

# 数字人文驱动下的核心作者发现 与主题挖掘研究\*

武 帅<sup>1,2</sup> 杨秀璋<sup>2</sup> 任天舒<sup>2</sup> 刘建义<sup>2</sup>

(1. 南京农业大学信息管理学院, 南京 210003;

2. 贵州财经大学信息学院, 贵阳 550025)

**摘要:** [目的/意义] 在数字人文驱动下, 从宏观层面对《红楼梦学刊》创刊四十年成果以科研者的角度进行统计分析, 从微观层面融合数据挖掘技术分析红学潜在的研究领域, 更好地促进红学研究的发展。[方法/过程] 主要运用了计量统计和主题挖掘的方法进行分析。首先计量统计出四十年红学研究的科研核心作者, 再对红学领域成果进行主题挖掘。从核心作者发现和主题演化分析两方面研究四十年红学的研究热点。[结果/结论] 红学研究整体可分为四个领域, 分别是人物关系研究、社会制度研究、当下学术研讨和《红楼梦》版本猜测。

**关键词:** 《红楼梦学刊》 主题演化 数字人文 主题挖掘 核心作者

**分类号:** G353.1

**DOI:** 10.31193/SSAP.J.ISSN.2096-6695.2023.01.05

## 0 引言

《红楼梦》作为中国四大名著之首, 是古代章回体长篇小说的代表, 被公认为中国古典小说的巅峰之作, 先后被翻拍了九次电视剧, 是研究中国古代文学的重要依据<sup>[1]</sup>。随着艺术文化的发展, 近年来红学研究和文艺评价层出不穷。红学研究者从不同角度进行分析研究, 学术成果呈现“百花齐放, 百家争鸣”之势<sup>[2]</sup>。

\* 本文系贵州省科技计划项目“基于大数据及图像识别的水族文献及濒危水书抢救性整理研究”(项目编号: 黔科合基础[2020]1Y279)、贵州财经大学项目“AI 大数据赋能的水族知识图谱构建和濒危水书文字统计及识别研究”(项目编号: 2021KYQN03)的研究成果之一。

**[作者简介]** 武帅 (ORCID: 0000-0002-1162-4308), 男, 助教, 博士研究生, 研究方向为文献计量学、信息资源管理, Email: 472191973@qq.com; 杨秀璋 (ORCID: 0000-0001-9648-9506), 男, 助教, 博士研究生, 研究方向为数字人文、主题挖掘, Email: 1455136241@qq.com (通讯作者); 任天舒 (ORCID: 0000-0001-5930-3653), 女, 硕士研究生, 研究方向为信息资源管理、文献计量学, Email: 413076113@qq.com; 刘建义 (ORCID: 0000-0003-1693-6631), 男, 硕士研究生, 研究方向为文献计量学、主题挖掘, Email: 1105053117@qq.com。

入选中文社会科学引文索引 (Chinese Social Sciences Citation Index, 简称 CSSCI) 的期刊论文, 在对应学科领域具有一定的指导作用, 代表着该领域学术性强、研究新颖、创意性高的学术成果。《红楼梦学刊》作为红学研究唯一入选 CSSCI 数据库的期刊, 四十年刊发大量期刊论文, 是红学研究的主要交流平台<sup>[3]</sup>。刊物自创刊以来一直秉承严谨的学术态度, 在追求学术性的同时兼顾普及性, 既有较高的学术水准, 又有一定的趣味性、可读性, 深受红学研究者以及文学爱好者喜爱。

《红楼梦学刊》作为红学研究的交流与传播的重要载体, 有效地推广了红学研究的发展。随着互联网技术应用于期刊出版, 研究成果层出不穷, 大量学者各抒己见, 从不同角度运用不同方法对《红楼梦》进行研究, 红学逐渐成为一种大融合的学派。

红学研究者对《红楼梦》的研究往往有着不同的见解, 从而不能精确地反映时代研究主题。针对这一不足, 本文以中国知网 (CNKI) 自 1979 年 4 月 2 日至 2019 年 4 月 2 日收录的 5 582 篇《红楼梦学刊》期刊论文数据为研究对象, 主要运用了计量统计和主题挖掘的方法进行分析。首先, 计量统计确认四十年红学研究的科研核心作者; 再次, 对红学领域成果进行主题挖掘。从核心作者发现和主题挖掘两方面探究四十年红学的研究热点。

## 1 相关研究

《红楼梦学刊》作为《红楼梦》学术研究的主要期刊, 刊发《红楼梦》相关思想研究、艺术价值、历史资料、红学研究、作者历史以及文物考证等内容, 在红学研究以及古典文学研究领域享有很高的学术声誉。《红楼梦学刊》代表着《红楼梦》研究的阶段性专业水平以及学术品位, 具有重要的参考依据以及研究价值。

### 1.1 传统文献研究现状

在大数据时代背景下, 数据挖掘技术发展迅捷, 学术研究成果层出不穷。目前, 国内外利用数据挖掘<sup>[4]</sup>、机器学习<sup>[5]</sup>算法深度挖掘期刊文献研究的相对较少。裴杰<sup>[6]</sup>运用传统文献计量方法从多维度探究《红楼梦》中日译学界研究的典型特征及发展规律。张庆善等人<sup>[7]</sup>运用传统文献计量的方法厘清红学的学科性质、划定学科范畴、建构学科框架。高淮生<sup>[8]</sup>运用传统文献计量的方法梳理红学研究历程, 发现红学每一个历程的发展都离不开文献。孙伟科等人<sup>[9]</sup>系统梳理 2017 年度红学研究成果, 发现红学研究注重多种方法的结合, 有效促进红学研究领域的整合, 进一步推动学科交叉融合。

传统的文献研究主要倾向于原著自我阅读、倾听专家讲解、核心文献阅读、参与论坛会议, 其中最常用的为核心文献阅读。核心文献阅读一般采用关键词查询以及下载量筛选, 方法比较单一, 不足以直观呈现红学研究深层次的主题, 缺乏利用数据挖掘方法研究《红楼梦学刊》的热点主题以及时间发展。

### 1.2 主题挖掘的文献研究现状

随着大数据技术的发展, 越来越多的学者发觉数据潜在价值的重要性, 致力于结合数据挖掘或机器学习的方法在海量文献数据中分析得出具有价值性的结论。沈琳<sup>[10]</sup>对文本内容

进行归纳总结, 系统建立了《红楼梦》家具的类型和组合规律。武迪等人<sup>[11]</sup>在内容上深挖《香艳丛书》中收录有关《红楼梦》的文献材料, 进一步了解红学研究的流布情况。陈骁<sup>[12]</sup>以清代《红楼梦》的图像世界为研究对象, 探讨文本与图像之间的联系, 拓宽红学数据挖掘的范畴。蔡永明等人<sup>[13]</sup>提出 CA-LDA 模型对中文短文本进行主题分析, 提升了具有相同搭配关系词汇划分同一主题的概率, 给短文本数据提供了新研究方法。武帅等人<sup>[14]</sup>运用数据计量和社交网络的分析方法梳理了图书情报学科发展。杨秀璋等人<sup>[15-16]</sup>运用文献计量和社交网络的分析方法梳理了清水江流域文献发展情况, 并利用综合指数和知识图谱发掘核心作者。

运用数据挖掘的方法深挖文献数据, 可以发现数据潜在价值。本文采用计量统计和主题挖掘的方法探究四十年《红楼梦学刊》期刊论文, 一定程度能够体现出《红楼梦》研究的阶段性专业水平以及学术品位。

## 2 研究框架

本文旨在对《红楼梦学刊》四十年刊发的 5 582 篇论文进行分析研究, 挖掘上述论文的高被引情况、核心作者、主要研究机构以及核心主题, 具体分析流程分为四步, 详见图 1。

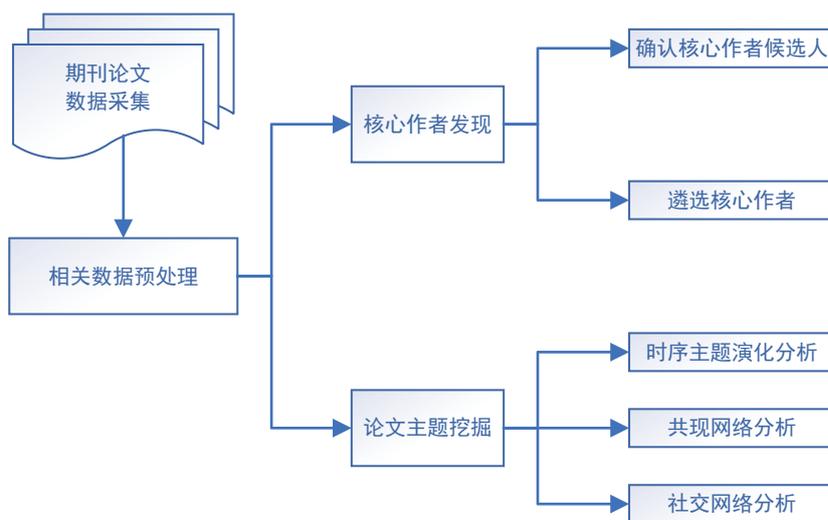


图 1 《红楼梦学刊》期刊论文分析流程图

(1) 基于 Python 环境下 Selenium 模块自定义爬取中国知网收录的《红楼梦学刊》期刊论文, 并存为 csv 文件。

(2) 将《红楼梦学刊》相关论文数据进行数据预处理, 主要包括: 数据清洗、关系抽取以及异常值处理, 并存为 csv 文件。

(3) 结合被引用量、发文量发现核心作者, 主要包括: 普赖斯定律预选核心作者候选人以及综合指数遴选核心作者。

(4) 对《红楼梦学刊》四十年期刊论文进行主题挖掘, 主要包括: 时序主题演化分析、共词网络分析和社交网络分析。

## 2.1 数据采集

本文旨在深度挖掘中国知网收录《红楼梦学刊》期刊论文的核心作者以及核心主题。依托 Python 环境下的 Selenium 模块抓取中国知网收录 1979 年 4 月 2 日至 2019 年 4 月 2 日期间《红楼梦学刊》的期刊论文。所抓取的字段主要包括: 论文标题、论文作者、论文发布时间、该论文被引用量以及下载量、论文关键词以及论文摘要。

## 2.2 数据清洗

中国知网收录的《红楼梦学刊》期刊论文中部分数据信息不完整, 需对抓取数据进行预处理, 规范数据格式。本文采用的数据预处理主要包括: 数据清洗、异常值判定处理以及相关数值提取。

# 3 核心作者群体发现

大数据在提供海量多元化数据信息的同时, 也带来了信息过载的烦恼, 在学术研究领域尤为突出。在线投稿方式取代传统投稿方式的同时, 也带来学术成果快速增长的现象, 精准识别文献核心作者越发困难。核心作者是学科研究的坚实基础<sup>[17]</sup>, 决定着学科研究方向以及学术成果的质量。传统思维识别方法是通过发文量来认定, 而忽略论文质量。对此, 本文采用一种基于普赖斯定律和综合指数的《红楼梦学刊》核心作者确认方法。先依据普赖斯定律确认《红楼梦学刊》期刊论文的核心作者候选人, 再结合作者发文量以及被引用量的综合指数遴选《红楼梦学刊》期刊论文的核心作者。

## 3.1 普莱斯定律分析

本文结合《红楼梦学刊》期刊论文的第一作者的发文量和被引用量来筛选核心作者候选人, 再结合普赖斯定律预选《红楼梦学刊》核心作者候选人, 具体操作如下。

### (1) 确认最低被引用量

《红楼梦学刊》被引用次数最高的期刊论文是王金波于 2010 年发表的“被忽视的第一个《红楼梦》120 回英文全译本——邦斯尔神父《红楼梦》英译文简介”, 共被引用 71 次, 记作  $N_{c_{max}}$ 。结合普赖斯最低被引用量计算公式(1), 计算最低被引用量, 记为  $M_c$ 。被发文引用累计达到 7 次或 7 次以上的作者可入选核心作者候选人。

$$M_c = 0.749\sqrt{N_{c_{max}}} = 0.749 \times \sqrt{71} \approx 6.31 \quad (1)$$

### (2) 确认最低发文量

《红楼梦学刊》发文量最多的作者为冯其庸, 共计发文 124 篇, 记作  $N_{p_{max}}$ 。结合普赖斯最低

发文量计算公式 (2), 计算最低发文量, 记作  $M_p$ 。发表 9 篇或 9 篇以上的作者可入选核心作者候选人。

$$M_p = 0.749\sqrt{N_{p_{max}}} = 0.749 \times \sqrt{124} = 8.34 \quad (2)$$

### (3) 核心作者候选人确认

筛选符合公式 (1) 或者公式 (2) 的《红楼梦学刊》期刊论文作者, 并进行去重统计, 最终选出符合标准的 363 位核心作者候选人, 共发文 2 957 篇, 占《红楼梦学刊》总收录论文的 52.97%; 所发论文共被引用 10 324 次, 占《红楼梦学刊》总收录论文被引用量的 66.41%。

## 3.2 综合指数遴选核心作者

结合综合指数从普赖斯定律确认的《红楼梦学刊》363 位核心作者候选人中设置综合指数阈值为 2, 遴选出 35 位《红楼梦学刊》收录论文的核心作者, 具体步骤如下:

### (1) 确认平均发文量

普赖斯定律预选的《红楼梦学刊》363 位核心作者候选人总发文量, 记作  $X_{总}$ ;《红楼梦学刊》预选的核心作者候选人总人数, 记作  $n$ 。结合平均发文量计算公式 (3), 计算核心作者候选人的平均发文量, 记作  $\bar{x}$ 。

$$\bar{x} = \frac{X_{总}}{n} = \frac{2957}{363} \approx 8.15 \quad (3)$$

### (2) 确认平均被引量

普赖斯定律预选的《红楼梦学刊》363 位核心作者候选人所发论文的总被引用量, 记作  $Y_{总}$ ;《红楼梦学刊》预选的核心作者候选人的总人数, 记作  $n$ 。结合平均被引用量计算公式 (4), 计算核心作者候选人的平均被引用量, 记作  $\bar{y}$ 。

$$\bar{y} = \frac{Y_{总}}{n} = \frac{10324}{363} \approx 28.44 \quad (4)$$

### (3) 综合指数遴选

《红楼梦学刊》核心作者候选人的平均发文量, 记作  $\bar{x}$ ;《红楼梦学刊》核心作者候选人的平均被引用量, 记作  $\bar{y}$ , 结合综合指数计算公式 (5) 计算第  $i$  位核心作者候选人的综合指数得分, 记作  $score_i$ 。计算过程中, 第  $i$  位核心作者候选人的总发文量记作  $x_i$ ; 第  $i$  位核心作者候选人的总被引用量记作  $y_i$ 。设置综合指数阈值为 2, 遴选出 35 位《红楼梦学刊》收录论文的核心作者。

$$score_i = \frac{x_i}{\bar{x}} \times 0.5 + \frac{y_i}{\bar{y}} \times 0.5 \quad (5)$$

表1 《红楼梦学刊》核心作者

第一作者	发文次数	篇均被引	综合指数	单篇最高被引文献	单篇最高被引次数
冯其庸	124	1.89	11.72	解读《红楼梦》	14
洪涛	25	12.92	7.21	《红楼梦》英译与东西方文化、语言	58
胡文彬	50	4.18	6.74	《红楼梦》与中国姓名文化	60
吕启祥	56	3.32	6.71	花的精魂 诗的化身 ——林黛玉形象的文化蕴含和造型特色	46
张庆善	53	2.68	5.75	桐花凤阁主人陈其泰《红楼梦》评点浅谈	24
沈治钧	47	3.45	5.73	《金瓶梅》和《红楼梦》之间的三段“链环”	22
朱淡文	28	5.71	4.53	林黛玉形象探源	33
蔡义江	43	2.14	4.26	“石头”的职能与甄、贾宝玉 ——《红楼梦论佚》中有关结构艺术的一章	17
梅新林	21	7.00	3.87	“旋转舞台”的神奇效应 ——《红楼梦》的宴会描写及其文化蕴义	34
刘世德	37	2.19	3.69	四种《红楼梦》程乙本的差异	11
赵建忠	31	2.81	3.43	二十世纪红学流派的冲突对垒与磨合重构	10
陈维昭	17	7.59	3.31	余英时红学观点的意义及其负面影响	23
曹立波	24	4.33	3.30	阮籍对《红楼梦》的影响举隅	20
王人恩	29	2.97	3.29	森槐南的题《红楼梦》诗词 ——《森槐南与红楼梦》补论（之二）	13
饶道庆	16	7.88	3.20	《红楼梦》影视改编与传播研究述评	29
段江丽	16	7.63	3.13	从小说叙事到影视叙事的改编空间 ——关于新版《红楼梦》电视剧的思考	30
杜景华	26	3.08	3.00	《红楼梦》与禅宗	11
王金波	12	10.42	2.93	被忽视的第一个《红楼梦》120回英文全译本	71
刘相雨	19	5.11	2.87	被污辱与被损害的女性 ——论赵姨娘与中国古典长篇小说中的妾妇形象	18
张书才	25	2.88	2.80	曹雪芹生父新考	15
俞晓红	14	7.36	2.67	《红楼梦》花园意象解读	27
孙玉明	18	4.44	2.51	大观园漫谈	14
张云	20	3.60	2.49	薛宝琴论	9
张锦池	18	4.28	2.46	论林黛玉性格及其爱情悲剧	18
杜春耕	14	5.71	2.27	程甲、程乙及其异本考证	21
詹丹	17	3.71	2.15	从说开去到说进去 ——谈中学语文教材中的“香菱学诗”	14
胡绍棠	17	3.59	2.12	曹寅与《全唐诗》	10
孙逊	16	4.00	2.11	“红楼文化”论纲	11
杜志军	13	5.69	2.10	《红楼梦》与《花月痕》	14
周思源	17	3.53	2.10	重拍电视连续剧《红楼梦》刍议	9
刘梦溪	16	3.94	2.09	读《红楼梦人物论》	10
刘永良	9	9.33	2.03	《红楼梦》回目语言探美	37
李晶	12	6.08	2.02	杨宪益、戴乃迭的《红楼梦》英译本底本研究初探	32

续表

第一作者	发文次数	篇均被引	综合指数	单篇最高被引文献	单篇最高被引次数
高淮生	12	6.08	2.02	关于《红楼梦》影视改编的思考	15
段启明	21	1.95	2.01	《红楼梦》与中国传统美学观	5

结合综合指数遴选核心作者候选人得出表1《红楼梦学刊》35位核心作者。从表中可以看出遴选出的核心作者中,第一位核心作者为冯其庸,共发文124篇,篇均被引1.89次,综合指数为11.72,单篇最高被引文献为“解读《红楼梦》”,被引用次数为14次;第二位核心作者为洪涛,共发文25篇,篇均被引12.92次,综合指数为7.21,单篇最高被引文献为“《红楼梦》英译与东西方文化、语言”,被引用次数为58次;第三位核心作者为胡文彬,共发文50篇,篇均被引4.18次,综合指数为6.74,单篇最高被引文献为“《红楼梦》与中国姓名文化”,被引用次数为60次。

结合篇均被引情况分析,《红楼梦学刊》核心作者中有八位篇均被引次数达到7次及以上。排第一位的是洪涛,篇均被引次数为12.92次;第二位是王金波,篇均被引次数为10.42次;第三位是刘永良,篇均被引次数为9.33次;第四位是饶道庆,篇均被引次数为7.88次;第五位是段江丽,篇均被引次数为7.63次;第六位是陈维昭,篇均被引次数为7.59次;接着是俞晓红,篇均被引次数为7.36次;最后是梅新林,篇均被引次数为7.00次。

## 4 文献主题挖掘

关键词作为期刊论文的核心内容所在,可大致体现研究主题、研究方法以及热点话题等内容。对《红楼梦学刊》收录期刊论文的关键词进行主题挖掘,能够明确《红楼梦学刊》领域的主要研究方向、研究方法以及热点主题。本文的文献主题挖掘主要包括时序主题演化分析、共词网络分析和社交网络分析。

### 4.1 时序主题演化分析

CiteSpace 时间序列主题演化是主要以时间发展为主题轴进行的主题演化分析。本文基于《红楼梦学刊》创刊四十年收录的5582篇期刊论文生成图2所示的时间序列主题演化分析图。图中每一个节点代表一个主题,两点之间的连线表示各主题词之间的共现关系。时间轴跨度为1979年至2019年。

通过考察词频出现的时间分布,将各个主题词和关键词相结合,将变化率高的词从中挖掘出来,依托词频变化趋势,确认《红楼梦学刊》前沿主题以及各时期的发展趋势。其核心主题主要包括“刘心武”“黛玉”“怡红公子”“版本”“作者”“墓石”“社会科学家”等。

总而言之,四十年我国红学研究者对《红楼梦》的相关研究经历了由点到线、由线到面的发展过程,不仅深度挖掘期刊文献,同时也进行考古研究以及影视作品的拍摄。从而更加客观、准确地还原《红楼梦》原著的主题、把握作品的时代背景以及作者的创作倾向,为红学研究的进一步发展提供了理论依据。

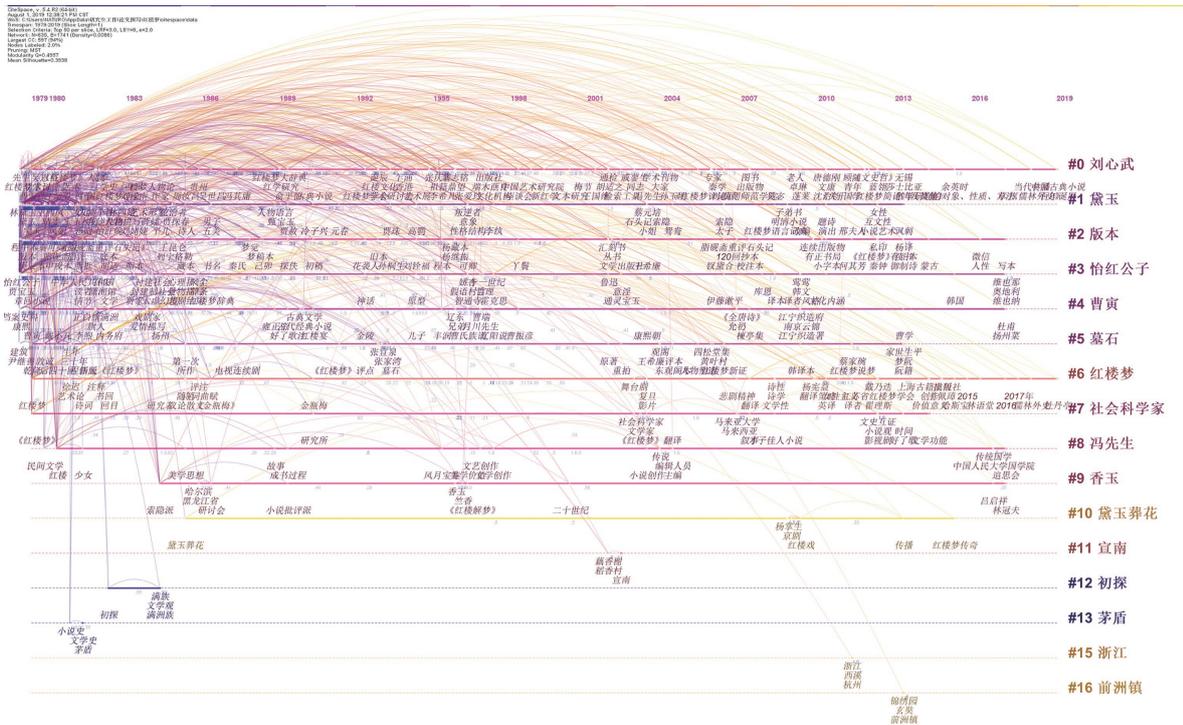


图2 CiteSpace 基于时间序列的《红楼梦学刊》主题演化分析

### 4.2 共词网络分析

本文依托 Python 环境对《红楼梦学刊》自创刊以来收录的 5 582 篇论文构建关键词共现矩阵。当两个关键词共同出现在同一篇期刊文献中时，视为共现并构建一条关系边，其边对应的权重 +1；反之，两关键词之间不存在共现关系，无法构建关系边，该条关系权重为 0。其具体共现矩阵构建公式如计算公式（6）。

$$score_i = \begin{cases} +1, a, b \text{ 关键词共同出现在同篇文献中} \\ 0, a, b \text{ 关键词未共同出现在同篇文献中} \end{cases} \quad (6)$$

在共现矩阵分析中，两个关键词共同出现的次数记为词频，词频越高说明关键词越紧密，越能体现主题的研究内容相关联；如果共现次数为 0，说明两个关键词之间没有关系。

为通过共现矩阵更好体现出红学研究的核心内容，在此剔除以下关键词：“红楼梦”（本文研究主题）、“曹雪芹”（《红楼梦》作者）、“红楼梦学刊”（本文研究期刊）、“章回小说”（《红楼梦》形式）。根据《红楼梦学刊》论文关键词共现分析，得出表 2 所示的文献共现高频词表。

排名前 20 位的分别是：“宝玉”和“黛玉”共现 582 次，“贾宝玉”和“怡红”共现 369 次，“黛玉”和“宝钗”共现 184 次，“宝玉”和“宝钗”共现 178 次，“宝玉”和“凤姐”共现 121 次，“宝玉”和“贾母”共现 102 次，“黛玉”和“凤姐”共现 88 次，“宝玉”和“贾府”共现 83 次，“宝玉”和“晴雯”共现 83 次，“黛玉”和“贾母”共现 77 次，“甲戌本”和“庚辰本”共现 76

次,“大观园”和“宝玉”共现 70 次,“贾母”和“凤姐”共现 66 次,“抄本”和“版本”共现 65 次,“贾府”和“黛玉”共现 65 次,“宝玉”和“贾政”共现 62 次,“宝玉”和“庚辰本”共现 55 次,“先生”与“红学”共现 53 次,“宝玉”和“王夫人”共现 52 次,“贾母”和“王夫人”共现 51 次。

表 2 《红楼梦学刊》文献共现高频词表

序号	关键词 1	关键词 2	共现频次	序号	关键词 1	关键词 2	共现频次
1	宝玉	黛玉	582	11	甲戌本	庚辰本	76
2	宝玉	怡红	369	12	大观园	宝玉	70
3	黛玉	宝钗	184	13	贾母	凤姐	66
4	宝玉	宝钗	178	14	抄本	版本	65
5	宝玉	凤姐	121	15	贾府	黛玉	65
6	宝玉	贾母	102	16	宝玉	贾政	62
7	黛玉	凤姐	88	17	宝玉	庚辰本	55
8	宝玉	贾府	83	18	先生	红学	53
9	宝玉	晴雯	83	19	宝玉	王夫人	52
10	黛玉	贾母	77	20	贾母	王夫人	51

从共现高频词可以看出,《红楼梦学刊》整体研究方向主要可分为两类,分别是人物关系研究和《红楼梦》版本研究。人物关系的主体人物以“宝玉”和“黛玉”为主,其他人物关系与之相关联。版本研究主要以对《甲戌本》和《庚辰本》的探究为主。

### 4.3 社交网络分析

社交网络算法是一种可以识别强弱关系网络的近聚类算法,运用知识图谱的方式将彼此之间的关系进行形象化表示。节点代表关系点,边代表节点之间关系。社交网络算法可将关系较为密切的节点聚类到类似区域,将关系稀疏的节点扩散到外围,直观发现核心关系点。

由于存在大量权重系数为 1、2、3 的零散节点,影响社交网络的整体效果,在此结合普赖斯定律进行节点删选,如计算公式(7)所示。

$$M_f = 0.749\sqrt{N_{f_{max}}} = 0.749 \times \sqrt{582} \approx 18.06 \quad (7)$$

公式(7)中  $M_f$  表示共现高频词最低频次,  $N_{f_{max}}$  表示普赖斯定律统计的《红楼梦学刊》收录期刊论文的共现高频词的最高频次。结合普赖斯定律可知共现高频词的频次须大于等于 19。因此设置共现阈值为 19,进行社交网络分析相关共现高频词,生成如图 3 所示的主题关键词共现知识图谱。该图谱包含核心节点数 54 个、核心节点产生 108 条关系边。经计算该社交网络的模块化系数为 0.549,符合模块化效果。

从整个关键词关系图谱可看出,《红楼梦学刊》收录期刊论文的研究整体可大致分为四个模

块,分别是人物关系研究、社会制度研究、当下学术研讨和《红楼梦》版本猜测。整体而言,四个模块领域中,人物关系研究以“宝玉”和“黛玉”两位人物研究为主,其他人物关系与之相关联,刘姥姥关系相对单一。社会制度研究以“封建社会”为主,映射成《红楼梦》创作所处社会背景。当下学术研讨以“冯其庸”先生作为红学研究代表,其在《红楼梦学刊》上共发文124篇,代表作为“解读《红楼梦》”。《红楼梦》版本猜测以晚清状元、协办大学士徐鼐(号颂阁)旧藏的《脂砚斋重评石头记》(又名《石头记》)为主,也存在部分学者认为的《甲戌本》、《庚辰本》、《程甲本》和《程乙本》。

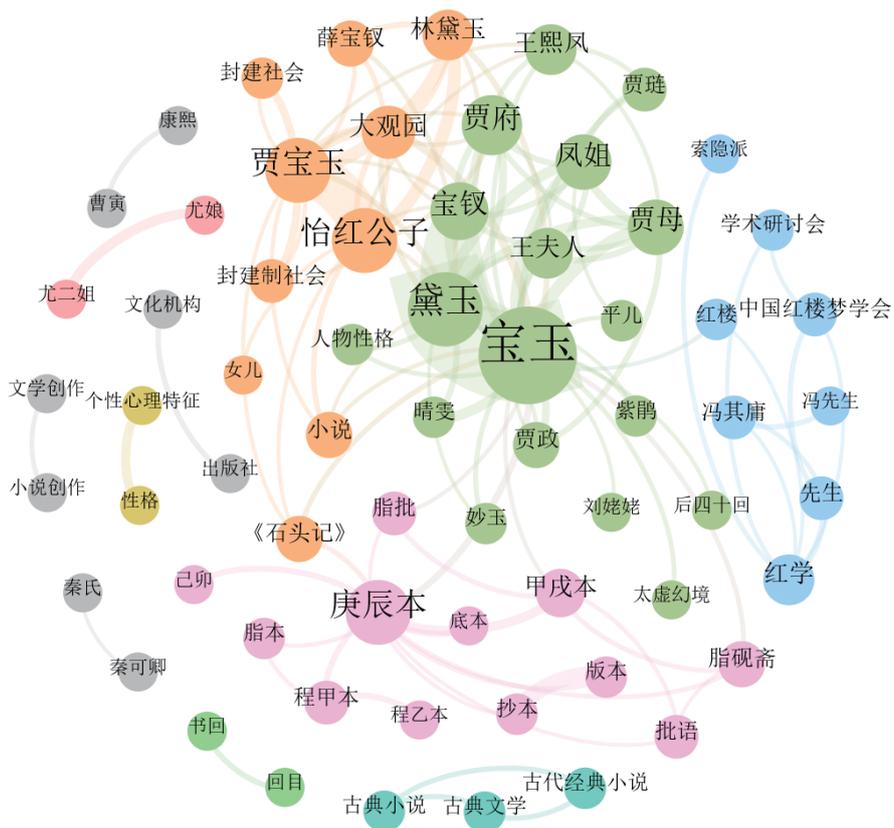


图3 《红楼梦学刊》主题关键词共现知识图谱

## 5 结语

随着艺术文化的发展,近年来红学研究和文艺评价层出不穷。红学研究者从不同角度对其进行分析研究,学术成果呈现“百花齐放,百家争鸣”之势。《红楼梦学刊》作为红学研究主要学术期刊,四十年刊发大量期刊论文,是红学研究的主要交流载体。然而,红学研究者对《红楼梦》进行深度研究时,往往建立在对内容本身及其个人理解上,研究结果相对单一,需要充分考

虑《红楼梦》的精髓所在以及核心学者对原著的认知层面。

本文主要运用了计量统计和主题挖掘的方法进行分析。先计量统计出四十年红学研究的核心作者, 再对红学领域成果进行主题挖掘。从核心作者发现和主题演化分析两方面研究四十年红学的研究热点。研究表明, 四十年红学研究以人物关系研究、社会制度研究、当下学术研讨和《红楼梦》版本猜测为主。该研究成果主体上确定了《红楼梦学刊》的核心研究主题以及发展历程, 可以给《红楼梦学刊》编辑部提供审稿标准以及相关主题建议, 为红学研究者提供更好的学术研究方向建议。

### 【参考文献】

- [1] 温庆新. 作为一种文学现象: 现代“《红楼梦》化”的生产性批评接受 [J]. 中国文学研究, 2021 (2): 155-162.
- [2] 赵建忠. “新红学”百年之际对《红楼梦》“作者”及“曹学”的反思 [J]. 明清小说研究, 2021 (1): 4-24.
- [3] 王慧. 奠基守正 砥砺前行——“红楼梦研究所建所、《红楼梦学刊》创刊四十周年学术研讨会”综述 [J]. 红楼梦学刊, 2020 (1): 8-16.
- [4] 杨秀璋, 武帅, 夏换, 等. 大数据视域下2019年中国电影产业分析研究 [J]. 电影文学, 2020 (23): 17-29.
- [5] 武帅. 科学研究前沿主题识别 [D]. 贵阳: 贵州财经大学, 2021.
- [6] 裴杰. 梦渡扶桑 [D]. 上海: 上海外国语大学, 2020.
- [7] 张庆善, 乔福锦, 苗怀明, 等. 《红楼梦》文献学研究笔谈 [J]. 中国矿业大学学报 (社会科学版), 2016, 18 (5): 89-96.
- [8] 高淮生. “历史回顾与未来展望——《红楼梦》文献学研究高端论坛”学术综述 [J]. 河南教育学院学报 (哲学社会科学版), 2016, 35 (3): 3-11.
- [9] 孙伟科, 何卫国, 胡晴, 等. 2017年度中国红学发展研究报告 [C] // 2017年度中国艺术发展研究报告, 2018: 363-385.
- [10] 沈琳. 程甲本《红楼梦》家具品类研究 [D]. 长沙: 中南林业科技大学, 2021.
- [11] 武迪, 吴佳儒. 《香艳丛书》所收《红楼梦》资料考述 [J]. 红楼梦学刊, 2017 (6): 175-189.
- [12] 陈骁. 清代《红楼梦》的图像世界 [D]. 杭州: 中国美术学院, 2012.
- [13] 蔡永明, 长青. 共词网络LDA模型的中文短文本主题分析 [J]. 情报学报, 2018, 37 (3): 305-317.
- [14] 武帅, 任天舒, 刘建义, 等. 基于数据计量和社交网络的图书情报学科发展探究 [J]. 情报探索, 2022 (1): 28-40.
- [15] 杨秀璋, 武帅, 夏换, 等. 基于文献计量和社交网络的清水江流域文化探究 [J]. 现代计算机, 2019 (35): 19-26, 37.
- [16] 杨秀璋. 基于水族文献的计量分析与知识图谱研究 [J]. 现代计算机 (专业版), 2019 (1): 25-32.
- [17] 杨秀璋, 夏换, 于小民, 等. 基于综合指数和知识图谱的水族文献核心作者群分析研究 [J]. 计算机时代, 2019 (4): 13-17.

# Research on Core Author Discovery and Topic Mining Driven by Digital Humanities

Wu Shuai<sup>1,2</sup> Yang Xiuzhang<sup>2</sup> Ren Tianshu<sup>2</sup> Liu Jianyi<sup>2</sup>

(1. College of Information Management, Nanjing Agricultural University, Nanjing 210003, China;  
2. School of Information, Guizhou University of Finance and Economics, Guiyang 550025, China)

---

**Abstract:** [ **Purpose/significance** ] Driven by digital humanities, the achievement of the 40th anniversary of *Studies on "A Dream of Red Mansions"* were statistically analyzed from the perspective of researchers from the macro level, and the potential research fields of Redology were analyzed from the micro level by integrating data mining technology, so as to better promote the Redology research. [ **Method/process** ] The methods of measurement statistics and topic mining were used in turn for analysis. First, the core authors of Redology research over the past four decades were counted. Then, the topics mining of the achievements in the field of Redology were carried out. The research hotspots of Redology research over the past four decades from the two aspects of core author discovery and topic evolution. [ **Result/conclusion** ] The research on Redology can be divided into four areas as a whole, which are the study of character relationships, the study of social systems, the current academic discussion and the guessing of the version of "*A Dream of Red Mansions*".

**Keywords:** *Studies on "A Dream of Red Mansions"*; Topic evolution; Digital humanities; Topic mining; Core author

---

( 本文责编: 魏 进 )