

基于超图的中药方剂超网络中药材群组信息挖掘

单雨濛^{1,2}, 张科^{1,2*}, 胡文军^{1,2}, 何英³, 郑泽宇¹

1. 湖州师范学院信息工程学院, 浙江 湖州 313000

2. 湖州师范学院 浙江省现代农业资源管理与应用研究重点实验室, 浙江 湖州 313000

3. 湖州师范学院医学院、护理学院, 浙江 湖州 313000

摘要: **目的** 应用基于超图的超网络理论, 构建中药方剂超网络并探究其群体的拓扑特性, 以期挖掘药材的群组信息及其高阶关联关系。 **方法** 收集整理书籍《实用三味中医方药》和中医药信息数据库 (TCM-ID) 中的中药方剂, 分别得到 2 个中药方剂数据库。以每首中药方剂作为超边, 相应方剂中出现的每味药材作为节点, 构建中药方剂超网络模型。通过计算中药方剂超网络的拓扑指标, 重点对组度序列及组度分布规律进行分析, 并借助 Python 软件中的 Pyplot 功能呈现可视化结果。 **结果** 2 个数据集共收录 9 234 首方剂, 根据 2 个数据集的不同来源分别构建均匀超网络模型和非均匀超网络模型。其中, 均匀超网络模型中共有 2 719 首方剂, 涉及 1 404 味药材, 频数最大的二元药材群组是 {黄连, 黄芩}, 黄芩虽然是二元药材群组中的重要组成药材, 但却未出现在排名前 10 的单味药材中; 非均匀超网络模型中共有 6 515 首方剂, 涉及 2 381 味药材, 频数最大的三元药材群组是 {黄连, 黄芩, 甘草}, 黄连虽然是排名第 1 的三元药材群组中的重要组成药材, 但其单味药材排名却是第 7。2 个中药方剂超网络的各类组度分布服从幂律分布, 表明药材群组的分布具有无标度特性, 提示频数越大的药材群组越重要。 **结论** 中药方剂中的药材从药理上具有互补性, 仅评估单味药材或药对重要性的视角较单一, 使用基于超图的超网络对中药方剂系统进行建模, 突破了使用普通复杂网络建模中基于点对关联的局限, 可更好地描述中药方剂这一现实复杂系统中的高阶交互关系, 从而能够有效地挖掘出中药方剂中重要的药材群组, 且组度值大的药材群组中药材之间具有紧密的多重关联性可为新药研发和创新用药等提供理论参考。

关键词: 超图; 中药方剂超网络; 组度; 组度分布; 无标度分布; 药材群组

中图分类号: G350; R283.21 **文献标志码:** A **文章编号:** 0253-2670(2024)11-3816-09

DOI: 10.7501/j.issn.0253-2670.2024.11.023

Mining of group information of medicinal materials in traditional Chinese medicine prescriptions hypernetwork based on hypergraph

SHAN Yumeng^{1,2}, ZHANG Ke^{1,2}, HU Wenjun^{1,2}, HE Ying³, ZHENG Zeyu¹

1. School of Information Engineering, Huzhou University, Huzhou 313000, China

2. Zhejiang Provincial Key Laboratory of Smart Management & Application of Modern Agricultural Resources, Huzhou University, Huzhou 313000, China

3. School of Medicine & Nursing, Huzhou University, Huzhou 313000, China

Abstract: Objective The hypernetwork theory based on hypergraph is applied to construct the hypernetwork of traditional Chinese medicine prescriptions, explore the topological characteristics of its group, and mine the group information and high-order association relationship of medicinal materials. **Methods** Sorting out the prescriptions in the book "Practical Three-flavor Traditional Chinese Medicine Prescriptions" and traditional Chinese medicine information databas (TCM-ID), and obtaining two traditional Chinese medicine prescription databases, respectively. Each traditional Chinese medicine prescription as super edge, every medicinal material as the node to build a hypernetwork model of traditional Chinese medicine prescriptions. By calculating the topological index of the hypernetwork, the composition sequence and distribution law are analyzed, and the visual results are presented with the help of Pyplot function in Python

收稿日期: 2023-02-23

基金项目: 国家自然科学基金基金区域创新发展联合基金 (U20A20228); 湖州市科技计划自然科学基金项目 (2022YZ53); 湖州师范学院研究生科研创新项目课题 (2024KYCX74)

作者简介: 单雨濛, 女, 硕士研究生, 研究方向为超网络、数据挖掘。E-mail: 1434194591@qq.com

*通信作者: 张科, 男, 讲师, 研究方向为复杂网络、超网络、数据挖掘。E-mail: hbsanli@163.com

software. **Results** A total of 9 234 prescriptions are included in the two datasets, and uniform and non-uniform hypernetworks models were constructed depending on the different sources of the two datasets. In the uniform hypernetwork model, there were 2 719 prescriptions and 1 404 herbs, and Huanglian (*Coptidis Rhizoma*) and Huangqin (*Scutellariae Radix*) were the largest frequent herbs in the two-tuple group. Although *Scutellariae Radix* was an important herb in the two-tuple group, it did not appear in the top 10 single herbs. In the non-uniform hypernetwork model, there were 6 515 prescriptions and 2 381 herbs. The three tuple herbal groups with the largest frequency were *Coptidis Rhizoma*, *Scutellariae Radix* and Gancao (*Glycyrrhizae Radix et Rhizoma*). Although *Coptidis Rhizoma* was important herb in the three-tuple herbal group ranked first, its single herbal medicine ranking was the seventh. The group degree distribution of the two traditional Chinese medicine prescription hypernetwork obeys the power-law distribution, indicating that the distribution of drug group has the scale-free property, which mean that the drug group with greater frequency was more important. **Conclusion** The medicinal herbs in the prescription system are pharmacologically complementary, the perspective of evaluating the importance of single medicinal materials or herb pairs is relatively simple, the use of hypergraph-based hypernetwork modeling of traditional Chinese medicine prescription system breaks through the limitation of point-pair association in ordinary complex network modeling. It is conducive to better describe the high-order interactions in the real-world complex system of prescriptions and effectively excavate the important medicinal material groups in the prescriptions. In addition, there are close multiple correlations among the medicinal materials in the medicinal materials groups with large grouping value, which can provide theoretical reference for the research and development of new prescriptions and innovative drugs. **Key words:** hypergraph; traditional Chinese medicine prescriptions hypernetwork; collective degree; distribution of collective degree; scale-free distribution; traditional Chinese medicinal materials group

中医药是中华文明的瑰宝^[1], 在中华民族数千年历史长河中提供了独特的医药理论和方法体系^[2-3]。目前, 对中医药的研究受到了越来越多的关注, 即使是在人工智能等热门研究领域也涌现出相关的研究成果, 如 Liu 等^[4]提出了一种 2 阶段的迁移学习模型, 从病历和中医文献资源中生成中医处方。

中药方剂是一个复杂系统, 复杂网络是研究复杂系统的重要工具。在网络科学视域内, 已有众多研究成果使用网络技术对中药方剂的配伍规律以及“病-症-药”关联关系进行分析, 对于指导中药新方开发和临床诊治等具有重要意义。随着对中医药的深入研究, 学者们发现方剂中的药与药、药与病症等存在大量模糊、非线性的关系, 这种关系可以映射为复杂网络^[5]。复杂网络是对实际复杂系统的抽象, 用于刻画系统中个体间的相互作用关系, 是研究复杂系统性质和功能的基础工具^[6]。周雪忠等^[7]利用复方药物配伍的无尺度网络规律, 实现了基于图论网络分析的处方核心药物配伍知识发现; 王世琤等^[8]基于复杂网络技术和点式互信息分析慢性肾脏病本虚标实证中药配伍规律。复杂网络理论已广泛地应用于解决中医药领域中的诸多问题^[9-11]。

药材群组是指 2 种及以上药材的组合。每个药材群组的药材组成不同, 功能也不尽相同^[12]。根据 2 种药材是否包含在同首方剂中的二元关联关系构建的普通复杂网络模型, 难以直观地揭示多种方剂中存在的药材之间的高阶复杂关联关系, 普通复杂网络不能全面刻画和揭示方剂网络中药材群组

信息及其内在规律, 基于超图的超网络方法的相关研究应运而生^[13]。超图允许由多个节点组成更一般的交互^[14], 可以更好地描述中药方剂中存在的药材群组之间的高阶复杂关联关系。Estrada 等^[15]认为基于超图拓扑结构构建的网络为超网络 (hypernetwork), 超网络模型与之前研究过的大多数复杂网络具有相似的定性特征^[16]。从理论上讲, 超图可以推广一般图上的某些结论^[17]。关于超网络的研究呈现出了快速发展的趋势, 吸引了大量学者从交叉应用的角度展开深入研究。Johnson^[18]认为超网络提供了一种表示多层次系统的新方法, 其目标是整合它们的微观和宏观动态, 如 Percy 等^[19]将生物代谢中渗流过程的概念扩展到超网络, 采用超网络的形式来研究细菌代谢超网络的鲁棒性; Pan 等^[20]将循环特征转移到超链接预测算法中, 提出了一种基于循环的超链接预测方法。在中医药领域运用超网络理论和方法的研究处于探索阶段, 不同于俞成诚等^[21]构建的基于图的超网络 (supernetwork) 的分析方法, 符康等^[22]基于超图理论建立中医药方剂网络, 对重要的单味药材或药对进行挖掘。

本研究运用基于超图的超网络对中药方剂中药材的多元关联进行建模, 将药材映射为节点, 方剂映射为超边, 在保证节点同质性的同时, 能有效地显示众多中药方剂中不同群组规模药材的高阶关联关系, 有利于系统地识别出核心的药材群组及药材之间的相互作用模式, 为中药方剂系统中的天然药材群组信息挖掘提供科学方法, 以期探究中药方剂作用机制及临床研发提供参考。

1 资料与方法

1.1 资料收集

本研究使用的数据来源于《实用中医三味药方》^[23]和 TCM-ID 中医药信息数据库 (<https://www.bidd.group/TCMID/>)；前者收集整理了中药方剂 2 719 首，后者收集整理来自包括《中国药典》、经典中药处方以及国家药品监督管理局批准的中药方剂共计 7 443 首。

1.2 数据采集与规范

纳入中药方剂的基本信息包括方剂名称和组成药材。若含中药提取物则将该中药提取物转换成对应的中药名称。名称相同的方剂只保留 1 首。排除方剂组成不完整或为单味药的方剂以及药味数大于 15 的方剂。参照《中国药典》2020 年版^[24]和全国中医药行业高等教育“十二五”规划教材《中药学》^[25]对纳入处方药物名称及性味归经进行规范化处理。同一药物因习惯或地域不同具有多种名称者进行统一，如“法半夏”“制半夏”和“姜半夏”均统一为“半夏”。同种药材名称有差异的进行规范化处理，“白芍药”规范为“白芍”，“仙灵脾”规范为“淫羊藿”，“山茱萸”规范为“山萸肉”等。炮制前后功效无明显差异者仍用生药名称，如生附子、熟附子统称为“附子”；功效差异较大的则分别录入，如“生地黄”和“熟地黄”。

将规范后的数据进行编码并建立 Excel 表，即得到所构建的中药方剂超网络的关联矩阵。方剂数据库的收集和整理由 2 名研究人员共同完成，然后独立进行数据审核，保证不一致的数据记录占比控制在 3% 以下。根据研究计划和内容，对数据进行集成、清洗和预处理等。按数据的来源分别建立数据集 1 和数据集 2。

1.3 中药方剂超网络模型构建

本研究对象是基于超图的中药方剂超网络，其拓扑结构采用了超图作为数学表示形式^[17]。设节点集合 $V = \{v_1, v_2, v_3, \dots, v_n\}$ ，超边集合 $\varepsilon = \{E_1, E_2, E_3, \dots, E_m\}$ 均是有限集合，且 $E_i \neq \emptyset$ ， $\bigcup E_i = V$ ，则称 V 和 ε 之间存在二元关系 H ，则 H 是一个超图。

为了说明中药方剂超网络的构建方法，本研究以 5 首中药方剂为例构建小规模的中药方剂超网络 H_L ，如图 1 所示。这 5 首方剂的名称及其各自的组成药材分别为：(1) 麻黄汤，由麻黄、桂枝、甘草、杏仁组成；(2) 大陷胸丸，由葶苈子、芒硝、杏仁、大黄组成；(3) 桂枝汤，由桂枝、炙甘草、白芍、

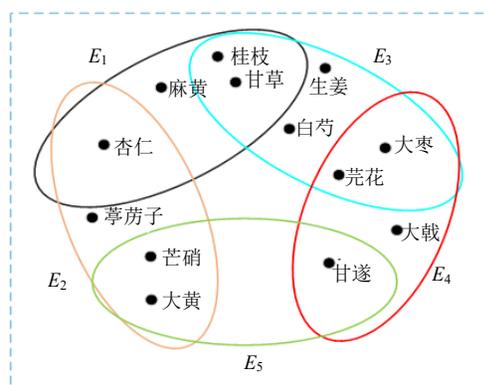


图 1 中药方剂超网络

Fig. 1 Traditional Chinese medicine prescription hypernetwork

生姜、大枣组成；(4) 十枣汤方，由大枣、芫花、大戟、甘遂组成；(5) 大陷胸汤，由芒硝、大黄、甘遂组成。将每一首中药方剂都作为 1 条超边（超边用封闭的曲线表示），将相应方剂中出现的每味药材作为节点，可得到图 1 中具有 13 个节点和 5 条超边的中药方剂超网络 H_L 。其中，节点用符号 v 来表示，超边用符号 E 表示。

在图 1 所示的超网络中，超边 E_1 、 E_2 、 E_3 、 E_4 和 E_5 分别表示麻黄汤、大陷胸丸、桂枝汤、十枣汤方和大陷胸汤。结合超图理论的基本知识，易知图 1 中每条超边的节点数和不同节点所属的超边数。然而，由于桂枝和甘草这 2 味药材同时出现在麻黄汤和桂枝汤中，意味着桂枝和甘草 2 个节点既存在于超边 E_1 中，也存在于超边 E_3 中；同理表示大黄和芒硝的 2 个不同节点既存在于超边 E_2 中，也存在于超边 E_5 中。超网络 H_L 体现了药材群组信息，需要新的方法进行信息挖掘。

1.4 超网络拓扑结构特征

在超网络中，节点超度表示该节点存在于多少条超边中，即其被包含的超边数目。在超网络的关联矩阵中，节点的超度也可通过统计相应行中非零元素的个数来计算。超度分布是节点超度的概率分布或频率分布，表示为超网络中超度的对应节点数量在整体节点总数中所占比例。为了分析中药方剂超网络中药材的群组信息，超网络中新的挖掘群体信息的概念介绍如下。

1.4.1 紧密相关集 (tightly related set)^[26] 设 $H = (V, \varepsilon)$ 是具有 m 条超边的 n 阶超图，若存在超边 E_i ($i \in 1, 2, \dots, m$) 使得集合 F 是 E_i 的非空子集，则称 F 是超图 H 的 1 个紧密相关集。超图 H 的所有紧密

相关集组成的集合记为 $\Phi_c(H)$ 。特别地, 当 F 的元素个数为 t 时, 称其为超图 H 的 t 元紧密相关集, H 的所有 t 元紧密相关集组成的集合记为 $\Phi_t(H)$ 。

基于超边 E_i 的 t 元紧密相关集的数量为 $s_t(E_i) = \binom{|E_i|}{t}$ 。超图 H 的 t 元紧密相关集为 $\Phi_t(H) = \bigcup_{i=1}^m \binom{E_i}{t}$, 其中, $\binom{E_i}{t}$ 表示基于超边 E_i 的 t 元紧密相关集。

1.4.2 组度 (collective degree) [26] 在超网络 H 中, 相关集 F 的组度 d_F 表示与紧密相关集 F 相关联的超边数。若 $F = \{v_{i_1}, v_{i_2}, \dots, v_{i_t}\}$, 则 d_F 记为 $d_{v_{i_1}, v_{i_2}, \dots, v_{i_t}}$ 。

1.4.3 组度分布 (collective degree distribution) [26] 在超网络 H 中, t 元组度分布 $p_t(k)$ 是组度为 k 的 t 元紧密相关集个数 $N_t(k)$ 占有所有 t 元紧密相关集总数 N_t 的比例, 具体公式为 $p_t(k) = \frac{N_t(k)}{N_t}$ 。

1.4.4 完全组度分布 [26] 在超网络 H 中, 完全组度分布 $p_c(k)$ 是组度为 k 的紧密相关集的数量 $N_t(k)$ 占有所有相关集总数 N_c 的比例, 具体公式为 $p_c(k) = \frac{N_c(k)}{N_c}$ 。

1.5 数据分析

利用收集的中药方剂数据集, 依据中药方剂超网络的构建方法, 使用 NumPy 库处理多维数组和矩阵, 得到对应超网络的关联矩阵。采用 Python 3.10 软件进行数据分析, 分别对超度、超度分布、 t 元组度、 t 元组度分布, 以及完全分布这些拓扑指标进行计算。将 Pandas 库导入 Python 3.10 中对计算结果进行处理, 并运用 Matplotlib 库中的 Pyplot 模块创建静态、交互性的网络图, 从而对结果进行可视化展示。

超图的特点是允许多个节点组成 1 条超边, 从而形成更为丰富和复杂的关联结构, 能为群组关系的描述提供最一般且无约束的数学表示 [26]。组度可以反映超网络中小群体的局部特性, 从而有利于挖掘出多种药材间潜在的、有价值的依赖关系。

2 结果

2.1 数据筛选结果

本研究创建 2 个数据集, 共收集 10 162 首中药方剂数据, 对数据进行清洗及规范化处理后最终得到 9 234 条有效数据。数据集基本指标统计如表 1 所示。

表 1 数据集基本指标统计

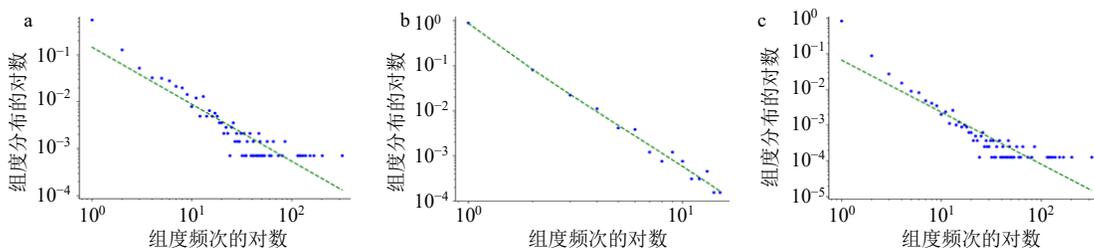
| 数据集 | 来源 | n | m |
|-------|------------|-------|-------|
| 数据集 1 | 《实用中医三味药方》 | 1 404 | 2 719 |
| 数据集 2 | 中医药信息数据库 | 2 381 | 6 515 |

2.2 均匀中药方剂超网络分析

由数据集 1 构建均匀中药方剂超网络 H_{sw} 。其中, 以相关的 1 404 味药材作为节点, 以这些药材组成的 2 719 首方剂为超边。因为每首方剂均含有 3 味药材, 所以超网络 H_{sw} 是均匀的。

2.2.1 超网络 H_{sw} 的组度分布规律分析 计算相关集的组度、组度分布和完全组度分布, 然后在双对数坐标下对超网络节点组度分布进行可视化, 最后用最小二乘法进行拟合。超网络 H_{sw} 的组度分布及线性拟合见图 2, 其中横坐标表示组度 (一元组度即超度) 频次的对数, 纵坐标表示组度分布的对数。

由图 2-a 可知, 超网络 H_{sw} 的超度分布呈现出明显的幂律分布特性。由图 2-b 可知, 超网络 H_{sw}



a-超度分布图; b-二元组度分布图; c-完全分布图。

a-hyper distribution map; b-binary degree distribution map; c-full distribution map.

图 2 超网络 H_{sw} 的双对数坐标系节点超度分布拟合曲线图

Fig. 2 Fitted curves of nodal hyperdegree distribution in double logarithmic coordinate system of hypernetwork H_{sw}

的二元组度分布呈现出明显的幂律分布特性。由图 2-c 可知,均匀超网络 H_{sw} 的完全组度分布也呈现出明显的幂律分布特性。由图 2 中的 3 个线性拟合结果可以看出,超网络 H_{sw} 的 3 个不同类型的组度分布都可以用最小二乘法拟合出 1 条直线,意味着每个分布都具有长尾效应。说明只有少部分节点(集)的组度较大,而大部分节点(集)的组度相对较小,表现出无标度特性。

2.2.2 超网络 H_{sw} 的高频药材群组分析 由组度分布规律研究结果可知,超网络 H_{sw} 规模不同的组度分布遵循幂律分布,是不均匀的。从而组度越大的药材集合在方剂超网络 H_{sw} 中的影响力越大。依据构建超网络 H_{sw} 的方法可知,组度即为对应药材群组被包含的方剂的首数。

本研究分别对影响力较大的不同规模的药材群组进行详细分析。超网络 H_{sw} 中超度排名前 10 的药材见图 3,它们都是十分常见的中药材。甘草是超度最大的药材,超度为 322,表明甘草出现在相应数据集的 322 首方剂中。甘草有清热解毒、去痰止咳、补脾益气、缓急止痛、调和诸药的功效^[27],其种植和应用非常广泛。超度排名 2~5 名的依次为黄连、当归、大黄和人参。排名第 10 的黄柏的组度也高达 86。

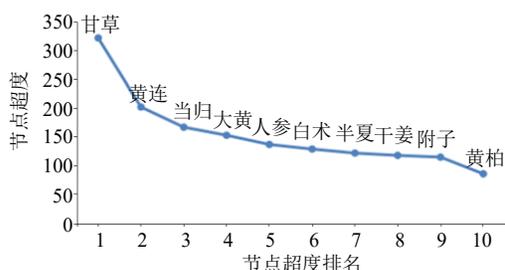


图 3 中药方剂超网络 H_{sw} 中超度前 10 的药材
Fig. 3 Top 10 herbs in traditional Chinese medicine prescription hypernetwork H_{sw}

组度 ≥ 7 的 25 个二元药材群组的词云图见图 4。排名第 1 的二元药材群组是{黄连, 黄芩}, 组度为 15, 表明黄连和黄芩同时包含在 15 首方剂中, 这 2 味药材配伍在相应方剂数据集中出现的频率最高。黄芩味苦、性寒; 黄连性苦、性寒; 2 味药皆以清热燥湿、泻火解毒为主, 常于方剂中配伍使用^[28]。排名第 2 的二元药材群组是{干姜, 附子}, 组度为 14, 表明干姜和附子同时包含在 14 首方剂中。干姜味辛, 性温、大热, 有辛散里寒、温助中阳的功效^[29]; 附子辛热燥烈, 补火散寒, 有温通周身阳气的功效^[30]。



图 4 超网络 H_{sw} 中二元药材群组词云图

Fig. 4 Word cloud map of two tuple traditional Chinese medicinal materials group in hypernetwork H_{sw}

它们常配伍使用,如含有这 2 味药材的方剂姜附汤, 主要治疗脾虚腹胀、呕吐痰饮或食不进等症状^[31]。排名第 3 的二元药材群组有{甘草, 人参}和{大黄, 甘草}, 组度均为 13, 表明这 2 对组合同时出现在 13 首方剂中。人参甘、微苦, 有益气健脾、燮理药性的功效^[29]; 大黄有下瘀血、调中化食及安和五脏的作用^[32]。以甘草和人参为主的方剂温中丸, 主要治疗中气虚热、不喜饮冷或肢体倦怠等症状^[31]。以大黄和甘草为主的方剂大黄汤, 主要治疗大便不畅或散风活血等症状^[31]。

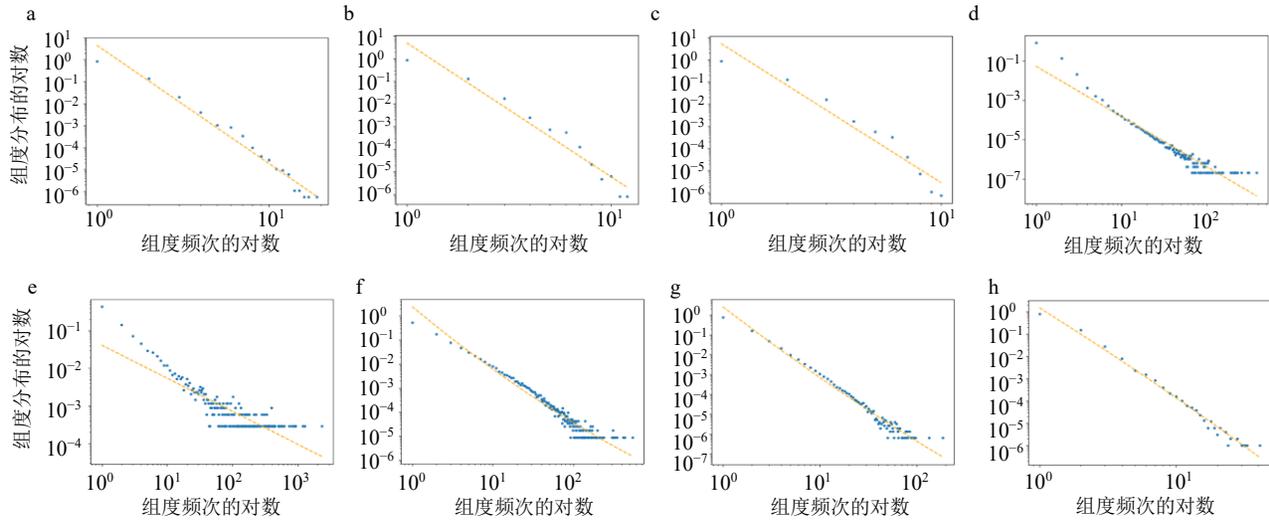
综上分析可知,黄芩和附子虽然是排名前 2 的二元药材群组的重要组成药材,但是这 2 味药都没有出现在超度排名前 8 的药材中。当归虽然是超度排名第 3 的药材,但是却没有出现在组度排名前 3 的二元药材群组中。

2.3 非均匀中药方剂超网络分析

由数据集 2 构建非均匀中药方剂超网络 H_{TC} , 以相关的 2 381 味药材作为节点, 以这些药材组成的 6 515 首方剂为超边。因每首方剂均含有的药材数量 >1 且 <16 , 所以超网络 H_{TC} 是非均匀的。通过计算可知,其超边的平均节点数为 8.98。该数据集相较于数据集 1 规模更大。

2.3.1 超网络 H_{TC} 组度分布规律分析 图 5 为非均匀超网络 H_{TC} 在双对数坐标系下的组度分布和完全分布, 以及用最小二乘法进行线性拟合的示意图。其中横坐标表示组度(一元组度即超度)频次的对数, 纵坐标表示组度分布的对数。

由图 5-a 可知,超网络 H_{TC} 的超度分布呈现出明显的幂律分布特性。由图 5-b~g 可知,超网络 H_{TC} 的二至七元组度分布也都呈现出明显的幂律分布特性。由图 5-h 可知,超网络 H_{TC} 的完全分布也呈现出明显的幂律分布特性。由图 5 中的 8 个线性拟合结果可以看出,超网络 H_{TC} 的 8 个不同类型的组度



a-超度分布图; b-二元超度分布图; c-三元超度分布图; d-四元超度分布图; e-五元超度分布图; f-六元超度分布图; g-七元超度分布图; h-完全分布图。

a-hyper distribution map; b-binary degree distribution map; c-triplet degree distribution map; d-quadruple degree distribution map; e-quintuple degree distribution map; f-hexuple degree distribution map; g-heptuple degree distribution map; h-complete distribution map.

图 5 超网络 H_{TC} 的双对数坐标系节点超度分布拟合曲线图

Fig. 5 Fitted curves of nodal hyperdegree distribution in double logarithmic coordinate system of hypernetwork H_{TC}

分布都可以用最小二乘法拟合出 1 条直线，且尾部节点分布较多，说明只有少部分节点（集）的组度较大，而大部分节点（集）的组度相对较小，表现出无标度特性。

2.3.2 超网络 H_{TC} 的高频药材群组分析 由组度分布规律结果可知，超网络 H_{TC} 的规模不同的组度的分布遵循幂律分布，是不均匀的。从而组度越大的药材集合在方剂超网络 H_{TC} 的中影响力越大。

本研究分别对影响力较大的不同规模的药材群组进行详细分析。非均匀超网络 H_{TC} 中超度排名前 20 的药材见图 6。其中，甘草是超度最大的药材，超度为 2353。超度排名 2~5 的依次为乳香、黄芩、牡蛎和黄连。排名第 20 的茴香的组度也高达 238。

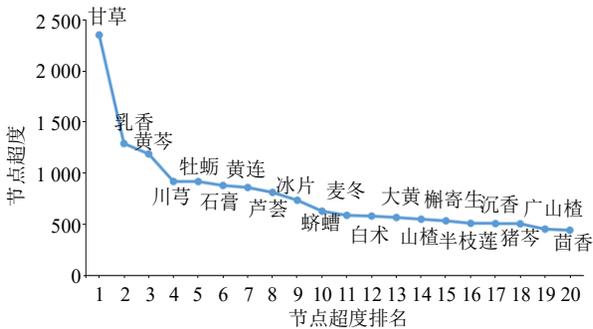


图 6 中药方剂超网络 H_{TC} 中排名前 20 的药材

Fig. 6 Top 20 herbs in traditional Chinese medicine prescription hypernetwork H_{TC}

组度排名前 5 的二至七元药材群组见表 2。本研究分别对不同规模药材群组的组度排名第 1 的群组进行分析。组度排名第 1 的二元药材群组是 {黄芩, 甘草}, 组度为 544, 表明黄芩和甘草这 2 味药同时包含在 544 首方剂中。它们常搭配使用, 如包含这 2 味药的清肺排毒汤具有抗炎和抗病毒等作用^[33]。组度排名第 1 的三元药材群组是 {黄连, 黄芩, 甘草}, 组度为 187, 表明黄连、黄芩和甘草这 3 味药同时包含在 187 首方剂中。它们常配伍使用, 如含有这 3 味药的方剂甘草泻心汤主要治疗脾胃虚弱和呕吐等症^[34]。组度排名第 1 的四元药材群组是 {党参, 白术, 茯苓, 甘草}, 组度为 41, 表明党参、白术、茯苓和甘草这 4 味药同时包含在 41 首方剂中。党参性甘, 有补中益气等功效; 白术味苦, 性甘、温, 归脾、胃经, 具有健脾益气、燥湿利水之功效; 茯苓药性甘淡平, 有健脾宁心、利水渗湿等功效。它们常配伍使用, 如含有这 4 味药的方剂八珍汤, 主要治疗脾虚和腹泻等症状^[31]。组度排名第 1 的五元药材群组是 {白术, 茯苓, 甘草, 生姜, 大枣}, 组度为 19, 表明白术、茯苓、甘草、生姜和大枣这 5 味药同时包含在 19 首方剂中。生姜有解表散寒、温中止呕和温肺止咳的作用; 大枣有补中益气 and 养血安神的作用。它们常配伍使用, 如含有这 5 味药的

表 2 超网络 H_{TC} 组度排名前 5 的二至七元药材群组

Table 2 Two to seven tuple traditional Chinese medicinal materials groups ranked in top five groups degree of hypernetwork H_{TC}

| 序列 | 1 | 2 | 3 | 4 | 5 |
|------|------------------------------|------------------------------|------------------------------|------------------------------|---------------------------------|
| 二元组度 | {黄芩, 甘草} | {黄连, 甘草} | {乳香, 甘草} | {乳香, 川芎} | {芦荟, 甘草} |
| 三元组度 | {黄连, 黄芩, 甘草} | {乳香, 牡蛎, 甘草} | {甘草, 生姜, 苍术} | {乳香, 黄连, 甘草} | {乳香, 黄芩, 甘草} |
| 四元组度 | {党参, 白术, 茯苓, 甘草} | {白术, 茯苓, 甘草, 生姜} | {茯苓, 甘草, 生姜, 大枣} | {白术, 甘草, 生姜, 大枣} | {白术, 茯苓, 当归, 白芍} |
| 五元组度 | {白术, 茯苓, 甘草, 生姜, 大枣} | {白术, 茯苓, 甘草, 当归, 白芍} | {白术, 茯苓, 当归, 白芍, 熟地黄} | {党参, 白术, 茯苓, 当归, 白芍} | {柴胡, 当归, 白芍, 茯苓, 薄荷} |
| 六元组度 | {羌活, 防风, 苍术, 白芷, 黄芩, 甘草} | {羌活, 防风, 苍术, 川芎, 黄芩, 甘草} | {党参, 白术, 茯苓, 甘草, 当归, 白芍} | {党参, 白术, 茯苓, 甘草, 当归, 川芎} | {党参, 白术, 茯苓, 当归, 白芍, 熟地黄} |
| 七元组度 | {川芎, 白芷, 羌活, 细辛, 防风, 薄荷, 甘草} | {川芎, 白芷, 羌活, 细辛, 防风, 荆芥, 甘草} | {羌活, 防风, 苍术, 细辛, 川芎, 白芷, 甘草} | {羌活, 防风, 苍术, 细辛, 川芎, 白芷, 甘草} | {金银花, 连翘, 薄荷, 牛蒡子, 桔梗, 淡竹叶, 甘草} |

方剂六君子汤, 主要治疗气血两虚、神疲肢倦和食欲不振等症状^[31]。组度排名第 1 的六元药材群组是 {羌活, 防风, 苍术, 白芷, 黄芩, 甘草}, 组度为 12, 表明羌活、防风、苍术、白芷、黄芩和甘草这 6 味药同时包含在 12 首方剂中。羌活和防风有解表散寒和祛风胜湿的作用; 苍术有燥湿健脾和祛风散寒的作用; 白芷有解表散寒、祛风止痛、通鼻窍和燥湿止带的作用。它们常配伍使用, 如含有这 6 味药的方剂九味羌活汤, 主要治疗感冒、发烧等症状^[29]。组度排名第 1 的七元药材群组是 {川芎, 白芷, 羌活, 细辛, 防风, 薄荷, 甘草}, 组度为 10, 表明这 7 味药同时包含在 10 首方剂中。川芎有活血行气和祛风止痛的作用; 细辛有解表散寒、祛风止痛和温肺化饮的作用; 薄荷有疏散风热、清利头目、利咽透疹和疏肝行气的作用。它们常配伍使用。如含有这 7 味药的方剂金不换膏, 有祛风散寒和活血止痛的功效^[31]。

为了直观地显示超网络 H_{TC} 的药材群组的频数大小, 使用词云技术展示不同规模药材群组的词云图。图 7 为组度大于 169 的 28 个二元药材群组词云图, 图 8 为组度大于 64 的 24 个三元药材群组词云图。

由上述分析可知, 超度排名第 2 的乳香和第 5 的牡蛎, 均没有出现在排名第 1 的二至七元药材群组中。超度排名第 3 的黄芩也没有出现在排名第 1 的四、五和七元药材群组中。川芎超度排名第 4, 但没有出现在二至五元药材群组排名前 5 的所有群组中。黄连超度排名第 7, 但是却出现在排名第 2 的二元药材群组中以及排名第 1 的三元药材群组中。



图 7 超网络 H_{TC} 中二元药材群组词云图

Fig. 7 Word cloud map of two tuple traditional Chinese medicinal materials group in hypernetwork H_{TC}



图 8 超网络 H_{TC} 中三元药材群组词云图

Fig. 8 Word cloud map of three tuple traditional Chinese medicinal materials group in hypernetwork H_{TC}

3 讨论

中医药全面振兴已成为国家战略, 很多新的科学技术与方法已广泛地应用于中医药研究中, 其中在中医药信息挖掘方面, 复杂网络理论是分析和处理传统中药方剂数据的有效方法。方剂是依据病情在辨证立法的基础上遵循“君、臣、佐、使”的基本组织结构, 选择合适的药物配伍而成, 含有丰富的复杂性规律^[35]。依据丰富的中医药数据进行信息挖掘, 对于阐明方剂配伍的科学内涵、完善中药药性理论和指导中医药新方剂开发等具有深刻意义^[36]。

本研究通过基于超图的超网络模型对方剂间多

元的药材群组进行分析。在探索药材群组信息时，将每首方剂视为超边，每种药材视为节点，多种药材同时使用可以看作它们之间存在高阶交互进而构成核心药材群组。构建中药方剂超网络模型，能更好地理解中药方剂的配伍规律以及中药材之间的相互作用模式。

研究表明，通过基于超图的超网络方法建模，能够挖掘出中药方剂和药材之间更多的隐藏信息，特别是包含药材味数大于2的群组信息。构建均匀超网络模型结果显示，甘草是使用频率最高的单一药材，{黄连，黄芩}是最常用的二元药材群组。构建非均匀超网络模型结果同样显示甘草是最常用的单一药材，{黄芩，甘草}是最常用的二元药材群组，{黄连，黄芪，甘草}是最常用的三元药材群组，{党参，白术，茯苓，甘草}是使用最多的四元药材群组。通过在双对数坐标系下进行药材组群分布规律统计，可知2个超网络模型的组度分布均遵循幂律分布，具有无标度特性，意味着对应方剂数据库中出现频率越高的药材组群越重要。探究二元药材群组或三元药材群组乃至更多元的药材群组的配伍使用，对中药方剂的配伍规律和中药材属性的研究具有重要意义，可为遣药组方等提供理论参考，对于医生临床组方等也能起到辅助作用。

传统复杂网络方法在处理中药方剂时难以有效地捕捉到多个药材同时出现在方剂中的情况。超网络突破了描述点对关联的局限，能够有效地描述中药方剂这一现实复杂系统具有的高阶交互关系。运用超网络的理论和方法对中药方剂系统进行建模，通过拓扑特性研究对应超网络结构功能有利于挖掘中药方剂系统中的组群信息。本研究在处理高阶的复杂关联关系具有一定的系统性和普适性，可用于对中药方剂系统的深入研究。

利益冲突 所有作者均声明不存在利益冲突

参考文献

[1] 国家中医药管理局. 中华文明的瑰宝习近平这样说中医药 [EB/OL]. (2019-10-26) [2024-01-31]. <http://www.satcm.gov.cn/xinxifabu/meitibaodao/2019-10-26/11636.html>.

[2] Normile D. Asian medicine. The new face of traditional Chinese medicine [J]. *Science*, 2003, 299(5604): 188-190.

[3] 宋大涵. 中华人民共和国中医药法 [M]. 北京: 法律出版社, 2017: 56-58.

[4] Liu Z, Luo C, Fu D, *et al.* A novel transfer learning model for traditional herbal medicine prescription generation from unstructured resources and knowledge [J]. *Artif*

Intellig Med, 2022, 124: 102232.

[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] 周建林, 樊瑛, 狄增如. 复杂网络进展 [J]. *北京师范大学学报: 自然科学版*, 2023, 59(5): 691-704.

[7] 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.

[8] 王世琤, 杜华, 张宁. 基于复杂网络技术和点式互信息分析慢性肾脏病本虚标实证中药配伍规律 [J]. *北京中医药*, 2020, 39(6): 548-552.

[9] Li S, Zhang B, Jiang D, *et al.* Herb network construction and co-module analysis for uncovering the combination rule of traditional Chinese herbal formulae [J]. *BMC Bioinformatics*, 2010, 11(Suppl 11): S6.

[10] 田园, 马良宵, 于文颜, 等. 应用复杂网络及数据挖掘探讨常见内脏痛针灸处方特点 [J]. *中国中医药信息杂志*, 2023, 30(6): 47-53.

[11] Wang N, Du N L, Peng Y H, *et al.* Network patterns of herbal combinations in traditional Chinese clinical prescriptions [J]. *Front Pharmacol*, 2020, 11: 590824.

[12] 杨薇, 谢雁鸣. 基于电子医疗病例的缺血性中风病药物群组模块思路探讨 [J]. *中国中药杂志*, 2014, 39(18): 3430-3434.

[13] 胡瑞学, 王正中, 李慧敏, 等. 超网络方法在医学相关领域应用的介绍与可行性分析 [J]. *现代中医临床*, 2023, 30(1): 56-60.

[14] Young J G, Petri G, Peixoto T P. Hypergraph reconstruction from network data [J]. *Commun Physics*, 2021, 4(1): 135.

[15] Estrada E, Rodríguez-Velázquez J A. Subgraph centrality and clustering in complex hyper-networks [J]. *Phys A Stat Mech Appl*, 2006, 364: 581-594.

[16] Larsen B, Aone C. Fast and effective text mining using linear-time document clustering [A] // *Proceedings of the fifth ACM SIGKDD international conference on knowledge discovery and data mining* [C]. San Diego: ACM, 1999: 16-22.

[17] Bretto A. Hypergraph theory: An introduction [J]. *Mathematical Engineering* [M]. Cham: Springer, 2013: 1-38.

[18] Johnson J. Hypernetworks and design [A] // *Hypernetworks in the Science of Complex Systems* [M]. London: Imperial College Press, 2013: 275-295.

[19] Percy N, Chuzhanova N, Crofts J J. Complexity and robustness in hypernetwork models of metabolism [J]. *J Theor Biol*, 2016, 406: 99-104.

[20] Pan L M, Shang H J, Li P Y, *et al.* Predicting hyperlinks via

- hypernetwork loop structure [J]. *EPL Europhys Lett*, 2021, 135(4): 48005.
- [21] 俞成诚, 王念, 关双, 等. 已上市中成药治疗肾阳虚证的“病-症-药”超网络分析及核心处方发现 [J]. *中草药*, 2023, 54(24): 8153-8164.
- [22] 符康, 闫光辉, 罗浩. 基于超图理论的中医方剂网络药对挖掘方法 [J]. *计算机工程与应用*, 2024, 60(10): 320-331.
- [23] 田代华. 实用三味中医药方 [M]. 北京: 人民卫生出版社, 2019: 1-749.
- [24] 中国药典 [S]. 一部. 2020.
- [25] 钟赣生. 中药学 [M]. 第3版. 北京: 中国中医药出版社, 2012: 140-144.
- [26] 张科, 高靖宇, 胡文军, 等. 超网络中的紧密相关集和组度分布特性研究 [J]. *中国科学: 物理学 力学 天文学*, 2023, 53(7): 110-126.
- [27] 王苗苗. 甘草的化学成分及生物活性研究 [D]. 济南: 山东大学, 2020.
- [28] 张博荀. 基于“肠道菌群-黏膜屏障”研究“黄芩-黄连”药对治疗 T2DM 的疗效机制及配伍效应 [D]. 成都: 成都中医药大学, 2020.
- [29] 石雪华, 郭宏伟. 干姜、人参共为理中丸君药刍议 [J]. *中医药学报*, 2023, 51(10): 82-86.
- [30] 赵庆香, 贾君叶, 余铮, 等. 麻黄细辛附子汤在新型冠状病毒感染恢复期运用心得 [J]. *浙江中医杂志*, 2023, 58(11): 856-857.
- [31] 李冀. 方剂学 [M]. 第2版. 北京: 中国中医药出版社, 2016: 18-136.
- [32] 刘朋军, 张冰仪, 司国民. 经方中大黄配伍应用规律探析 [J/OL]. *亚太传统医药*, 2024, 20(3): 202-206.
- [33] 文旭清. 基于液质联用与目标网络药理学探究黄芩-甘草药对治疗新冠病毒肺炎的作用机制 [D]. 石家庄: 河北医科大学, 2023.
- [34] 许宏平, 黄古叶, 梅露, 等. 甘草泻心汤临床研究概况 [J]. *中国民族民间医药*, 2023, 32(20): 78-83.
- [35] 梁力伟, 丁长松, 黄辛迪, 等. 基于重叠社区的“方-药”网络经方配伍规律分析 [J]. *中草药*, 2020, 51(2): 496-506.
- [36] 吕庆莉. 数据挖掘与复杂网络的融合及其在中医药领域应用 [J]. *中草药*, 2016, 47(8): 1430-1436.

[责任编辑 潘明佳]