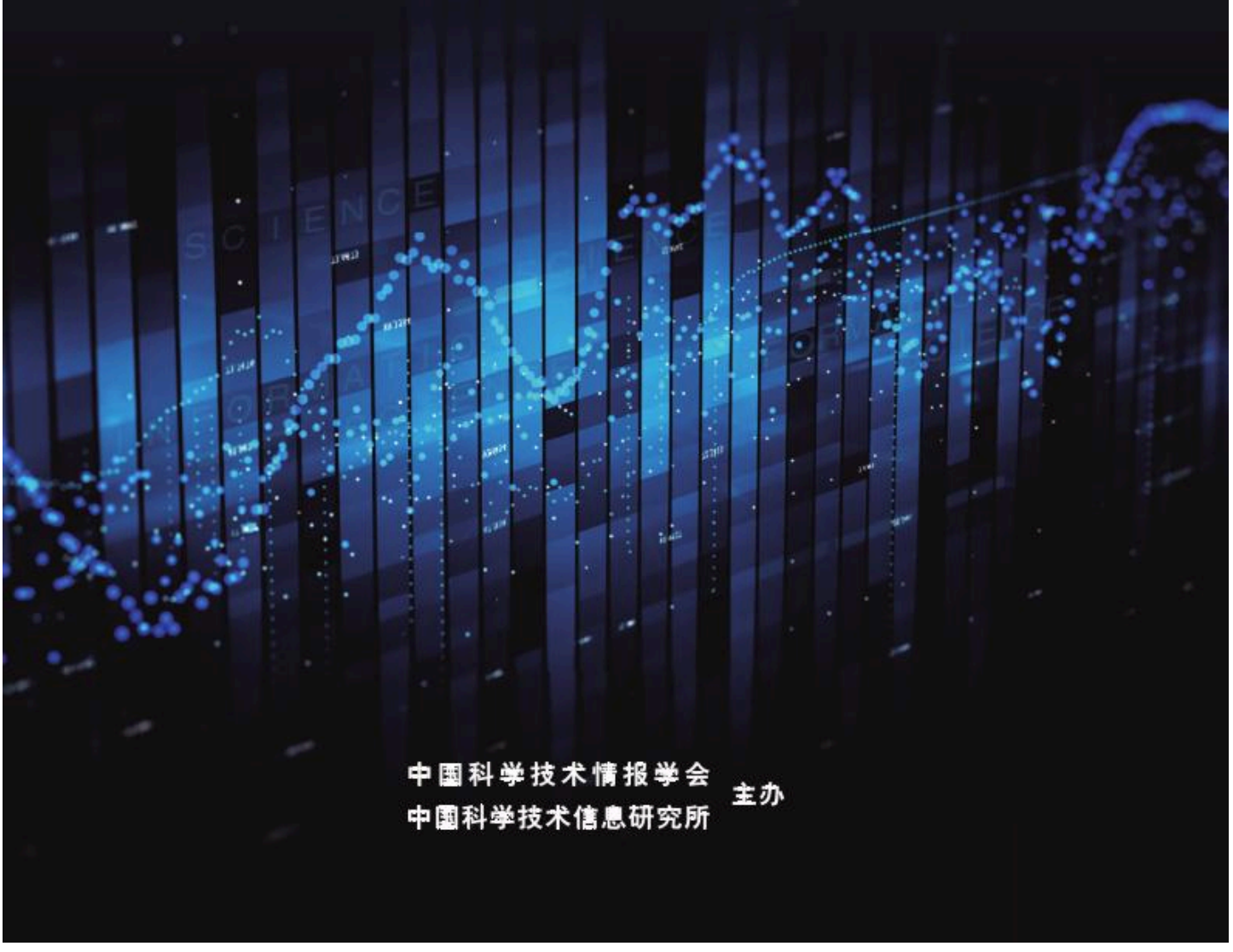
JOURNAL OF
THE CHINA SOCIETY FOR
SCIENTIFIC AND TECHNICAL INFORMATION

ISSN 1000-0135
CN 11-2257/G3

情报学报

第 43 卷 第 7 期 Volume 43 Number 7 2024

- 合贡献者网络的结构特征及其合作群体识别应用研究
- 科学-技术关联对高技术产业创新绩效的影响研究——对外技术依存度的调节作用



中国科学技术情报学会
中国科学技术信息研究所 主办

情 报 学 报

Qingbao Xuebao

（月刊 1982年创刊）

第43卷 第7期

2024年7月24日出版

JOURNAL OF THE CHINA SOCIETY FOR SCIENTIFIC AND TECHNICAL INFORMATION

(Monthly, Founded in 1982)

Vol. 43 No. 7 July 2024

主管单位 中国科学技术协会

主办单位 中国科学技术情报学会
中国科学技术信息研究所

编辑出版 《情报学报》编辑部

地址 北京市复兴路15号（100038）

电话 （010）68598273，58882172，68598285

网址 https://qbxb.istic.ac.cn/

E-mail qbxb@istic.ac.cn

印刷单位 北京科信印刷有限公司

总发行处 北京报刊发行局

订 购 处 全国各地邮局

邮发代号 82-153

国外发行 中国国际图书贸易总公司（北京399信箱）

国外代号 BM4134

定 价 48.00元

Supervised by

China Association for Science and Technology

Sponsored by

China Society for Scientific and Technical Information
Institute of Scientific and Technical Information of China

Edited & Published by

Editorial Board of Journal of the China Society for Scientific and Technical Information
No. 15, Fuxing Road, Beijing 100038, China

Distributed by (except in China)

China International Book Trading Corporation (Guoji Shudian), BM4134, P. O. Box 399, Beijing, China

ISSN 1000-0135

CN 11-2257/G3

《情报学报》是国家自然科学基金委员会管理科学部认定的A类学术期刊，被如下数据库系统收录：

科学文摘（INSPEC，英国电气工程师学会），图书馆和信息科学文摘（LISA，美国），文摘杂志（РЖ，俄罗斯），中文社会科学引文索引（CSSCI，南京大学），中国科技核心期刊（CSTPCD），中文核心期刊要目总览，中国期刊全文数据库（CNKI，中国知网），中国学术期刊文摘（CSAC，中国科学技术协会），数字化期刊全文数据库（万方数据），国家哲学社会科学学术期刊数据库（中国社会科学院）。

情报学报

Qingbao Xuebao

第 43 卷　第 7 期　2024 年 7 月

目　　次

情报理论与方法

情报技术与应用

Journal of the China Society for Scientific and Technical Information

Vol. 43 No. 7 July 2024

Contents

Intelligence Theories and Methods

Intelligence Technology and Application

情报学报　2024 年 7 月　第 43 卷　第 7 期
Journal of the China Society for Scientific and Technical Information, Jul. 2024, 43(7): 773-788

DOI: 10.3772/j.issn.1000-0135.2024.07.002

合贡献者网络的结构特征及其合作群体识别应用研究

卢　超[1]，李梦婷[1]，陈秀娟[2]，董　克[3,4]，魏瑞斌[5]

（1. 河海大学商学院，南京　211100；2. 南京师范大学新闻与传播学院，南京　210097；
3. 南京大学数据管理创新研究中心，苏州　215163；4. 南京大学数据智能与交叉创新实验室，南京　210023；
5. 安徽财经大学管理科学与工程学院，蚌埠　233030）

摘　要　有组织科研团队建设有赖于对科研合作现象和规律的科学认识。常用于科研合作模式研究的合著者网络默认同一成果的合作者间贡献均等，但这通常与科研合作实践相左。作者贡献声明数据的出现为揭示更细粒度的合作实践提供了重要素材。为此，本研究提出一种利用贡献声明数据构建的新型合作网络——合贡献者网络，为深入研究科研合作问题提供新工具。本研究以 PLoS（Public Library of Science）上的药学论文数据为例，以合著者网络为基准，从合贡献者网络的网络结构特征入手，认识此新型合作网络的物理性质；选取当前重要研究方向之一的“合作群体识别”为切入点，进一步认识合贡献者网络的应用价值。研究结果表明：①在网络结构形态上，合贡献者网络比合著者网络更稀疏；②在合作群体识别上，两种网络的群体识别结果部分一致，重合度约为 57%；约 32% 的合作群体在合贡献者网络上发生了重组；③合贡献者网络中的合作群体发文主题比合著者网络更为聚焦，但检验结果并不显著。总体来看，在本研究的数据集上，合贡献者网络较之合著者网络显示出更良好的社区结构；合贡献者网络有助于识别出更细粒度的合作群体，且在所识别的合作群体上发文主题的一致性更高。

关键词　科研合作；合著者网络；作者贡献声明；合贡献者网络

Structural Characteristics of the Co-contributorship Network and Its Application in Collaborative Group Identification

Lu Chao[1], Li Mengting[1], Chen Xiujuan[2], Dong Ke[3,4] and Wei Ruibin[5]

(1. *Business School, Hohai University, Nanjing*　211100; 2. *School of Journalism and Communication, Nanjing Normal University, Nanjing*　210097; 3. *Research Institute for Data Management & Innovation, Nanjing University, Suzhou*　215163; 4. *Laboratory of Data Intelligence and Interdisciplinary lnnovation, Nanjing University, Nanjing*　210023; 5. *School of Management Science and Engineering, Anhui University of Finance and Economics, Bengbu*　233030)

Abstract: The construction of organized scientific research teams relies on a scientific understanding of the phenomena

收稿日期：2023-03-17；**修回日期**：2023-10-19

基金项目：国家自然科学基金青年科学基金项目“劳动分工视角下科研合作者的科研效能研究”（72004054）；江苏省高等学校基础科学（自然科学）研究面上项目“重大突发公共卫生事件中疫苗研发领域的国际科研合作布局与产出效益研究”（21KJB630008）；中央高校基本科研业务经费专项资金项目“科研团队多样性对科研绩效的因果效应研究”（B220201058）；江苏省社会科学基金后期资助项目“我国NSFC国际合作研究项目国际合作特征与产出规律研究”（21HQ038）。

作者简介：卢超，男，1991年生，博士，副教授，硕士生导师，研究方向为科学学、科研合作；李梦婷，女，2000年生，硕士研究生，研究方向为学术文本挖掘；陈秀娟，女，1989年生，博士，副教授，硕士生导师，研究方向为科研合作、开放科学；董克，男，1986年生，博士，副教授，博士生导师，研究方向为文献计量；魏瑞斌，通信作者，男，1973年生，博士，教授，硕士生导师，研究方向为科学计量、信息组织，E-mail：rbwxy@126.com。

and patterns of research collaboration. The commonly used research model for scientific collaboration, the co-authorship network, assumes equal contributions among co-authors for the same research output, which often contradicts actual research collaboration practices. The emergence of author contribution statement data provides valuable material for revealing more detailed collaboration practices. This study proposes a novel collaboration network, which is called the "co-contributorship network" and constructed using contribution declaration data, to provide a new tool for investigating scientific collaboration issues in depth. Using article data in the field of medicine from *PLoS* as an example and the co-authorship network as a baseline, we explore the physical properties of this new collaboration network through its network structure characteristics. Furthermore, we focus on identifying collaboration groups, an essential research direction, to better understand the practical value of the co-contributorship network. The study finds the following: in terms of the network structure, the co-contributorship network is sparser than the co-authorship network. The results of both networks regarding the identification of collaboration groups partially coincide, with an overlap of approximately 57%. Approximately 32% of the collaboration groups experienced restructuring in the co-contributorship network. The evaluation results show that collaboration groups in the co-contributorship network tend to be more focused on research topics compared to those in the co-authorship network, but the difference is not statistically significant. Overall, based on our dataset, the co-contributorship network exhibits a more favorable community structure compared to the co-authorship network. It helps identify finer-grained collaboration groups with higher consistency in their research topics among the identified groups.

Keywords: scientific collaboration; co-authorship network; author contribution statement; co-contributorship network

0 引 言

科学研究经历着从关注单一学科问题到多学科交叉融合问题的发展历程。随着研究问题的不断深入和复杂，学者难以独立完成整项研究，特别是极具创新性的研究。科学研究逐渐从以往的独立研究形式过渡到如今更为普遍的团队合作形式[1-2]。从劳动分工的角度来看，科研合作理论上有助于提升科学研究的效率，推动科学研究的前进和发展。因此，精准认识科研合作中的分工模式以及成员间的互动关系，对我国有效构建科研合作团队、推动有组织科研建设、促进我国科研创新水平具有重要意义。

一般而言，科研合作模式研究途径主要有两种。一种是定性方法，通过对科研群体或团队进行长期跟踪，认识团队合作的模式及其劳动分工，并据此优化团队合作[3-4]。这一类研究能很好地揭示科研合作的形式、具体内容与团队结构等；但消耗的时间和经济成本通常较高，样本数量有限，研究结论缺乏普适性。另一种则是文献计量方法，以大规模科研成果中的署名数据揭示科研合作模式[1,5]。这一类研究虽然很好地使用了海量科研成果数据驱动科研合作认识，但也存在一定问题。从文献计量的角度来看，传统文献计量方法分析科研合作是将学者之间在共同完成科研成果过程中的贡献等而视之，忽略了学者在科学活动中的劳动分工，这与实际的科学研究明显不符；特别是通过大规模数据进行科研合作研究时，粗粒度的合作次数累计方法在表征科研合作过程中存在明显短板，即知其然而不知其所以然，限制了文献计量的有效性。从科技政策和科技管理的角度来看，在当前大科学时代背景下，准确认识科研合作模式、揭示科研合作团队的发展规律，在微观上是实现对科研群体精准化管理的基础，在宏观上更是制定有效科研组织政策的基本前提，这就要求对科研合作关系的揭示结果更接近现实。因此，如何克服现有分析方法的不足，既能实现对科研合作关系的有效表征，又能开展基于海量科研成果数据的定量研究，是当前科研合作研究亟待解决的问题。

作者贡献声明是论文合作者为披露其在论文研究工作中所承担的具体任务（或贡献类型）所做的说明，该机制能够有效应对过于强化论文合著关系、忽视合作者具体贡献而引发的诸多问题[6-7]，并得到了国内外大量期刊的采用。通过对作者贡献声明的解析，能够精准识别学者在合作过程中的分工模式以及贡献细节，对学者的学术贡献评估具有重要作用[8-10]。但现有研究较少关注作者贡献声明数据在合作网络模型构建方面的应用，揭示科研合作模式[8,11]。Corrêa 等[11]通过构建作者-贡献二部图网络，实施了作者实际贡献的细粒度测度。Lu 等[8]利用贡献声明数据构建了单篇论文的合贡献者网络子网，研究了论文团队内部的劳动分工模式，但未将其拓展到整体网络中，用于认识合作群体层面的科研合作模式研究。以作者贡献声明为依据，区分作者在科研成果完成过程中的实际贡献，并构建合贡献者网络，为实现对科研合作更为精准的抽象与建

模，并进一步识别和分析科研合作团队具有重要理论意义和实践价值。

为此，本研究提出利用作者贡献声明数据构建一种新型的科研合作网络——合贡献者网络。从理论上来看，在合贡献者网络中，节点用于表征科研成果的合作者，合作者间共同承担具体研究任务关系被抽象为节点之间的边。相较于合著者网络简单地使用合作次数作为边权重，合贡献者网络中边的权重表征了研究人员间共同完成相关贡献类型的总次数，其网络密度理论上应当小于经典的合著者网络密度。因此，本研究以药学为例，以合著者网络为基准，从合贡献者网络的结构特征入手，认识此新型合作网络的物理性质；选取当前重要研究方向之一的“合作群体识别”为切入点，进一步认识合贡献者网络的应用价值。

1　研究进展与研究问题

本节从合著者网络相关研究、合作群体识别相关研究以及作者贡献声明相关研究三个方面综述当前相关研究的最新进展，对当前研究现状进行总结，并提出本研究关注的问题。

1.1　合著者网络相关研究

合著者网络最早由 Newman[12]提出完整的构建思路，他发现合著者网络属于小世界网络。后续研究又发现合著者网络同时具备无标度网络特性[13]，网络具有同质性、传递性和合作偏好[14]。合著者网络也出现了较多在节点属性[15]、边的权重计算[16]与方向[17]方面的改进研究。随着对合著者网络的认识逐渐深入，它也成为研究科研合作问题最常使用的网络模型之一[15,18]。合著者网络被用来探究学者影响力评估[19-20]、合作群体发现[18,21]以及合作机会预测[22-23]等问题。这些研究都表明合著者网络在科研合作问题上能提供良好的工具性作用，但这一类的研究未能充分揭示每个合作者在团队中的具体工作方式，科研合作的微观信息不明确。因此，科研合作中团队成员间复杂的协作关系被忽视了[24]。

1.2　合作群体识别相关研究

在基于合著者网络的众多应用中，合作团队识别是一个重要的应用方向[18,21,25]。合作团队识别是将合作团队从合作数据中识别出来的研究任务。根据不同的合作团队认定层次不同，合作团队的识别方法也不同。合作团队的认定方式可分为三种：论文团队（article-based team）、项目团队（project-based team）以及合作群体（community-based team/collaborative group）[26]。本研究中所指的合作团队是第三种类型，即合作群体，针对这类合作团队的研究更为广泛[18,21]。合作群体通常指利用合作者间合著关系构建的合作网络中存在的社区结构[12,21,27]。在这些社区结构中，学者间的合作频率显著高于与外部学者间的合作[26]。此类合作群体的识别通常借助于复杂网络中的社区发现算法[21,25]、网络密度指标[18]以及其他相关算法[28]进行团队识别。因此，基于合著者网络的社区结构发现是合作群体识别的一个重要基础。

1.3　作者贡献声明相关研究

近年来，作者贡献声明逐渐被众多期刊广泛采纳，用于披露合作者们在其研究成果中的实际贡献。学者们也开始利用作者贡献声明进行科研合作的相关研究，主要包括作者贡献评估[11,29]、科研合作模式研究[2,8]以及作者贡献与署名位置关系研究[2,11,30]三个方向。作者贡献评估通常利用作者贡献数据对作者的具体贡献整体评估，以期取得比署名位置更精准的贡献评价，如 Corrêa 等[11]通过作者贡献的细粒度评估，发现作者贡献大小与署名位置间三种主要的相关关系。科研合作模式研究主要通过披露作者贡献细节探究科研合作中不同合作者间的任务分配和专业化方面的分布规律，如 Lu 等[8]发现科研合作中有三种类型的合作者，分别为全能型、专业型和合作型，并各自承担不同类型的合作任务。通过对作者贡献的细粒度挖掘，相关学者还识别出作者具体贡献及其署名位置存在一定的相关关系[2,11]，但在大团队中作者的具体贡献无法在署名位置中得到清晰反映[2]。

1.4　现状总结与研究问题

综上，首先，科研合作相关研究大量使用合著者网络作为基础工具开展较大数据规模的科研合作问题研究，但此类研究很少关注合作者之间内部的分工状况，对科研合作实践缺乏清晰的表征，难以充分揭示科研合作现象及规律。其次，科研合作群体识别研究通常基于对合著者网络社区结构的发现而实现，社区结构发现算法为相关研究的开展提供了重要支撑。最后，作者贡献声明也不断在科研合作研究中发挥其独特价值。因此，基于上述研究现状，提出以下两个研究问题。

（1）在网络结构形态上，合贡献者网络有何特

点，较合著者网络有何异同？

基于作者贡献声明构建科研合作网络，主要利用作者声明中任务分工上的共现关系[8,11]。相较于合著者网络基于合著关系，合贡献者网络对合作者间合作关系的识别更精细化。因此，有必要了解两种不同粒度的网络如何反映科研合作的结构特征，如节点数、网络密度、连通性等。这为认识合贡献者网络提供了微观层面的物理结构证据，也是进一步应用该网络研究科研合作问题的基础。

（2）在合作群体发现上，合贡献者网络有何特点，较合著者网络有何异同？

通过复杂网络和社会网络分析等手段识别具有共同科学研究兴趣的团队是相关研究的重点内容[21]。近年来，随着科研团队成为科学研究主力军[1,8]，学者间自发的合作行为带来研究主题的集聚乃至学科的发展[31]。因此，本研究通过探究合贡献者网络与合著者网络在社区结构发现上的差异，深入了解合贡献者网络的结构特征，并且通过合作者群体所著文献的主题进一步评估社区划分质量，分析合贡献者网络在合作群体发现中的应用价值。

2 实验数据与方法

本研究的整体思路如图 1 所示。①编写 Python 爬虫，获取 PLoS（Public Library of Science）官网 2016—2020 年所有学术论文 XML（extensible markup language）格式的全文数据；②从所有论文的全文数据中筛选按照 CRediT（contributor roles taxonomy）格式标注作者贡献的学术论文，并从符合要求的全文数据中抽取每篇学术论文的基本信息，包括标题、作者、作者贡献声明等；③利用微软学术知识图谱（Microsoft academic graph，MAG）完成论文作者消歧、学科标签信息获取；④从 PubMed 获取 MeSH（medical subject headings）主题词表用于构建 MeshTerm 网络；选取目标学科标签下的论文数据，分别构建合著者网络和合贡献者网络，并获取这些论文对应的 MeSH；⑤对两种合作网络进行社区划分及相应社区的关键词抽取，并对两种网络结构和识别进行对比分析。在实验证明部分，利用 MeSH 范畴表构建 MeSH 网络，学习每个 MeSH 主题词的词向量。合作群体发现效果的一致性指标（coherence）主要用词向量间的余弦相似度进行计算。下文将从实验数据和实验方法两个方面进行详述。

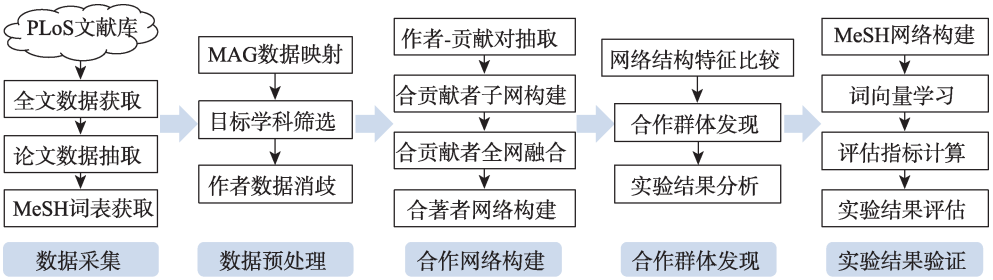


图 1 研究思路①

2.1 实验数据

2.1.1 数据来源

本研究使用的实验数据包括三个来源：PLoS 开源期刊全文数据集，用于构建合著者网络和合贡献者网络；微软学术知识图谱，用于数据集的学科范围筛选和作者姓名消歧；医学主题词表的收集是为了寻找可替代的专家知识，用于从合作群体发文主题的一致性角度评估合作群体识别的效果。

（1）PLoS 开源期刊全文数据集

PLoS 是目前全球最大最早的开放期刊集团之一，每年发文量 2 万~4 万篇，为相关学者提供了大量的学术资源，对生物学、医学等学科具有重要影响。由于其开放数据的理念，学者可以通过其官方提供的 API（application programming interface）或检索入口获取学术论文的全文数据和元数据，为全文文献计量工作的开展提供大量的数据资源[32-33]。此外，PLoS 期刊是最早提供作者贡献披露数据的出版商之一，为相关研究开展作者贡献挖掘工作提供了独特的优势资源[2,8,11]。因此，本研究利用在 PLoS 期刊群中采集的作者元数据和贡献数据用于合作网络的构建。

（2）微软学术知识图谱

微软学术知识图谱（MAG）是目前最具规模和

① 全文彩图见 http://c.nxw.so/ahtEl

影响力的开源学术图谱之一，可通过微软提供的Azure云服务进行免费下载。它由1800—2018年超过2.5亿作者和收录在5万余个期刊及会议中的约2.1亿篇出版物组成，共计19个主要研究领域，包括生物学、计算机科学和物理学等①。本研究利用MAG中包含的PLoS期刊作者的消歧数据构建作者合作网络。

（3）医学主题词表

医学主题词表（MeSH）是生物医学领域广泛使用的标引检索系统，为相关领域提供主题标引服务。MeSH范畴表为MeSH主题词提供了树形的网络结构，可用于定位主题词间的相互关系。本研究采用的PLoS期刊文献均被MeSH所标注，为学术论文的主题揭示提供了专家知识检验依据。本研究使用的词表于2023年1月15日在美国国家医学图书馆官网上获取②，词表文件中包含MeSH主题词30 452个。

2.1.2　数据采集

PLoS全文数据的获取包括两个步骤[33]：①构造检索式“publication_date:[2016-01-01T00:00:00Z TO 2020-12-31T23:59:59Z]”，在PLoS网站检索2016年1月1日③—2020年12月31日的所有文献，共113 336篇，其中研究论文105 875篇，均为XML格式全文，其时间分布和期刊分布信息分别如图2和表1所示。②以研究论文作为分析对象，根据检索结果，共获得1 765个研究论文的分页网址。利用分页源码获得所有PLoS论文绝对链接，爬取XML格式文件，文件记录了正文、作者与被引文献等各类信息。

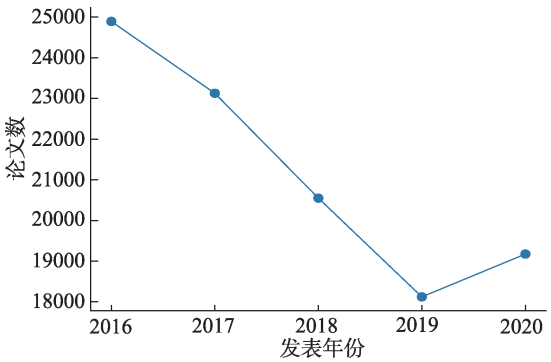


图2　PLoS论文数据的时间分布

表1　PLoS论文数据的期刊分布

期刊名称	论文数	占比(%)
PLoS ONE	91 621	86.5
PLoS Neglected Tropical Diseases	3 843	3.6
PLoS Computational Biology	2 956	2.8
PLoS Genetics	2 755	2.6
PLoS Pathogens	2 655	2.5
PLoS Biology	1 026	1.0
PLoS Medicine	1 010	1.0

2.1.3　数据预处理

（1）论文数据抽取

从2016年开始，PLoS要求作者使用CRediT对所有作者的贡献信息进行标准化标注。为保证研究数据的一致性，将未采用该标注标准的论文过滤掉，共得到68 390篇标准化标注作者贡献的研究论文，并利用规范化的XML语义标注标签，抽取筛选后的所有论文的元数据。①论文信息：包括论文DOI（digital object identifier）、出版年份、期刊名称、文章标题、文章摘要、所属学科；②作者信息：论文DOI、作者姓名、是否为通信作者、作者所属机构；③贡献信息：论文DOI、贡献类型、作者姓名。

（2）学科信息的匹配

PLoS期刊对每一篇学术论文均有较复杂的多学科标注信息[2,34]，因此需要借助MAG知识图谱确定PLoS期刊论文的唯一学科信息。首先，以论文DOI字段为映射，在文献数据表中查找每篇论文的唯一ID（identity）；其次，根据论文的ID信息在论文学科信息表中查找论文所属的学科信息；最后，68 390篇研究论文中的22 476篇文章成功匹配到了一级学科标签，其余45 914篇论文因没有在MAG知识图谱中匹配到唯一论文ID而被排除。通过统计可知，匹配到一级学科标签的22 476篇论文共涉及18个学科，按论文篇数排名前3位的学科分别为药学、生物学和化学（表2），本研究选择占比最大的药学开展后续实验。

（3）作者信息的匹配与消歧

考虑后续构建网络时会出现“同名不同人”或“同人不同名”的问题，利用MAG确定作者的唯一

① https://www.microsoft.com/en-us/research/project/open-academic-graph/

② https://nlmpubs.nlm.nih.gov/projects/mesh/MESH_FILES/xmlmesh/desc2022.xml

③ 本文选择2016年作为PLoS数据采集时间的起始点，是因为PLoS从2016年开始采用贡献角色分类法（CRediT）进行作者贡献声明的标注，作者提供的声明格式更加规范。

表 2 22 476 篇论文的学科分布情况统计表

学科	论文数	占比(%)	学科	论文数	占比(%)
Medicine(药学)	7 890	35.1	Materials Science(材料科学)	260	1.2
Biology(生物学)	6 003	26.7	Physics(物理学)	150	0.7
Chemistry(化学)	2 379	10.6	Political Science(政治学)	101	0.4
Psychology(心理学)	1 844	8.2	Economics(经济学)	99	0.4
Computer Science(计算机科学)	1 588	7.1	Geology(地质学)	76	0.3
Geography(地理学)	856	3.8	Sociology(社会学)	67	0.3
Environmental Science(环境科学)	471	2.1	History(历史学)	29	0.1
Mathematics(数学)	369	1.6	Art(艺术学)	6	<0.1
Business(商学)	292	1.3	Engineering(工程学)	5	<0.1

身份标识符。在得到论文 ID 的基础上，本研究以论文 ID 及作者姓名为映射，在作者信息表中匹配每个作者的 ID。最终，药学的 7 890 篇文章中有 7 570 篇文章的所有作者成功匹配到 AuthorID，而其余的 320 篇论文因作者信息在 MAG 中无法匹配而被排除。

因此，本研究最终使用 7 570 篇包含作者消歧数据的药学学科论文作为最终的数据集。

2.2 实验方法

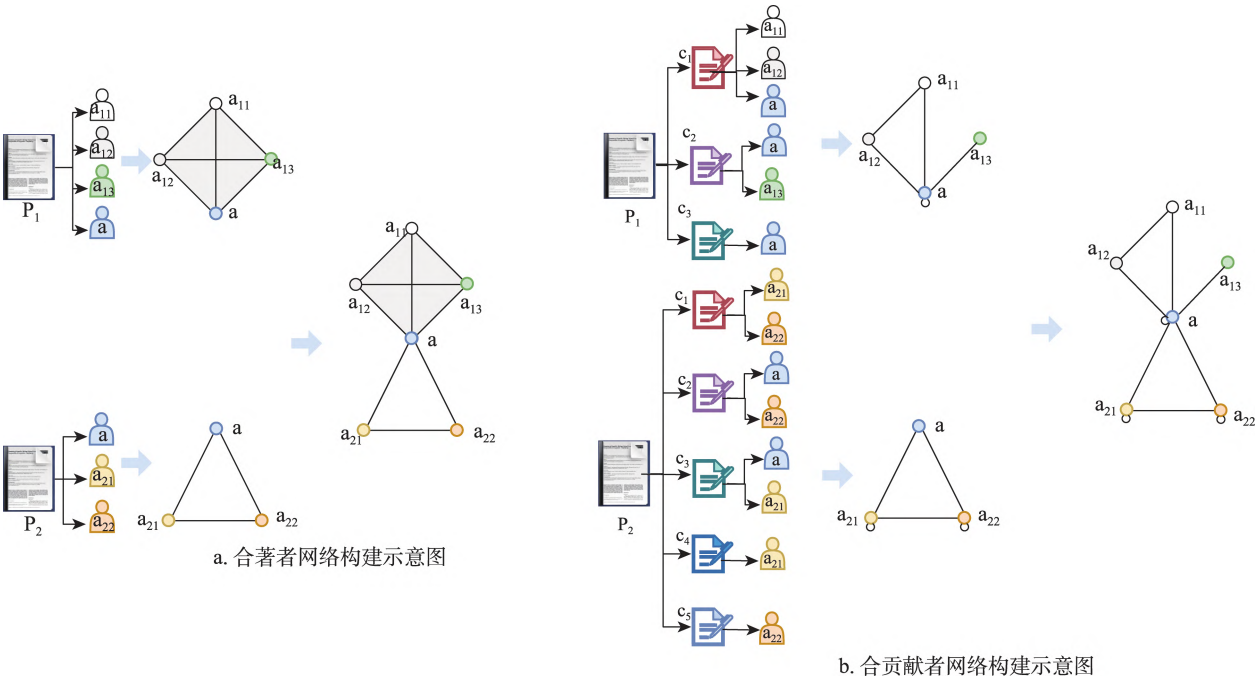
2.2.1 合作网络构建

（1）合著者网络构建

合著者网络构建思路如图 3a 所示。首先，利用单篇文章中合作者间的合作关系构建子网；其次，根据消歧数据将所有合作子网中的节点进行合并，构建合著者网络。本研究涉及的 7 570 篇文章共得到 53 757 个消歧后的作者，合计边数 314 854 条。

（2）合贡献者网络构建

合贡献者网络构建思路如图 3b 所示。首先，利用作者贡献数据在单篇文章中以贡献的合作关系构建合贡献者网络子网。在该子网络中，节点间的边是作者间基于具体贡献（如论文写作）的合作关系。更为具体的子网构建方法参见文献[8]。其次，利用作者消歧数据，对所有子网中的相同作者节点进行合并，构建合贡献者网络。本研究涉及的 7 570 篇论文包含节点 53 757 个、边 272 430 条，比合著者网络中的边数少 13.5%。



a. 合著者网络构建示意图

b. 合贡献者网络构建示意图

图 3 合著者网络和合贡献者网络构建示意图

2.2.2　合作群体发现

（1）合作群体结构发现

在无标签网络数据集上，识别社区结构一直是相关领域中的难题，相关研究通常利用模块度、轮廓系数等网络指标来近似表征网络中社区结构划分的准确程度，模块度是其中最为常用的指标之一[35]。本研究在得到两种合作网络的基础上，提取了各自的最大连通子图，并利用 Gephi 中的模块度计算模块度[36]，利用“手肘准则”获得当前网络较为满意的社区大小[37]。模块度（modularity）是一种目前常用的衡量网络社区结构强度的方法，最早由 Newman[38]提出。通过对模块度的优化，可以实现对网络中社区结构发现结果的优化，保证在识别出的社区中，其内部节点间的边数多于其内部节点与外部节点间的边数。Gephi 采用经改进后的模块度优化算法 Louvain 算法[36]实现社区结构发现，该方法被广泛应用在社区结构发现与合作群体识别相关研究中。通过对 Gephi 软件中参数的人工调节发现，合著者网络的最优社区个数为 76，合贡献者网络的最优社区个数为 85。

（2）合作群体结构变化分型

本研究针对两种合作者网络中合作群体结构划分结果，按照社区结构“一对一”“一对多”和“多对多”三种基本情况，将两种网络间群体变化形式总结为“一致”“细分”和“重组”三种形式，如图 4 所示。“合作群体一致”，即社区结构“一一对应”，是指某一网络中某个合作群体 90% 以上成员构成另一网络中另一个合作群体。“合作群体细分”，即社区结构“一对多”，是指某一网络中的某个合作群体构成另一网络中多个合作群体。“合作群体重组”，即社区结构“多对多”，是指某一网络中的多个合作群体成员在另一网络中被重新分组识别成不同的多个研究群体。

（3）合作群体研究主题识别

得到群体结构划分结果后，利用群体内所有学者合著文章的标题和摘要，进行关键词抽取，并用抽取的关键词作为该学者群体研究主题的标签词，以便捕捉研究主题的内容[39]。这也为研究两种网络合作群体发现结果差异提供了主题证据。本研究采用的关键词抽取算法为 KeyBERT（key bidirectional encoder representation from transformers），该算法利用 BERT 嵌入和余弦相似度来查找文档中与文档本身最相似的子短语[40]。具体做法如下：首先，使用 BERT 提取文档向量（嵌入）以获取文档级表示；

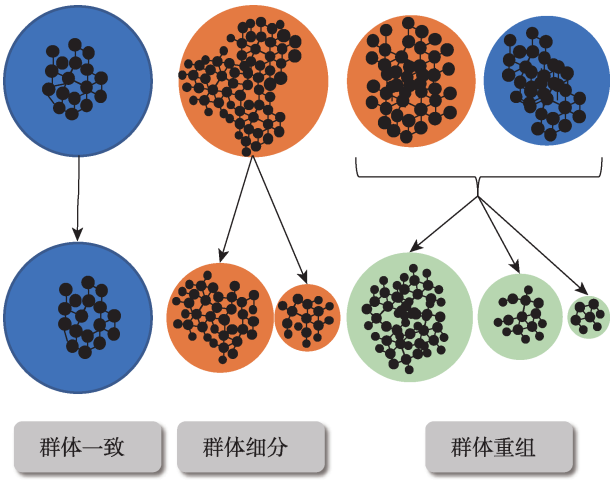


图 4　合作群体变化分型示意图

其次，针对 N 元语法词/短语提取词向量；最后，通过去除停用词，使用余弦相似度查找与文档最相似的 Top N 个词/短语。本研究选取前 10 个关键词作为每篇文章的关键候选词，然后对得到合作群体所有发文的候选词集合，通过提取词干，合并同义词、近义词，最终选取相似度值最高的词语作为合作群体的研究主题。

2.2.3　实验结果验证

基于对实验结果进行的主题分析，本研究借助 MeSH 主题词表，通过计算合作群体发文主题的一致性或相似性，从计算层面分析合贡献者网络相对于合著者网络在合作群体识别上是否更具优势。首先，本研究将基于 MeSH 词表构建树形的语义网络；其次，利用 node2vec 方法计算该网络中每个 MeSH 主题词的词向量；再其次，利用专家对每篇文章标引的主题词，对两种网络上存在结构差异的合作群体（细分类和重组类）的发文主题的主题一致性进行计算；最后，借助主题一致性指标，比较分析两种网络在合作群体识别上的优劣。

（1）MeSH 词表获取

利用美国国家医学图书馆官网获取结构化存储的 MeSH 词表，并对词表文档中存储的主题词（Term）、概念（Concept）以及树形编码（TreeNumber）等信息进行解析并抽取，用于 MeSH 网络的构建[41]。进一步通过人工方式，搜集本研究选取的 7 570 篇论文的 MeSH 主题词及其相应的 TreeNumber，用于对合作群体的主题一致性进行评估。

（2）MeSH 语义网络构建

利用 MeSH 主题词之间的上下位类关系，可将 MeSH 词表组织成一棵语义树[41]。具体如下：首先，

从词表（下载时间为2023年1月15日）中抽取所有的MeSH主题词（包含TreeNumber），共计30 452个；其次，将MeSH主题词所对应的TreeNumber根据上下位关系构建主题词间的边，其中只有具备父子关系的主题词之间才形成边[41]。通过此法，本研究构建的MeSH语义网络共含节点30 452个，边41 487条。

（3）MeSH主题词的词向量计算

利用node2vec算法[42]，按照默认的参数设定，对本研究构建的MeSH语义网络进行表示学习，将每个节点学习为200维的向量。

（4）主题一致性计算

首先，利用每篇文章标注的MeSH主题词，计算每篇文章的主题向量$\boldsymbol{v}_i$，计算方法为

$$\boldsymbol{v}_i = \frac{1}{n}\sum_{k=0}^{n}\vec{e}_k^i \tag{1}$$

其中，$\boldsymbol{v}_i$是其所有标注主题词的平均向量；$\vec{e}_k^i$表示文章i中第k个MeSH主题词的embedding。

对某个合作群体g的所有发文，两两求主题向量的余弦相似度，再求均值。所得到的相似度均值作为该合作群体g发文主题的一致性指标，记为C_g，其计算公式为

$$C_g = \frac{n(n-1)}{2}\cdot\sum_{i,j=0}^{n}\frac{\boldsymbol{v}_i^g\cdot\boldsymbol{v}_j^g}{\left\|\boldsymbol{v}_i^g\right\|\cdot\left\|\boldsymbol{v}_j^g\right\|},\quad i\neq j \tag{2}$$

其中，$\boldsymbol{v}_i^g$表示合作群体g所发表文章i的主题向量。

通过对合著者网络和合贡献者网络在细分和重组两类合作群体发文主题一致性的分析检验，进一步从专家知识角度验证本研究提出的合贡献者网络在合作群体识别上的特性。

3 实验结果

3.1 合贡献者网络的结构特征

表3列出了合贡献者网络与合著者网络最大连通子图的主要结构特征[43]。合贡献者网络的网络密度为0.000 19，表明在本研究的数据集上构建的合作网络较为稀疏。较之合著者网络，合贡献者网络密度更低，主要源于合贡献者网络中合作者间形成边的概率会低于合著者网络，这与合贡献者网络的构建原理保持了一致。合贡献者网络的平均邻居数为10.136，比合著者网络（11.714）略低，表明了合贡献者网络更为稀疏的特点。合贡献者网络的节点间平均距离为11.349，最长距离为32；合著者网

表3 合著者网络与合贡献者网络结构特征指标统计表

指标	合贡献者网络	合著者网络
论文总数	7 570	7 570
作者总数	53 757	53 757
每个作者平均论文数	0.14	0.14
每篇文章平均作者数	7.1	7.1
节点数	53 757	53 757
边数	272 430	314 854
网络密度	0.000 19	0.000 22
平均邻居数	10.136	11.714
平均节点距离	11.349	11.041
最长节点距离	32	30
模块度	0.978	0.972
连通子图数量	5 912	4 762
最大连通子图节点数	9 233	9 501
最大连通子图占比	17.2%	17.7%
最大连通子图边数	94 098	113 335
最大连通子图节点平均邻居数	20.383	23.857
最大连通子图最长节点距离	32	30
最大连通子图平均路径长度	11.439	11.131
最大连通子图模块度	0.873	0.835
最大连通子图密度	0.002	0.003
第二大连通子图节点数	306	314

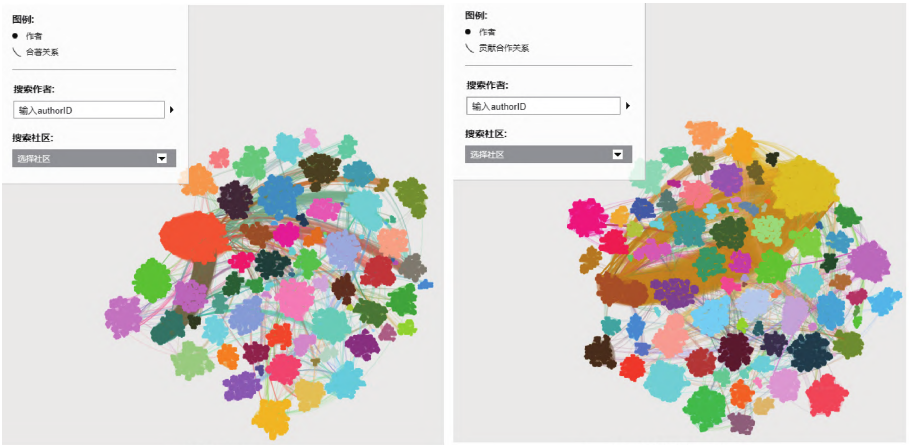
络的节点间平均距离为 11.041，最大距离为 30。这表明较之合著者网络，合贡献者网络的连通性相对较差。合贡献者网络的模块度为 0.978，比合著者网络的模块度（0.972）稍高，这表明两种合作网络的社区结构性质都比较强，其原因可能是在本研究的数据集上构建的合作网络相较于其他研究中的网络都比较稀疏[43]，网络中有较多的子图数量。

此外，表 3 还给出了合著者网络连通子图的结构特征。总体而言，合贡献者网络中共有子图 5 912 个。相较于合著者网络而言，合贡献者网络的子图数更多，显示了合贡献者网络稀疏的特点。从第二大连通子图的大小来看，尽管两种网络的子图数量较多，但除了最大连通子图以外，其他子图包含的节点数都偏小。从最大连通子图来看，合贡献者网络的最大连通子图包含的节点数、边数及子图密度都小于合著者网络的最大连通子图。同时，合贡献者网络的最大连通子图也显示出更长的平均节点距离。类似地，合贡献者网络最大连通子图的模块度为 0.873，较之合著者网络（0.835）也更高一些。总体来看，两种合作网络的最大连通子图间的区别与整体网络间的差异基本保持一致，即合贡献者网络比合著者网络更稀疏、连通性稍差，但网络的模块度指标更高。

3.2　合作群体识别结果分析

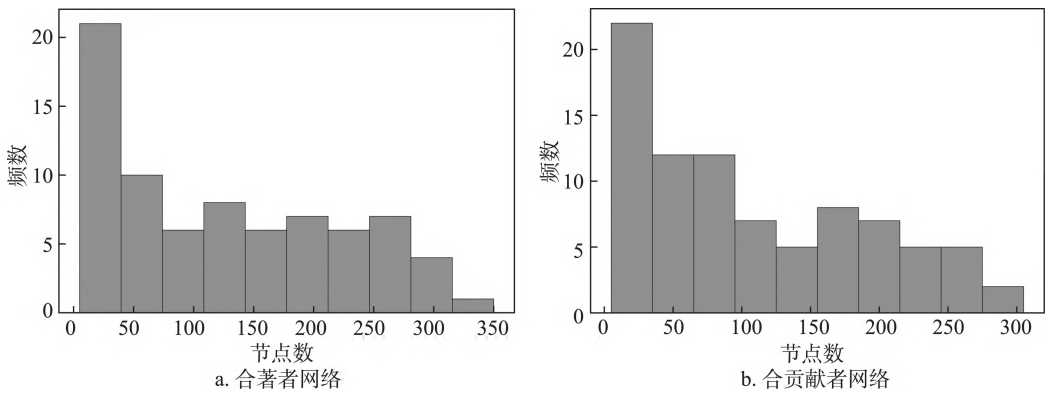
3.2.1　合作群体结构分析

本研究利用合著者网络和合贡献者网络的最大连通子图作为合作群体识别的对象。两种网络最大连通子图社区结构划分的可视化结果如图 5 所示。总体来看，两种合作网络中所识别的合作群体结构具有较高的相似性。从合作群体所包含作者数的统计结果（图 6）来看，合作群体结构大小相当，但合贡献者网络中识别出的合作群体比合著者网络稍小。其可能的原因在于：①合贡献者网络的最大连通子图较合著者网络的节点数更少，因此，合作群体的大小受到影响；②更为细化的合贡献者网络密度较小，导致该网络中的社区结构的间隙更加松散，所识别的社区结构更多、社区的平均节点数更少。



a. 合著者网络　　b. 合贡献者网络

图 5　合著者网络与合贡献者网络最大连通子图社区划分结果可视化



a. 合著者网络　　b. 合贡献者网络

图 6　合著者网络与合贡献者网络的各社区节点数分布

为了更清晰地揭示两种合作网络中的群体划分情况，两种网络中部分群体结构变化的桑基图如图7所示[①]。统计过程中，如果社区结构间节点差异的比例小于各自社区节点数目的10%，那么该分支在后续分析都忽略不计。研究结果发现，“合作群体一致”的共44组，如图7中的n1-39与n2-39（n1表示合著者网络，n2表示合贡献者网络；n1-39表示合著者网络39号社区，以此类推），占比约57%；“合作群体细分”的共7组，如图7中的n1-37与{n2-57, n2-78, n2-10, n2-55}，涉及合著者网络8个社区，占比约11%，这8个合作群体细分成了合贡献者网络中的15个社区；“合作群体重组”的共11组，如图7中的{n1-28, n1-58}与{n2-40, n2-61}，涉及合著者网络中的24个合作社群，占比约32%，这24个合作群体重组成为合贡献者网络中的26个合作群体。两种网络间合作群体的映射关系如图8所示。

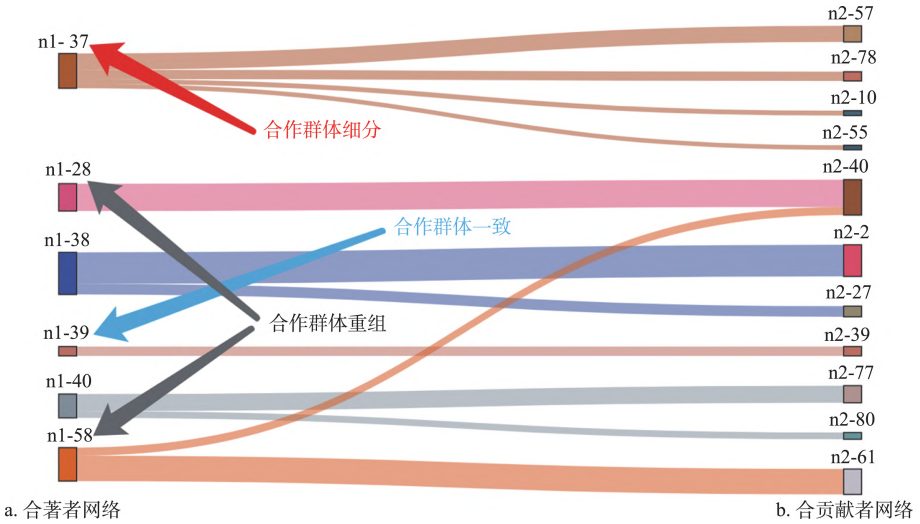


图7 合著者网络与合贡献者网络的社区结构桑基图(部分)

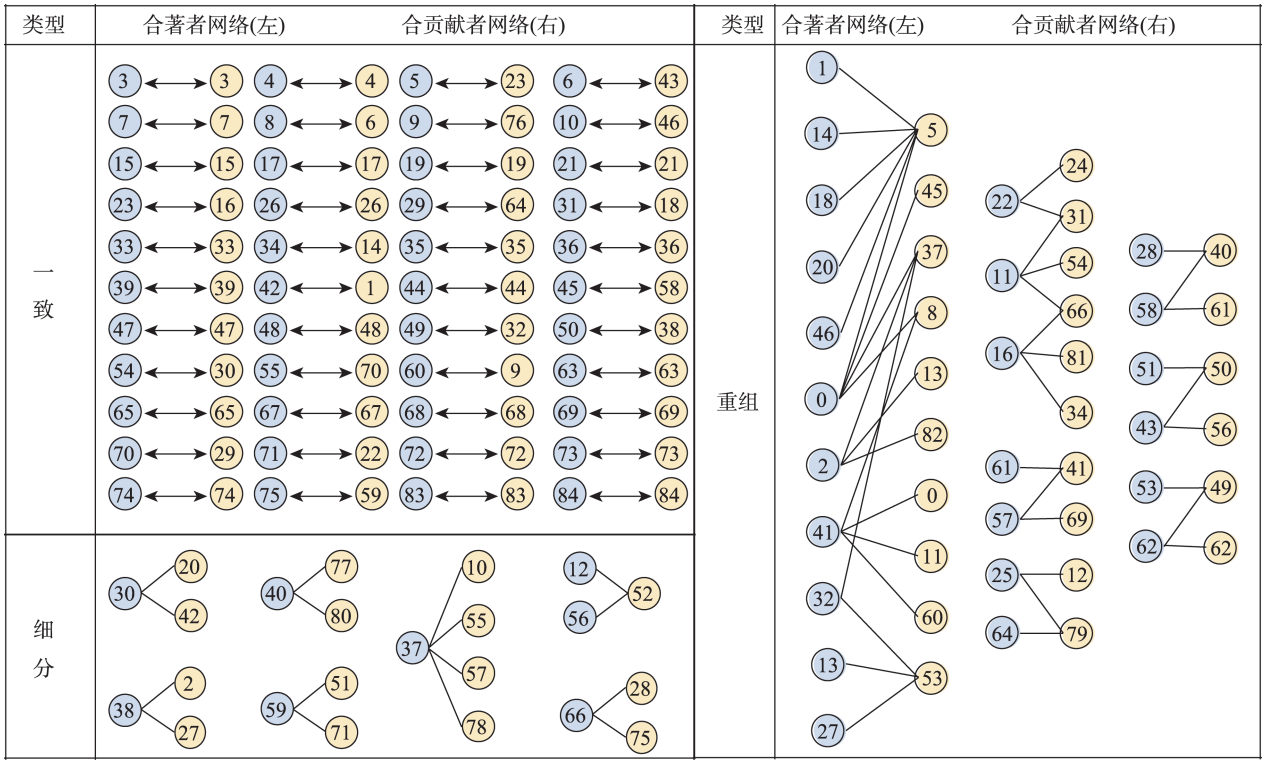


图8 两种合作网络的合作群体研究主题映射关系示意图

① 本研究的所有可视化结果见 http://c.nxw.so/c2qsj

3.2.2　合作群体发文主题分析

根据合作群体的分型，可将研究话题分型定义为“话题一致”“话题细分”和“话题重组”三种。通过抽取各合作群体所发文章的关键词，可以总结出所有合作群体的研究主题①。下文结合两种网络中话题变化的实例对三种关系进行示例描述。

（1）“话题一致”

示例：合著者网络的33号社区与合贡献者网络的33号社区。两个社区下均由31个作者构成同一个合作群体，共合作发表6篇文章。主题发现结果显示，这6篇文章的研究主题为“先天性心脏病”。

（2）“话题细分”

示例一：合著者网络的30号社区细分成合贡献者网络的20号社区和42号社区，各社区的结构与主题情况如表4所示。具体来看，合著者网络30号社区共216位作者，合作26篇文章，研究主题为“肾病与心血管疾病”。合贡献者网络20号社区共30位作者，合作6篇文章，研究主题为“肾病”；合贡献者网络42社区共169位作者，合作发表21篇文章，研究主题为“心血管疾病”。图9展示了该示例中两种网络合作群体的细分情况。可以发现，合著者网络中的30号社区结构包含两个较明显的子社区结构；而在合贡献者网络中，这两个子社区结构被划分成20号和42号两个独立社区。由此可见，合贡献者网络将合著者网络30号话题中的两个子话题做了分割，即将心血管疾病和肾病这两个子话题做了更细致的区分。

表4　“话题细分”示例一的合作群体网络结构特征及其主题标签

社区编号	节点数	边数	平均网络长度	主题标签
n1-30	216	1 197	4.952	肾病与心血管疾病
n2-20	30	110	2.483	肾病
n2-42	169	941	4.295	心血管疾病

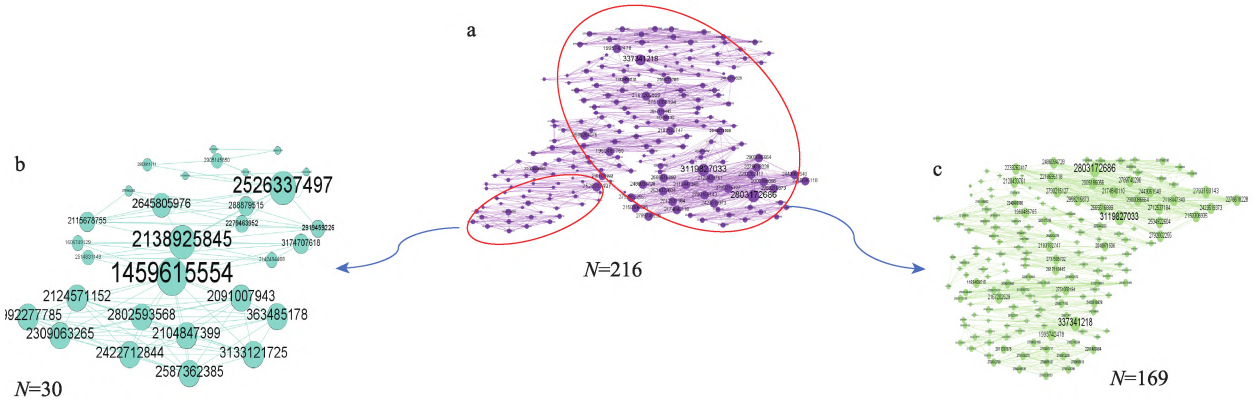


图9　“话题细分”示例一网络结构细分示意图

a代表合著者网络30号社区，b代表合贡献者网络20号社区，c代表合贡献者网络42号社区。

示例二：合著者网络40号社区在合贡献者网络中分成了77号社区和80号社区，其结构与主题信息如表5所示。具体来看，合著者网络40号社区共157位作者，合作21篇文章，研究主题为“心肌病”。合贡献者网络77号社区共110位作者，合作10篇文章，研究主题为“心血管疾病”；合贡献者网络80社区共43位作者，合作发表12篇文章，研究主题为“房颤”。两组社区结构如图10所示。可以发现，合著者网络40号社区包含两个较明显的子社区；在合贡献者网络中，这两个子社区结构被划分成77号和80号两个独立的社区。由此可见，合贡献者网络将合著者网络40号话题中的两个子话题做了分割，即将“心血管疾病”和“房颤”这两个子话题做了更细致的区分。

表5　“话题细分”示例二合作群体网络结构特征及其主题标签

社区编号	节点数	边数	平均网络长度	主题标签
n1-40	157	1 520	3.556	心肌病
n2-77	110	806	3.328	心血管疾病
n2-80	43	155	3.398	房颤

（3）“话题重组”

示例一：合著者网络43号合作群体下共265位作者，合作发表34篇文章，研究主题为“结核

① 合作群体的研究主题及关键词抽取结果见 http://c.nxw.so/axbrH

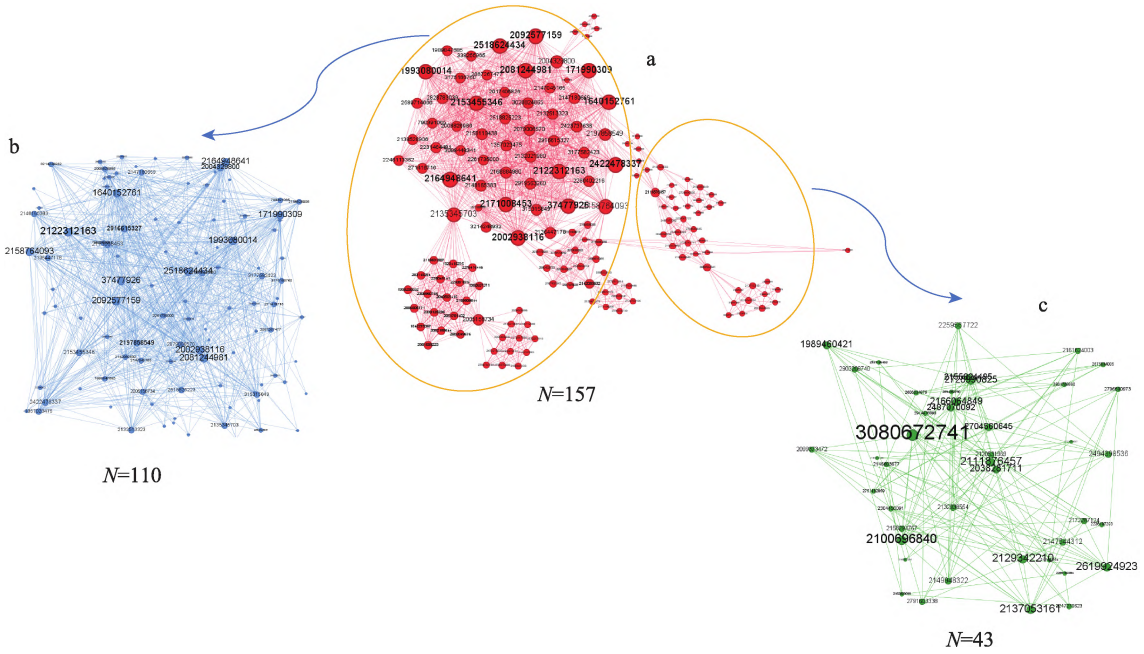


图 10 “话题细分”示例二网络结构细分示意图

a 代表合著者网络 40 号社区，b 代表合贡献者网络 77 号社区，c 代表合贡献者网络 80 号社区。

病”。其中，26 篇文章的合作者在合贡献者网络中成为一个单独的分支（第 56 号社区），其研究主题词为“结核病诊断”；另外 8 篇文章与合著者网络 51 号社区的 15 篇文章重组成为合贡献者网络 50 号社区下的文章，重组后新的研究主题为“HIV（human immunodeficiency virus）并发结核病”。合贡献者网络的 50 号社区关键词为合著者网络 43 号社区和 51 号社区关键词的综合（表 6），两组的社区结构如图 11 所示。

表 6 “话题重组”示例一合作群体网络结构特征及其主题标签

社区编号	节点数	边数	平均网络长度	主题标签
n1-43	265	2 654	3.486	结核病
n1-51	127	703	3.612	HIV 与 ART（atraumatic restorative treatment）治疗
n2-50	175	1 024	4.591	HIV 并发结核病
n2-56	203	2 172	3.335	结核病诊断

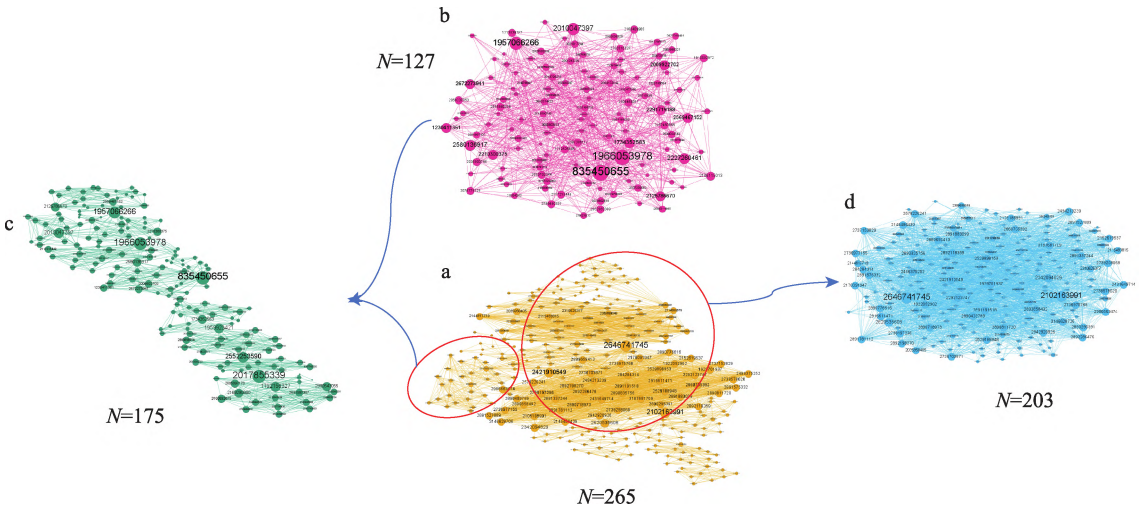


图 11 “话题重组”示例一网络结构重组示意图

a 代表合著者网络 43 号社区，b 代表合著者网络 51 号社区，c 代表合贡献者网络 50 号社区，d 代表合贡献者网络 56 号社区。

示例二：合著者网络 58 号合作群体下共 235 位作者，合作发表 26 篇文章，研究主题为“HIV 与寄生虫病”。其中，163 位作者在合贡献者网络中成为第 61 号社区，其研究主题词为“寄生虫病”；其余 52 位作者与合著者网络 28 号社区的 176 位作者重组成为合贡献者网络 40 号合作群体，重组后新的研究主题为“HIV”。合著者网络 58 号社区关键词为合著者网络 28 号社区和合贡献者网络 61 号社区关键词的综合（表 7）。两组的社区结构如图 12 所示。

表 7　“话题重组”示例二合作群体网络结构特征及其主题标签

社区编号	节点数	边数	平均网络长度	主题标签
n1-58	235	1 633	4.179	HIV 与寄生虫病
n2-61	163	1 104	4.013	寄生虫病
n1-28	179	1 964	2.838	HIV
n2-40	228	2 038	3.699	HIV

3.3　合作群体识别结果评估

为了检验合贡献者网络在合作群体识别应用上的效果，本研究利用 MeSH 词表，辅助构建合作群体发文的主题一致性指标 coherence。本研究将细分和重组两类合作群体的变化关系进行了分析和检验，如表 8 所示。合贡献者网络中，细分话题的主题一致性均值为 0.879，高于合著者网络中的主题一致性均值 0.866（p=0.18），不具备统计上的显著性；类似地，在重组群体中，合贡献者网络的话题一致性为 0.867，也稍高于合著者网络的主题一致性均值 0.865（p=0.78），但也不具备统计上的显著性。从整体上看，两种网络在合作群体识别上的主题一致性差异不大。

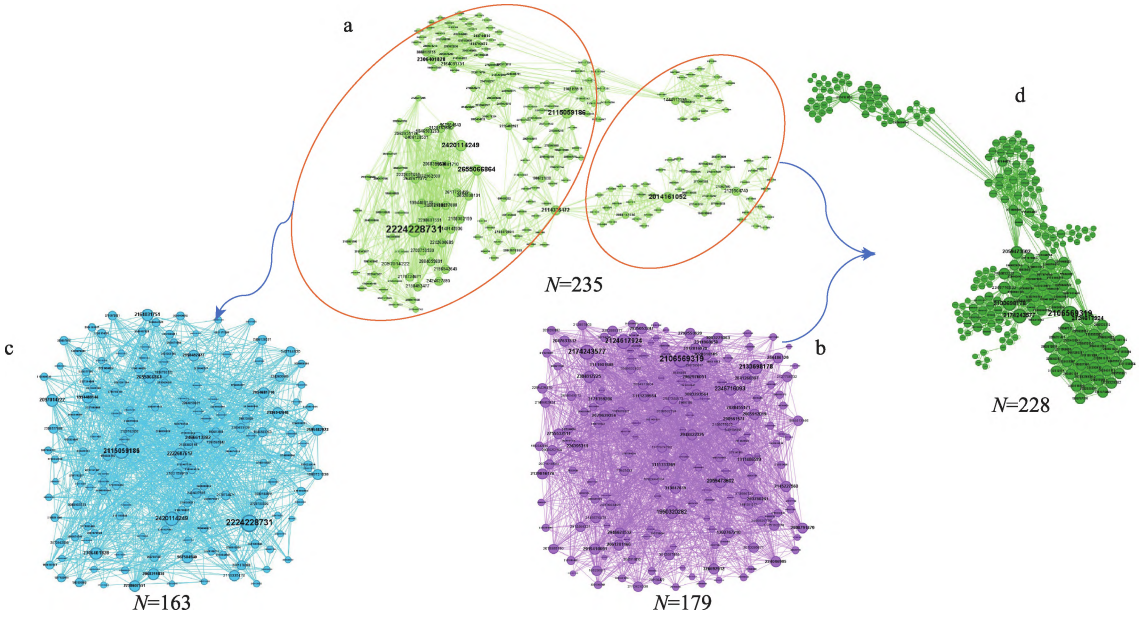


图 12　“话题重组”示例二网络结构重组示意图

a 代表合著者网络 58 号社区，b 代表合著者网络 28 号社区，c 代表合贡献者网络 61 号社区，d 代表合贡献者网络 40 号社区。

表 8　合著者网络与合贡献者网络研究主题一致性水平

组别	合贡献者网络			合著者网络			显著性水平
	均值	标准差	样本个数	均值	标准差	样本个数	
细分群体	0.879	0.010	15	0.866	0	8	0.18
重组群体	0.867	0.027	23	0.865	0.016	19	0.78
总体	0.868	0.024	38	0.865	0.015	27	0.57

注：本研究采用的是双样本异方差 t 检验（双边），合著者网络中有 5 个合作群体仅发表 1 篇论文，合贡献者网络中有 3 个合作群体仅发表 1 篇论文，在检验时，剔除了这 8 个样本。

4　讨论与结论

精准认识科研合作中的分工模式以及成员间的互动关系具有重要意义。常用于科研合作模式研究的合著者网络与科研合作实践相左，作者贡献声明数据的出现为揭示更细粒度的合作实践提供重要素

材。为此，本研究提出一种利用贡献声明数据构建新型合作网络——合贡献者网络，为深入研究科研合作模式问题提供新工具。本研究以药学为例，以合著者网络为基准，从合贡献者网络的网络结构特征入手，认识此新型合作网络的物理性质；从合作群体识别入手，进一步认识合贡献者网络的应用价值。

本研究的实验结果总结如下。

（1）合贡献者网络较之合著者网络显示出更良好的社区结构。由于引入了贡献合作关系，合贡献者网络的网络密度更低，但这并没有破坏合贡献者网络的整体结构，反而在社区结构的属性上展现了更显著的特征，为社区发现类应用预备了良好的条件。从后续的合作群体识别中可以看到，尽管有57%的社区结构在合贡献者网络以及合著者网络中都能够发现，但在合贡献者网络中还是发现了较多的更为精细或重组的社区结构。这意味着合贡献者网络的提出为这一类型的研究任务可能带来更多不同的社区识别结果，特别是其中较多有价值的细分和重组群体，表明这些不一样的识别结果具有一定的参考价值。这也显示出贡献合作关系更能揭示合作者之间密切的合作关系，而较为粗线条的合著关系并不能有效刻画学者间密切细微的合作关系[8,44]。

（2）合贡献者网络有助于识别出更细粒度的合作群体。从社区发现的结果来看，合贡献者网络上所识别的社区结构大小整体上低于经典的合著者网络。其原因可能在于合贡献者网络更为稀疏、社区结构较好，因贡献合作关系所提供的筛选机制使得网络能聚焦于学者的频繁研究行为，而不被偶然的合作行为所影响。这样的机制呈现了更为聚焦的合作群体识别效果：合著者网络中的6个合作群体在合贡献者网络上被识别为14个独立的合作群体。这为进一步分析合作群体的研究兴趣、主题迁移等问题[45]提供了良好的跟踪线索。

（3）合贡献者网络上，所识别的合作群体发文主题一致性更高。两种合作网络不仅在合作群体的数量以及体量上存在一定差异，而且在所识别出的合作群体发文主题上也存在一定差异。不管是在细分群体上或是在重组群体上，在合贡献者网络上识别的群体发文的主题一致性都略高于在合著者网络上的合作群体，因而合贡献者网络在本研究的合作群体识别任务上表现出更优的识别效果。但是，本研究的检验结果显示，在两种合作网络上所识别的合作群体发文主题一致性不存在显著的差异。其原因可能是在由MeSH词表网络得到的主题词的词嵌入向量具有一定的各向异性[46]，而导致词向量间相似度的计算结果具有普遍较高的相似度水平。

此外，本研究尚存在一定局限性。首先，社区结构发现算法的稳定性和有效性在一定程度上会影响相关研究的结论。本研究采用Gephi软件是考虑到该软件的广泛使用，在社区发现结果上具有一定的稳定性。后续研究可考虑利用其他算法（如图神经网络等社区发现算法）进行结果的性能分析。其次，本研究使用的微软学术知识图谱存在一定的PLoS数据收集不全的情况，其作者消歧数据正确率会在一定程度上影响研究结果。后续研究可考虑自行开发PLoS学科信息标注和作者消歧算法，以进一步提高研究数据的召回率。

基于本研究工作，提出三个未来可能的研究方向，为相关研究的进一步深化提供研究思路。①作者贡献声明预测模型构建。尽管当前许多影响力较高的刊物，如*Nature*、*Science*、*Cell*、*Lancet*等，均在披露作者贡献声明，并且会有更多的刊物会加入其中，但当前已发表作者贡献声明的期刊相对较少，这给作者贡献声明挖掘相关研究的深入，特别是应用研究的深度检验带来了障碍。其中一个可行的思路是利用人工智能算法来实现对大量无作者贡献数据的论文的合作者贡献进行预测[47-48]，以弥补数据方面的不足。②合贡献者网络的特性研究。本研究的合贡献者网络特性的研究仅建立在PLoS上的药学领域部分论文数据上。未来研究可进一步通过编制作者消歧算法、拓展数据集等方法，构建更大数据量、更多学科领域的合贡献者网络，探究该网络的物理性质，为进一步认识合作分工规律提供广泛证据。③基于合贡献者网络的学者合作机会预测等应用研究。本研究结果显示，合贡献者网络在揭示更精细、更聚焦的合作群体上具有其独特之处，未来研究可围绕合贡献者网络的这一特性开展更多的相关研究，如学者研究兴趣的形成及转移、学者在研究主题层面的影响力精细化评估等，进一步拓展合贡献者网络的应用研究范围。

参 考 文 献

[1] Larivière V, Gingras Y, Sugimoto C R, et al. Team size matters: collaboration and scientific impact since 1900[J]. Journal of the Association for Information Science and Technology, 2015, 66 (7): 1323-1332.

[2] Lu C, Zhang C W, Xiao C R, et al. Contributorship in scientific collaborations: The perspective of contribution-based byline or-

ders[J]. Information Processing & Management, 2022, 59(3): 102944.

[3] Birnholtz J P. What does it mean to be an author? The intersection of credit, contribution, and collaboration in science[J]. Journal of the American Society for Information Science and Technology, 2006, 57(13): 1758-1770.

[4] Leahey E, Reikowsky R C. Research specialization and collaboration patterns in sociology[J]. Social Studies of Science, 2008, 38 (3): 425-440.

[5] Amjad T, Ding Y, Xu J, et al. Standing on the shoulders of giants [J]. Journal of Informetrics, 2017, 11(1): 307-323.

[6] Allen L, Scott J, Brand A, et al. Publishing: credit where credit is due[J]. Nature, 2014, 508(7496): 312-313.

[7] Frische S. It is time for full disclosure of author contributions[J]. Nature, 2012, 489(7417): 475.

[8] Lu C, Zhang Y Y, Ahn Y Y, et al. Co-contributorship network and division of labor in individual scientific collaborations[J]. Journal of the Association for Information Science and Technology, 2020, 71(10): 1162-1178.

[9] Larivière V, Pontille D, Sugimoto C R. Investigating the division of scientific labor using the Contributor Roles Taxonomy (CRediT)[J]. Quantitative Science Studies, 2021, 2(1): 111-128.

[10] Yang S L, Xiao A X, Nie Y, et al. Measuring coauthors' credit in medicine field-based on author contribution statement and citation context analysis[J]. Information Processing & Management, 2022, 59(3): 102924.

[11] Corrêa E A, Silva F N, da F Costa L, et al. Patterns of authors contribution in scientific manuscripts[J]. Journal of Informetrics, 2017, 11(2): 498-510.

[12] Newman M E J. Coauthorship networks and patterns of scientific collaboration[J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101(Suppl 1): 5200-5205.

[13] Barabási A L, Jeong H, Néda Z, et al. Evolution of the social network of scientific collaborations[J]. Physica A: Statistical Mechanics and Its Applications, 2002, 311(3/4): 590-614.

[14] Zhang C W, Bu Y, Ding Y, et al. Understanding scientific collaboration: homophily, transitivity, and preferential attachment[J]. Journal of the Association for Information Science and Technology, 2018, 69(1): 72-86.

[15] Savić M, Ivanović M, Jain L C. Analysis of enriched co-authorship networks: methodology and a case study[M]// Complex Networks in Software, Knowledge, and Social Systems. Cham: Springer, 2019: 277-317.

[16] He C C, Wu J, Zhang Q P. Characterizing research leadership on geographically weighted collaboration network[J]. Scientometrics, 2021, 126(5): 4005-4037.

[17] Zhai L, Yan X B. A directed collaboration network for exploring the order of scientific collaboration[J]. Journal of Informetrics, 2022, 16(4): 101345.

[18] Yu S, Alqahtani F, Tolba A, et al. Collaborative team recognition: a core plus extension structure[J]. Journal of Informetrics, 2022, 16(4): 101346.

[19] Li E Y, Liao C H, Yen H R. Co-authorship networks and research impact: a social capital perspective[J]. Research Policy, 2013, 42 (9): 1515-1530.

[20] Amjad T, Daud A, Aljohani N R. Ranking authors in academic social networks: a survey[J]. Library Hi Tech, 2018, 36(1): 97-128.

[21] Ma G S, Qian Y H, Zhang Y Y, et al. The recognition of kernel research team[J]. Journal of Informetrics, 2022, 16(4): 101339.

[22] Chuan P M, Son L H, Ali M, et al. Link prediction in co-authorship networks based on hybrid content similarity metric[J]. Applied Intelligence, 2018, 48(8): 2470-2486.

[23] Orzechowski K P, Mrowinski M J, Fronczak A, et al. Asymmetry of social interactions and its role in link predictability: the case of coauthorship networks[J]. Journal of Informetrics, 2023, 17(2): 101405.

[24] Katz J S, Martin B R. What is research collaboration? [J]. Research Policy, 1997, 26(1): 1-18.

[25] 刘晓婷, 黄颖, 李瑞婧, 等. 内聚-耦合视角下科研团队合作模式识别与对比研究[J]. 情报科学, 2022, 40(12): 170-180.

[26] Zou B T, Wang Y F, Kwoh C K, et al. Directed collaboration patterns in funded teams: a perspective of knowledge flow[J]. Information Processing & Management, 2023, 60(2): 103237.

[27] 吕千千, 谭宗颖. 虚拟科研团队识别方法研究——以重症医学领域为例[J]. 图书情报工作, 2022, 66(15): 97-106.

[28] 李纲, 柳明飞, 吴青, 等. 基于蝴蝶结模型的科研团队角色识别及其特征研究[J]. 图书情报工作, 2017, 61(5): 87-94.

[29] 丁敬达, 王新明. 基于作者贡献声明的合著者贡献率测度方法[J]. 图书情报工作, 2019, 63(16): 95-102.

[30] 张梦莹, 章成志, 王杰. 不同学科的作者贡献分布差异研究——以图情和医学领域的四种期刊为例[J]. 图书馆论坛, 2018, 38(12): 112-119.

[31] Sinatra R, Deville P, Szell M, et al. A century of physics[J]. Nature Physics, 2015, 11(10): 791-796.

[32] 卢超, 章成志, 王玉琢, 等. 语义特征分析的深化——学术文献的全文计量分析研究综述[J]. 中国图书馆学报, 2021, 47(2): 110-131.

[33] 卢超, 董克. 文献耦合网络的引文内容加权研究——基于提及次数的方法[J]. 情报杂志, 2022, 41(11): 171-178.

[34] Lu C, Bu Y, Dong X L, et al. Analyzing linguistic complexity and scientific impact[J]. Journal of Informetrics, 2019, 13(3): 817-829.

[35] Lancichinetti A, Fortunato S. Community detection algorithms: a comparative analysis[J]. Physical Review E, 2009, 80(5): 056117.

[36] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast unfolding of communities in large networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2008, 2008(10): P10008.

[37] Syakur M A, Khotimah B K, Rochman E S, et al. Integration *k*-

means clustering method and elbow method for identification of the best customer profile cluster[J]. IOP Conference Series: Materials Science and Engineering, 2018, 336: 012017.

[38] Newman M E J. Fast algorithm for detecting community structure in networks[J]. Physical Review E, 2004, 69(6): 066133.

[39] Wang Q. A bibliometric model for identifying emerging research topics[J]. Journal of the Association for Information Science and Technology, 2018, 69(2): 290-304.

[40] Grootendorst M. MaartenGr/KeyBERT: BibTeX[CP/OL]. (2021-01-25). https://doi.org/10.5281/zenodo.4461265.

[41] Guo Z H, You Z H, Huang D S, et al. MeSHHeading2vec: a new method for representing MeSH headings as vectors based on graph embedding algorithm[J]. Briefings in Bioinformatics, 2021, 22(2): 2085-2095.

[42] Grover A, Leskovec J. node2vec: scalable feature learning for networks[C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2016: 855-864.

[43] Newman M E J. The structure of scientific collaboration networks [J]. Proceedings of the National Academy of Sciences of the United States of America, 2001, 98(2): 404-409.

[44] Devine E B, Beney J, Bero L A. Equity, accountability, transparency: implementation of the contributorship concept in a multisite study[J]. American Journal of Pharmaceutical Education, 2005, 69(4): 61.

[45] Jia T, Wang D S, Szymanski B K. Quantifying patterns of research-interest evolution[J]. Nature Human Behaviour, 2017, 1: Article No.0078.

[46] Li B H, Zhou H, He J X, et al. On the sentence embeddings from pre-trained language models[C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2020: 9119-9130.

[47] Robinson-Garcia N, Costas R, Sugimoto C R, et al. Task specialization across research careers[J]. eLife, 2020, 9: e60586.

[48] Xu F L, Wu L F, Evans J. Flat teams drive scientific innovation [J]. Proceedings of the National Academy of Sciences of the United States of America, 2022, 119(23): e2200927119.

（责任编辑 冯家琪）