM , 2007.