

融合研究主题与关系网络结构的科研合作者推荐研究 ——以 CNKI 平台情报学领域的学者推荐为例

黄雪雯^{1,2} 杨建林^{1,2*}

(1. 南京大学信息管理学院, 江苏 南京 210023;
2. 江苏省数据工程与知识服务重点实验室, 江苏 南京 210023)

摘要: [目的/意义] 强针对性、高自由度的合作者推荐服务能够促进科研工作者之间的知识和资源共享, 提高科研质量与效率。[方法/过程] 基于混合推荐方法的思想, 从学者科研合作动机角度切入设计科研合作者推荐方案的框架并给出具体的实现过程。首先, 针对寻求主题相似、主题指定或主题互补等合作者的需求, 设计多项基于学者研究主题的推荐指标并给出计算方案; 其次, 面向寻求权威合作者的需求, 设计多项基于关系网络结构的推荐指标并给出计算方案; 最后, 基于学者按需求提供的权重将两部分指标综合集成。[结果/结论] 以情报学领域为例对论文所提科研合作者推荐方案进行实证研究。研究结果表明, 结合领域内研究主题与学者关系网络结构的科研合作者推荐方案能够准确地识别出社会资本强度更高的备选合作者, 能够更好地适应科研合作者推荐的个性化需求。

关键词: 研究主题; 关系网络结构; 科研合作者; 推荐; BERTopic 模型

DOI: 10.3969/j.issn.1008-0821.2024.08.012

[中图分类号] G250.252 [文献标识码] A [文章编号] 1008-0821 (2024) 08-0124-13

Research on Scientific Collaborator Recommendation by Fusing Cooperation Networks and Research Topics

——Taking the Recommendation of Scholars in the Field of Information
Science on CNKI Platform as an Example

Huang Xuewen^{1,2} Yang Jianlin^{1,2*}

(1. School of Information Management, Nanjing University, Nanjing 210023, China;
2. Jiangsu Key Laboratory of Data Engineering and Knowledge Service, Nanjing 210023, China)

Abstract: [Purpose/Significance] Highly targeted and highly free collaborator recommendation services can promote the sharing of knowledge and resources among scientific researchers, and enhance the quality and efficiency of scientific research. [Methods/Process] Based on the idea of the hybrid recommendation method, the paper designed a framework of the recommendation scheme of scientific collaborators from the perspective of scholars' motivation for scientific cooperation and gave a specific implementation process. Firstly, the paper designed several recommendation indicators based on scholars' research topics to meet the needs of collaborators pursuing similarity, topic designation, or topic complementarity. Secondly, the paper designed some recommendation indicators based on the cooperation network to meet the needs of those seeking authoritative collaborators and provided a calculation scheme. Finally, the paper synthesized those indicators based

收稿日期: 2023-11-01

基金项目: 国家社会科学基金重点项目“统筹科技发展和科技安全的情报学理论与方法研究”(项目编号: 22ATQ007); 南京大学新时代文科卓越研究计划“中长期研究专项”项目“‘数智赋能’背景下的情报学理论、方法与应用研究”。

作者简介: 黄雪雯 (1999-), 女, 硕士研究生, 研究方向: 智能化信息处理。

通信作者: 杨建林 (1970-), 男, 教授, 博士, 博士生导师, 研究方向: 情报学基础理论、智能化信息处理、科技安全。

on the weights requested by scholars. [Result/Conclusions] Taking the field of information science as an example, the paper conducts an empirical study on the proposed scientific collaborator recommendation scheme. The results show that the scientific collaborator recommendation scheme combined with the research topics in the field and the relationship network of scholars can accurately identify the candidate collaborators with higher social capital intensity, and can better adapt to the personalized needs of the scientific research collaborators.

Key words: research topic; cooperation network; scientific collaborator; recommendation; BERTopic model

科研领域学者间的科研合作能够有效推动科学研究问题的高效解决并显著提高科研成果的质量^[1],同时促进学者的科研产出量的提高^[2],成为当前科研活动中的普遍现象。当代科学各分支学科在高度分化基础上的跨学科属性增强^[3],高质量的合作能够实现各学科领域资源共享与优势互补,有助于提升科研产出的学术影响力^[4-5]。因此,面向寻求潜在科研合作对象的现实需求,科研合作者推荐方法成为学界广泛关注的研究热点。当前科研合作者推荐方法多样化、推荐视角多元化,推荐目标多为发现潜在合作对象,但是没有充分考虑用户科研合作需求的多样性。例如,跨领域学者希望合作者能够弥补专业知识的不足,科研领域新人更希望与领域内权威学者合作,权威学者一般对新合作者的需求并不迫切等。合作需求的多样性源于合作动机的多样性,覆盖学者研究主题与学者合作关系网络两方面。为此,本文在梳理科研合作动机的基础上,以学者研究主题与关系网络结构为支点设计指标,构建个性化合作者推荐方案,为已有明确研究目标的科研工作者提供高自由度的个性化推荐服务,以期科研合作者个性化推荐方法的改进与完善提供可能的解决思路。

1 相关研究

1.1 科研合作动机

学者的科研合作动机复杂多样,受多种因素影响^[6]:一方面,国家政策、机构体制、经济文化发展水平、学科发展趋势等外部因素在促使学者寻求合作共赢^[7-8];另一方面,知识、技术、资源、目标等个人因素使学者寻求伙伴资源。例如,Beaver D D^[9]梳理了学者展开合作的18个潜在动机,包括获取专业知识、接触缺失的资源、提高声望促进职业发展、提高生产力、建立或保持学术链接等。Melin G^[10]综合半结构化访谈与问卷调查结果,认为学者的合作动机有合作者具有特殊能力、数据、设

备,想要尝试新方法以及出于社交关系、师生关系的需要。Birnholz J P^[11]认为,学者的合作倾向由两部分决定:一是合作能否提供专业的知识、仪器等研究所需的资源;二是研究人员本身是否认为合作能够帮助其提高个人声誉、维护职业道路。Garcia A等^[12]通过调研100名拉丁美洲学者建立合作关系的动机,得到学者合作动机来自寻求更重要的结果、提高生产力、提高成果知名度与认可度、获得新方法或研究资源4个目标。在国内,马凤等^[13]借鉴Melin G^[10]的结论构建问卷,验证了大部分学者的合作动机主要由两方面构成,一是合作者与其具有共同兴趣或合作者可弥补自身知识储备方面的不足。二是维持原有合作关系或为提高研究影响力寻求权威合作者。赵君等则将学者科研合作动机分为生产动机(提高效率)、经济动机(降低科研成本)以及社会动机(提高学者声誉)3个部分^[14]。

此外,全球化、科研跨学科属性增强等发展趋势下,学者的跨国合作动机与跨学科合作动机也成为相关研究内容之一。1998年,Georghiou L^[15]探究了国家层面的跨国合作动机,认为对研究的直接利好与对参与者的间接利益是促成国家间合作的主要动机。Dusdal J等^[16]对科研工作者的跨国合作动机进行研究,认为除获取资金、共享知识设备实验室等资源、扩展关系网络、提高知名度等外,更多的个人原因在于维持与同行的友谊与提高科研产出。Maglaughlin K L等^[17]通过半结构化访谈调研自然科学跨学科合作的影响因素,并将其分为个人、资源、动机以及共同点4类,共20个因素,囊括拥有专业知识、拥有资金设备等资源、获得发现并解决新问题的快乐等。

综合上述研究,合作者拥有更专业的知识,合作以共享更多的设备、技术等资源,维持原有关系,提升学者个人声誉与学术地位等合作动机得到许多学者的认同。

1.2 学者合作推荐方法

常见的学者合作推荐方法可分为3类：基于内容的推荐方法、基于结构的推荐方法以及结合前两种推荐路径的混合推荐方法。

基于内容的推荐方法通过对学者成果的内容进行分析以获得学者的研究兴趣，进而比较学者研究兴趣间的相似性，为学者推荐与其兴趣最相近的学者。常用的分析方法有文本挖掘、自然语言处理以及科学计量等，多基于LDA模型^[18-19]、Word2vec模型^[20]等主题模型进行分析。近年来，部分学者进一步从动态视角深入探究学者研究兴趣随时间的变化情况^[21-22]，为学者个性化推荐模型提供了新视角。此外，专家发现系统也被认为是一种重要的合作者推荐方法^[23]。

基于结构的推荐方法则是利用拓扑结构分析来挖掘学者间各种网络关系，通过比较节点相似性或预测节点链接的可能性以确定推荐与否，多采用基于节点的推荐模型与基于路径的推荐模型^[24]。基于节点的推荐模型常通过公共邻居索引^[25]、随机游走索引^[26]等方法计算节点间结构相似性，为关系网络中关系结构相近的学者推荐合作。例如，Chen Y等^[27]学者则改进随机游走算法，构建融合机构网络和合著网络的异构网络，通过“基于异构网络的随机游走”获得学者推荐结果。基于路径的推荐模型则是通过链接预测的方法为可能产生直接路径的学者推荐合作，多采用相似性算法、最大似然法、概率模型^[28]和堆叠模型等来实现。例如，Chi K等^[29]学者提出了一种基于节点间吸引力的动态网络链路预测方法。

结合前两种推荐路径的混合推荐方法旨在通过更全面的考虑提供更准确的建议，其结合方式较为多样。部分学者将主题内容嵌入学者关系网络，基于对异构网络的分析进行推荐，例如，Chen J等^[30]构建包含协作关系和潜在语义关系的内容增强网络嵌入模型，为目标研究人员推荐相似合作者；Liu X Y等^[31]将共同研究主题、学者间合著关系、引文关系以及现实地点关系嵌入异构网络以实现科学合作者推荐。部分学者利用权重分配、投票得分等方式结合两个方向的推荐结果，例如，Xi X W等^[32]利用Word2Vec模型基于上下文计算学者之间主题

的相似性，再利用Node2Vec模型提取学者的网络拓扑特征，最后使用CombMNZ方法融合两个相似性度量的结果获得潜在合作者推荐列表。还有部分学者借鉴其他领域知识与技术来实现推荐，例如，Li P等^[33]基于学者知识图、用户—学者交互网络构建高级知识图谱为学者推荐合作实体；董文慧等^[34]构建用户画像，综合学者研究主题兴趣与社会网络为学者推荐合作者。

基于内容的推荐方法存在两方面的问题：一是相似性分析的准确性会受到不同主题分析方法对文本语义理解程度差异的影响。二是仅仅关注内容信息而忽视学者关系这一重要信息。基于结构推荐的方法存在以下问题：其一，与基于内容的推荐方法相反，基于结构的方法常忽略学者节点的内容信息，如论文的语义、学者的学术课题等；其二，拓扑结构分析预测结果的解释性较低。结合前两种推荐路径的混合推荐方法充分考虑了学者研究主题与关系网络两方面的内容，但是其结合方法多样，援引理论繁多，目前尚未形成一个被学界统一接受的理论框架。此外，就当前已有推荐方法而言，如何真正实现对合作者的个性化推荐仍是亟需突破的一大难题。

2 科研合作者推荐方案设计

2.1 推荐方案框架设计

鉴于数据的可获取性以及指标的可量化性，本文只考虑两种代表性的合作动机，并将其转化为学者对合作者在研究兴趣与专业知识储备方面的需求，以及学者维持原有合作关系或寻求权威合作者的需求。为此，本文以学者研究主题内容与关系网络结构为支点构建科研合作者个性化推荐方案，所设计合作者推荐方案属于混合推荐方法，总体框架如图1所示。

首先，将学者对合作者的需求分解为研究主题、权威影响两个方面，再对学科领域内学者的研究主题与学者合著关系网络进行分析与利用。在研究主题方面，本文针对主题相似、主题互补以及主题指定3种主题合作类型分别设定一项推荐指标，并给出计算方案。在关系网络结构方面，本文设定学者社会资本强度、备选学者与需求学者的合作可能两项推荐指标，并给出计算方案。最终，综合前述指

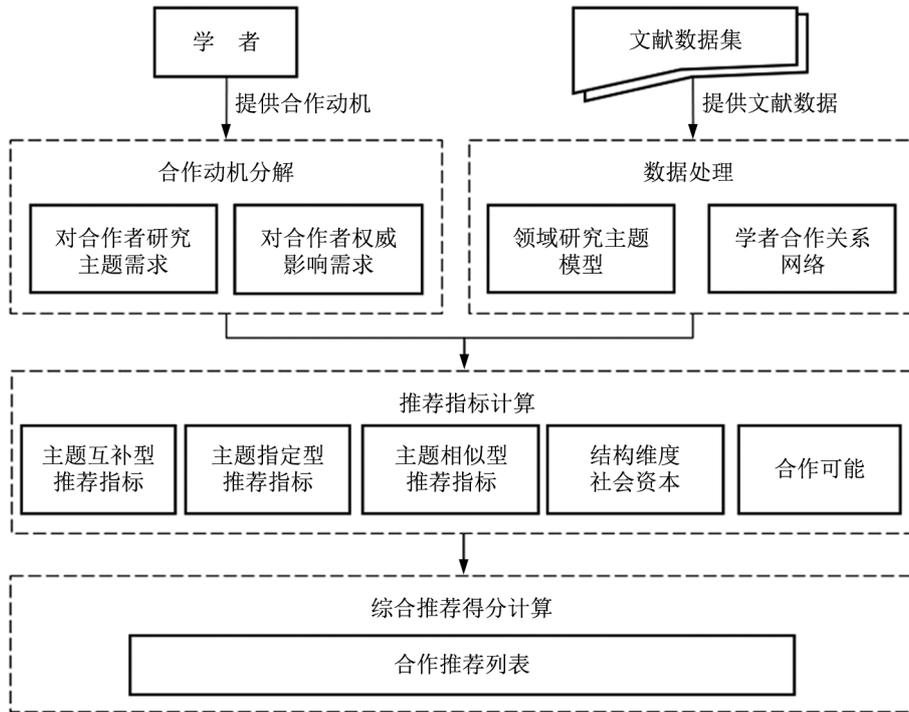


图1 科研合作者推荐方案的总体框架

Fig. 1 General Framework of the Scientific Collaborator Referral Program

标得出推荐得分最高的前 k 名推荐合作者与各指标得分靠前的学者列表，以供学者充分了解备选合作者信息并自主抉择对结果的采纳程度。

2.2 基于学者研究主题的推荐指标设计

学者在寻找合作者时，为弥补自身知识储备的不足，多对合作者擅长的主题有特定的要求。本文以主题相似、主题指定、主题互补3种需求情境设计基于学者研究主题的并列合作者推荐指标。由于指标依托具体的研究主题，需选用合适方法挖掘领域内科研主题。

Bert 是 2018 年 Google 提出的语言模型，与 LDA (Latent Dirichlet Allocation) 主题模型相比，Bert 模型单独对文档进行嵌入式编码，克服 LDA 主题模型忽略词顺序与深层语义的缺点，所得词向量泛化能力更强，对文本中语义关系捕捉更充分^[35]。基于共同的词向量，BERTopic 模型改进 Top2Vec 主题模型的主题获取方式，相较其能够更准确表达主题信息^[36]。同时，BERTopic 模型能够给出在新文本中发现不同主题的概率，帮助推荐方案进行针对性的主题分析，更好地满足其主题指定、主题互补的合作动机。因此，本文选用 Bert 的衍生主题模型 BERTopic 模型挖掘领域内研究主题。

BERTopic 模型可为每一篇文档生成文档向量，并通过对文档向量的无监督聚类挖掘领域内所有主题形成学科主题模型。向量间距离越短，在向量空间中位置越接近，本文利用模型所得文档向量，基于向量间的距离构建推荐指标。首先，确定距离公式，本文选用欧氏距离作为两向量间距离的计算公式，如式 (1) 所示：

$$d = |\vec{r}_x \vec{r}_y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

\vec{r}_x 、 \vec{r}_y 为文档向量空间中的向量， n 为向量 \vec{r}_x 、 \vec{r}_y 的维度， d 为向量 \vec{r}_x 、 \vec{r}_y 间的距离。

其次，丰富向量空间，基于表征单独文档的文档向量构建表征学科主题、学者的新向量，用以计算主题间、学者间以及主题与学者间的距离。

主题中心向量，用以表征主题在文档向量空间的位置，其计算方式如式 (2) 所示：

$$\vec{r}_{c,t} = \frac{1}{n} \sum_{i=1}^n \vec{r}_{i,t} \quad (2)$$

其中， n 为分布在主题 t 中的文档总数， $\vec{r}_{i,t}$ 为主题 t 中第 i 份文档的文档向量，对主题 t 中所有文档向量取平均得到主题 t 的中心，表征主题 t 在文档向量空间的位置。

学者主题中心向量，表征文档向量空间中学者

在每个主题下的位置，其计算方式如式(3)所示：

$$\vec{r}_{c,s,t} = \frac{1}{n} \sum_{i=1}^n \vec{r}_{i,s,t} \quad (3)$$

其中， n 为学者 s 分布在主题 t 中的产出文档总数， $\vec{r}_{i,s,t}$ 为学者 s 在主题 t 中第 i 份产出文档的文档向量，对主题 t 中所有学者 s 所产出文档的文档向量取平均得到主题 t 中学者 s 的中心，表征主题 t 中学者的位置。

最大主题距离，表征主题内最大距离，其计算方式如式(4)所示：

$$d_{t,max} = \text{Max} \left(\bigcup_{i=1}^m |\vec{r}_{c,t} - \vec{r}_{c,i,t}| \right) \quad (4)$$

其中， $\vec{r}_{c,t}$ 为主题 t 的主题中心向量， m 为主题 t 中有产出的学者总数， $\vec{r}_{c,i,t}$ 为在主题 t 中有产出的学者的学者主题中心向量， $d_{t,max}$ 为主题 t 最大主题距离。

最大主题学者距离，表征特定主题中所有学者与学者的最大距离，其计算方式如式(5)所示：

$$d_{t,n,max} = \text{Max} \left(\bigcup_{i=1}^m |\vec{r}_{c,n,t} - \vec{r}_{c,i,t}| \right) \quad (5)$$

其中， $\vec{r}_{c,n,t}$ 为学者 n 在主题 t 中的学者主题中心向量， m 为主题 t 中有产出的学者总数， $\vec{r}_{c,i,t}$ 为在主题 t 中有产出的学者的学者主题中心向量。 $d_{t,n,max}$ 为主题 t 中学者 n 的最大主题学者距离。

最后，分析主题相似、主题指定以及主题互补的具体需求情境，结合上述向量与距离，构建基于距离的学者研究主题推荐指标，为学者推荐与其或与其研究目标距离最短的学者。

2.2.1 主题相似型合作者推荐指标

主题相似的需求情境是学者期望找到与自己研究兴趣相近的合作者，主题相似型合作者推荐指标将筛选出在学者所涉及的研究主题中与学者距离最近的其他学者。对需求学者 n 寻求可能的合作者 k ，以需求学者 n 涉猎的所有主题为距离计算的范围，对在这些主题中有过成果产出的备选学者 k ，需求学者 n 与备选学者 k 在该范围内任意主题 t 中的距离的计算方法如式(6)所示：

$$d_{t,n,k} = \begin{cases} d_{t,n,max}, & \vec{r}_{c,k,t} = \vec{0} \\ |\vec{r}_{c,n,t} - \vec{r}_{c,k,t}|, & \vec{r}_{c,k,t} \neq \vec{0} \end{cases} \quad (6)$$

其中， $d_{t,n,k}$ 是在主题 t 中需求学者 n 与备选学者 k 的距离， $\vec{r}_{c,n,t}$ 为需求学者 n 在主题 t 中的学者

主题中心向量， $\vec{r}_{c,k,t}$ 为备选学者 k 在主题 t 中的学者主题中心向量。

当备选学者 k 在主题 t 中无产出时， $\vec{r}_{c,k,t}$ 为 $\vec{0}$ ， $d_{t,n,k}$ 等于主题 t 中需求学者 n 的最大主题学者距离 $d_{t,n,max}$ 。基于此，需求学者 n 与备选学者 k 在向量空间中的总距离计算如式(7)所示：

$$distance_{n,k} = \frac{\sum_{i=1}^h d_{i,n,k}}{h} \quad (7)$$

其中， h 是需求学者 n 所涉猎的主题数目， $d_{i,n,k}$ 是在需求学者 n 有产出的主题 i 中需求学者 n 与备选学者 k 的距离，可由式(6)获得， $distance_{n,k}$ 为需求学者 n 与备选学者 k 在向量空间中的总距离。

2.2.2 主题指定型合作者推荐指标

主题指定的需求情境是学者具有明确研究目标与研究描述，并基于此对合作学者的擅长研究主题有指定要求。该情境下，主题指定型合作者推荐指标将筛选出在研究描述相关主题中有产出且与相关主题距离总和最小的学者。利用已训练好的BERTopic模型，估算研究描述的主题文档概率 $\vec{p} = (p_1, p_2, \dots, p_i, \dots, p_h)$ (h 为研究描述所涉及的主题数)，以在研究描述所涉及主题中有产出的学者为备选学者。备选学者 k 与研究描述所涉及的任意主题 t 的距离计算方法如式(8)所示：

$$d_{t,k} = \begin{cases} d_{t,max}, & \vec{r}_{c,k,t} = \vec{0} \\ |\vec{r}_{c,t} - \vec{r}_{c,k,t}|, & \vec{r}_{c,k,t} \neq \vec{0} \end{cases} \quad (8)$$

其中， $d_{t,k}$ 为学者 k 与主题 t 的距离， $\vec{r}_{c,k,t}$ 为备选学者 k 在主题 t 中的学者主题中心向量， $\vec{r}_{c,t}$ 为主题 t 的主题中心向量。 $\vec{r}_{c,k,t}$ 为 $\vec{0}$ 时，备选学者 k 在主题 t 中无产出， $d_{t,k}$ 等于主题 t 的最大主题距离 $d_{t,max}$ 。

综上所述，备选学者 k 与研究描述的总距离计算方式如式(9)所示：

$$distance_k = \sum_{i=1}^h (p_i * d_{i,k}) \quad (9)$$

$distance_k$ 为主题指定情况下，备选学者 k 与研究描述的距离。其中， h 是研究描述所涉及的主题数， p_i 是研究描述中主题 i 分布的概率， $d_{i,k}$ 为学者与主题 i 的距离，由式(8)计算， i 仅与主题文档概率 \vec{p} 中出现的主题数量相关。

2.2.3 主题互补型合作者推荐指标

主题互补的需求情境是学者具有明确研究目标

与研究描述,并要求合作学者与己方在研究主题中擅长主题为互补关系。该情境下,主题互补型合作者推荐指标将两个学者视为已达成合作的团队,筛选使得团队到主题距离最短的学者。同理获取主题文档概率 $\vec{p}=(p_1, p_2 \cdots p_i \cdots p_h)$ (h 为研究描述所涉及的主题数)与备选学者列表。将需求学者 n 与任意

$$\vec{r}_{c,n\&k,t} = \begin{cases} \left(\sum_{i=1}^n \vec{r}_{i,n,t} + \sum_{j=1}^k \vec{r}_{j,k,t} \right) / (n+k), & n \neq 0 \mid k \neq 0 \\ \vec{0}, & n = k = 0 \end{cases} \quad (10)$$

其中, $\vec{r}_{c,n\&k,t}$ 为学者 n 与学者 k 在主题 t 中的位置, $\vec{r}_{i,n,t}$ 为需求学者 n 在主题 t 中第 i 份文档的文档向量,同理 $\vec{r}_{j,k,t}$ 为备选学者 k 在主题 t 中第 j 份文档的文档向量,公式中 n 为需求学者 n 分布在主题 t 中的文档总数, k 为备选学者 k 分布在主题 t 中的文档总数。当 n 或 k 为0时,相应文档向量为零向量 $\vec{0}$, n 与 k 同时为0时, $\vec{r}_{c,n\&k,t}$ 等于 $\vec{0}$ 。

基于此,合作团队与主题 t 距离计算方式如式(11)所示:

$$d_{t,n\&k} = \begin{cases} d_{t,max}, & \vec{r}_{c,n\&k,t} = \vec{0} \\ |\vec{r}_{c,t} - \vec{r}_{c,n\&k,t}|, & \vec{r}_{c,n\&k,t} \neq \vec{0} \end{cases} \quad (11)$$

其中, $d_{t,n\&k}$ 为两个学者所构成合作团队与主题 t 距离, $\vec{r}_{c,t}$ 为主题 t 的主题中心向量, $\vec{r}_{c,n\&k,t}$ 为需求学者 n 与备选学者 k 在主题 t 中的合作学者主题中心向量。

合作团队与研究目标的距离计算方式如式(12)所示:

$$distance_{n\&k} = \sum_{i=1}^h (p_i * d_{i,n\&k}) \quad (12)$$

其中, h 是研究描述所涉及的主题数, p_i 是研究描述中主题 i 的分布概率, $d_{i,n\&k}$ 是合作团队与主题 i 的距离,由式(11)计算, $distance_{n\&k}$ 是需求学者 k 与备选学者 n 之间在假定合作的情况下,合作团队与研究目标的距离。

2.3 基于关系网络结构的推荐指标设计

学者常为提高研究影响力寻求团队合作,既有想要维持原有合作关系目标,也有寻求权威合作者以提高个人声誉的目标^[37]。本文以学者在合作关系网络中的位置结构,衡量学者在领域中的重要性与权威,基于学者合作关系网络结构,为学者设计综合考虑学者已有合作关系与备选学者权威性的权威合作者推荐指标。传统合作关系网络图将单次多

备选学者 k 在同一主题中的文档视作一个团队整体,计算合作学者主题中心向量,表征向量空间中二者所构成的合作团队在主题中的位置。在研究描述涉及的任意主题 t 中,合作学者主题中心向量的计算方法如式(10)所示:

人合作记为多次一对一合作的处理方式,造成学者多人合作信息缺失以及学者节点度增大等问题,使得图结构与实际情况存在差别。超图能够通过可同时连接两个及以上节点的超边来体现多位学者合作的情况,更契合现实实际。故本文利用超图构建学者合作关系网络超图,并基于其设计推荐指标。

2.3.1 结构维度社会资本强度

结构维度社会资本是个体在社会交往中与他人建立联系与获得资源的能力,资源的数量和质量决定其高低^[38]。在基于学术关系网络的社会资本中,结构维度社会资本对学者研究成果的正向影响、间接影响也得到大多数学者的认可^[39-41]。通过这一指标筛选出的学者能够更好地满足学者寻求权威合作者、提高研究影响力的合作动机。因此,本文将其作为基于关系网络结构的合作者推荐指标之一。

以往研究常以学者合著关系网络中的中心性衡量学者结构维度社会资本^[42-44]。本文则选用Katz中心性作为结构维度的衡量指标。Katz中心性是特征向量中心的重要变体,特征向量中心性量化节点在网络中的影响力与重要性,认为节点的中心性是邻居节点中心性的函数,邻居节点越多、中心性越大,节点自身中心性越大^[45],Katz中心性在特征向量中心性的基础之上引入常数项,以避免特征向量中心性为零的情况,应用性更强。许多学者将中心性算法扩展到超图结构,构建了超图节点中心性的算法。节点 v_i 的Katz中心性定义如式(13)所示:

$$C_{katz}(v_i) = \alpha \sum_{j=1}^n A_{ij} C_{katz}(v_j) + \beta \quad (13)$$

其中, A_{ij} 为学者关系网络超图的邻接矩阵, C_{katz} 为表示所有节点Katz中心性的向量, n 为超图的节点总数, β 为常数项。

2.3.2 学者合作可能

基于合著关系网络的结构维度社会资本不受模

型、学者个人影响，在具体领域中具体时间点可被视为定值。然而，能否与高社会资本的备选学者达成合作不由备选学者的社会资本强度决定，出于提高学者合作可行性与成功率的考虑，本文试图为学者筛选关系可达范围内最强社会资本合作者，同时帮助学者审视原有合作关系。

聚类系数是一般图中重要的网络度量，有全局与局部之分，全局的聚类系数用于衡量节点邻域的聚合紧密程度或其可能达到的聚合紧密程度，一个节点的局部聚类系数体现其邻居节点也相通的可能性^[46]。本文试图利用聚类系数来衡量关系网络中任意两个学者之间合作的可能性，认为当两个学者间有路径相连并且路径上的学者与其邻居所构成的团体关系越紧密，两个学者的聚合紧密程度更大的可能性更高，彼此合作的可能性也就更大。学者 Gallagher S R 等^[47]在超图研究中提出新的聚类系数定义，借鉴其算法与定义，节点 u 的局部聚类系数计算方法如式 (14) 所示：

$$cc_u = \frac{2 \sum_{v,w \in N_u} I_E(v,w, \neg u)}{|N_u|(|N_u|-1)} \quad (14)$$

其中 N_u 为节点 u 的邻居集合，其中 E 为超图中所有连接集合 N_u 中节点的超边集合， $v, w \in N_u$ ，如果存在超边 $e_i \in E$ 使得 $v, w \in e_i$ 但是 $u \notin e_i$ ，则 $I_E(v,w, \neg u) = 1$ ，否则 $I_E(v,w, \neg u) = 0$ ， cc_u 为节点 u 的聚类系数，表示与节点 u 相连的任意两邻居合作的紧密程度。

对关系网络中彼此间最短路径较大的学者，基于六度分离理论，将最短路径长度超过 6 或没有路径相连的学者间合作可能性定为网络中全局聚类系数。对关系网络图 G ，其全局聚类系数计算方法如式 (15) 所示：

$$gcc = \frac{\sum_{v \in G} r_v}{\sum_{v \in G} \frac{d_v(d_v-1)}{2}} \quad (15)$$

其中， r_v 为节点 v 的邻居节点间的关系数， $\sum_{v \in G} r_v$ 为图中由点与边一起构成的封闭三角形的数量， d_v 为节点 v 的邻居节点数， $\sum_{v \in G} \frac{d_v(d_v-1)}{2}$ 为图中所有的边与点构成的三角形的数量，包括开放三角形与封闭三角形， gcc 为图 G 的全局聚类系数。

对于最短路径小于 6 且大于 1 的任意两学者，其关系可达程度由最短路径上学者间的局部聚类系数决定。具体情况为，对任意最短路径小于等于 6 的学者 s 与学者 t ，其合作可能计算方式如式 (16) 所示：

$$p = \frac{\sum_{u \in st} cc_u}{l_{st}-1} \quad (16)$$

其中， l_{st} 为学者 s 与学者 t 之间最短路径的长度， u 为该路径中的点 ($u \neq s, u \neq t$)，当两个学者间不止一条最短路径时，选用 p 最大的最短路径。

对于任意两学者，其合作可能计算方式如式 (17) 所示：

$$Pro = \begin{cases} gcc, & l_{st} = 0 \vee l_{st} > 6 \\ p, & 1 < l_{st} \leq 6 \\ 1, & l_{st} = 1 \end{cases} \quad (17)$$

若两个学者间存在直接相连的边 ($l_{st} = 1$)，则其合作可能视为 1。

2.4 融合研究主题与关系网络结构的推荐排名设计

本文的目标是为已有明确研究目标的科研工作者提供高自由度、强个性化的合作推荐服务，借鉴投票得分的方法，对基于研究主题内容与基于学者关系网络结构两部分推荐指标，本文采取以学者提供权重，二者按权重线性结合的方式。

首先，获得学者研究主题推荐得分。基于学者研究主题的推荐指标以向量间的距离筛选合作者，所设计的 3 个推荐指标间为并列关系。然而距离越小，推荐程度越高，与一般直觉认知相反，需对距离指标进行归一化，形成基于学者研究主题的推荐得分。具体计算方式如式 (18) 所示：

$$Topic = \begin{cases} \frac{\max(\text{distance}_{n,k}) - \text{distance}_{n,k}}{\max(\text{distance}_{n,k}) - \min(\text{distance}_{n,k})}, & \text{相似型指标} \\ \frac{\max(\text{distance}_k) - \text{distance}_k}{\max(\text{distance}_k) - \min(\text{distance}_k)}, & \text{指定型指标} \\ \frac{\max(\text{distance}_{n\&k}) - \text{distance}_{n\&k}}{\max(\text{distance}_{n\&k}) - \min(\text{distance}_{n\&k})}, & \text{互补型指标} \end{cases} \quad (18)$$

其次，获得学者关系网络结构推荐得分。基于关系网络结构的推荐指标目标是为学者筛选关系可达范围内最强社会资本合作者，以满足学者对合作者专业权威的需求。故需综合备选学者的结构维度社会资本强度 x_v 、需求学者与备选学者的合作可能 Pro 两项指标。本文以二者乘积计算学者关系网络结构推荐得分，在获得推荐得分之前需对二者进行归一化处理，计算方式如式 (19) 所示：

$$Stru = \frac{Pro - \min(Pro)}{\max(Pro) - \min(Pro)} * \frac{x_v - \min(x_v)}{\max(x_v) - \min(x_v)} \quad (19)$$

最后，获得学者综合推荐得分。前两部分内容将按照学者所提供权重线性结合，得到备选学者的最终综合推荐得分。学者提供所需合作者的研究主题方面权重 α 值，结合已获得的学者研究主题推荐得分 $Topic$ 、备选学者关系网络推荐得分 $Stru$ ，学者综合推荐得分计算方法如式 (20) 所示：

$$Score = \alpha * Topic + (1 - \alpha) * Stru \quad (20)$$

基于本文的推荐方案，学者可以通过自主选指标、定权重的方式个性化推荐过程。方案同时提供备选学者的综合推荐结果与 $Topic$ 、 $Stru$ 两部分的得分 ($Score_{\alpha=1}$ 、 $Score_{\alpha=0}$)，跨领域学者、领域新人以及权威学者可以从不同角度进行解读。例如，跨学科学者选择“主题指定型”指标并以综合推荐得分 $Score_{\alpha=1}$ 形成推荐列表，获得相关主题的专家型学者，解决进入新领域遇到的冷启动问题。已有稳定合作关系的权威学者可以横向、纵向比较主题推荐指标以判断是否需要以及谁组成团队。

3 实验与结果分析

本文基于国内情报学领域对科研合作者推荐方案进行实证，由于方案给出的综合推荐得分依赖需求学者对方案进行个性化处理；同时，方案试图展示学者全部推荐指标得分为学者提供决策支持，无法获得推荐成功与否、推荐方案被学者采纳程度的直接结果。因而，本文选择以案例实验的方式展示推荐过程与结果分析，以展示推荐方案可被利用的角度与方案的可行性。

3.1 数据来源及预处理

本文选择中国知网作为国内情报学领域期刊文

献的来源数据库。首先，检索中国知网中 2007—2023 年刊登在中文社会科学引文索引 (CSSCI) 来源期刊上且中图分类号为 G35 (情报学、情报工作) 的期刊文献，并导出其题录数据，共获得 17 030 条题录数据，形成“总文献数据集合”，检索时间截至 2023 年 4 月 25 日。其次，为所有文献与学者给定特定编号，并以隶属单位区分重名学者，得学者 17 594 名，形成“总学者记录集合”。学者的隶属单位将由题录中单位字段与学者字段获得，无法从字段记录中获得明确隶属单位的学者，则手动检索、补充。进而，筛选成果数量大于 3 的学者并依据文献题录作者字段记录符合条件的学者间的合作关系，用以构建学者合作关系网络。得符合条件的学者 1 943 名，其总产出文献 10 727 条，彼此间合作产出文献 4 706 条。筛选的目的在于减少向科研工作者们推荐硕士生、未毕业博士生以及偶然跨学科学者的情况，同时减轻基于关系网络结构的推荐指标运算负担。最后，将“总文献数据集合”中所有文献题录的标题字段与摘要字段作为研究文档，并对其分词、去标点符号以及去停用词等预处理。上述过程所得数据集与数据量如表 1 所示。

表 1 实验数据集的数据量

Tab. 1 The Amount of Data in the Experimental Dataset

数据集名称	数据数量
总文献数据集合	17 030
总学者记录集合	17 594
产出超 3 的学者集合	1 943
产出超 3 的学者产出文献集合	10 727
产出超 3 的学者间合作记录集合	4 706

3.2 情报学领域主题模型框架建立

本次实验选择来自国内哈工大与科大讯飞研究联合实验室的 Bert 中文预训练模型“chinese-RoBERTa-wwm-ext”，该预训练模型继承原 RoBERTa 预训练模型鲁棒性强的优势，同时能够更好地适应中文语境^[48]。在 Python 中将已处理好的研究文档嵌入 Bert 中文预训练模型“chinese-RoBERTa-wwm-ext”，并对所得文档向量进行降维与聚类，调节两环节参数，解决初始聚类结果粒度过细、情报学信息蕴含不清晰等问题，最终获得主题 94 个。模型

对情报学有关主题的挖掘较为准确,非情报学相关主题较少,94个主题覆盖情报学理论与方法、情报学教育、竞争情报、技术创新与识别、学科知识流动、政治政务公开与治理、国家安全、图书馆领域相关、文献计量与替代计量学、学者或期刊评价、舆情评论监控、智库相关、知识抽取、本体语言、信息检索与检索系统等多个方面,每个方面由多个主题进一步细分,主题粒度适中,不同主题间区分明确。

为此,针对当前模型中主题内容与分布情况,计算每个主题的主题中心向量 $\vec{r}_{c,t}$ 以及每位学者所有主题的中心向量 $\vec{r}_{c,s,t}$ 。

3.3 学者合作关系网络构建

选择“产出超3的学者间合作记录集合”作为学者关系网络超图的数据来源,以学者为节点、学者间合作记录为边,构建学者合作关系网络超图,得学者节点1766个,超边4706条。产出超3的学者有1943名,然而有177名学者的所有合作对象符合要求的产出都不足3,故其不在“产出超3的学者间合作记录集合”中,将其以孤立点加入学者合作关系网络超图。由于学者节点过多,仅展示学者邓三鸿与其他学者的部分合作记录,如图2所示。

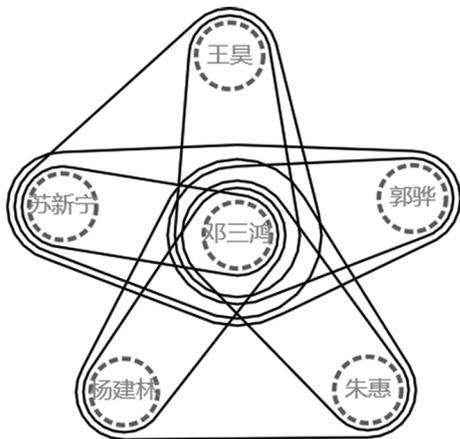


图2 学者邓三鸿的部分合作记录

Fig. 2 Part of Deng S H's Record of Collaboration

依据基于关系网络结构的推荐指标,计算关系网络超图中学者的结构维度社会资本强度。

3.4 案例分析与推荐结果

选择《科技情报研究》于2023年10月刊登的文献“多源异构数据视角下的学术评价:内涵、进展与展望”为实证案例^[49],其由学者邓三鸿、杨

杰、王昊以及沈健威共同产出,本文以该文献第一作者邓三鸿为需求学者,学者王昊为备选学者,由于杨杰为在读博士、沈健威为专业技术人员,不在备选学者数据库内,不将其作为备选学者。在本次实验案例中,将围绕两部分得分进行推荐结果分析,在实际应用中,应由寻求合作者的学者提供研究主题与关系结构两部分推荐得分相结合的权重 α 值。

首先,获得所有备选学者在研究主题方面推荐指标。将所选用文献的标题与摘要作为学者邓三鸿提出的研究描述,具体描述如下:

“多源异构数据视角下的学术评价:内涵、进展与展望。综合多源异构数据的学术评价与科学计量方法逐渐发展成为领域热点,目前对于多源异构数据视角下的学术评价缺乏系统的理论探讨。针对多源异构数据利用不充分、研究方法不妥当、研究目标不明确等问题,文章梳理多源异构数据视角下的学术评价发展状况、厘清领域内涵及外延、分析领域前沿和热点。研究发现,在多源异构数据视角下,学术评价领域呈现出数据维度增加、数据粒度细化、评价指标复杂化等趋势。多源异构数据视角下的学术评价内涵界定、现状分析与未来展望,对于澄清领域认知、辨析相关概念、明确研究路径与方法论基础,以及构建中国特色的学术评价体系有着重大意义”。

基于已构建好的BERTopic模型,获得上述研究描述所涉及的主题类别及其可能概率,如图3所示。

按主题类别从小到大的顺序,将对应的可能概率排序并形成向量,即为研究描述主题文档概率 $\vec{p}=(0.01509406, 0.01513423, 0.0193932, 0.01968422, 0.01786734, 0.03460608, 0.01712272, 0.0165137, 0.02054663, 0.0176196, 0.01595388, 0.01742257, 0.01765436, 0.02613466, 0.01946489, 0.01602702, 0.01887774)$ 。按照基于研究主题的推荐指标计算方案,计算3个指标下前15名距离最短的备选学者,学者列表如表2所示。

学者王昊在主题相似型指标中排名较前,同时在主题指定型、主题互补型指标中距离为1.057787、1.048493,分别排在第28、32名。在主题指定的要

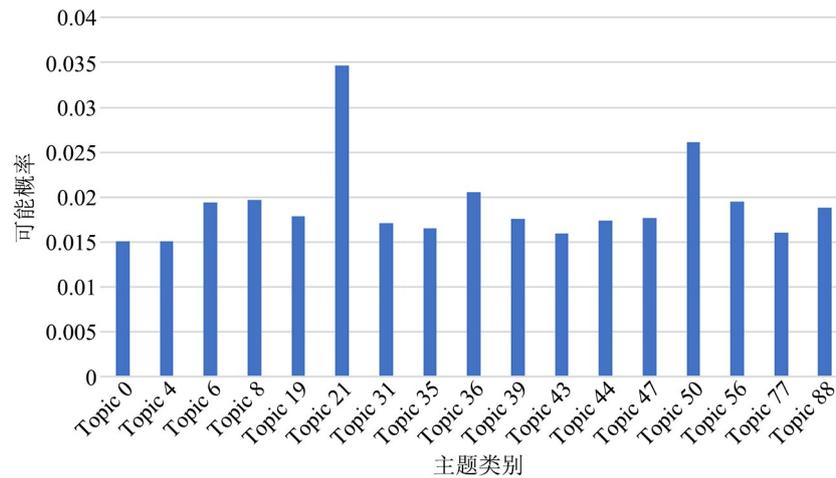


图3 研究描述的主题类别与可能概率

Fig. 3 The Topic Categories and Probability of the Description of the Study

表2 前15名距离最短的学者

Tab. 2 Top 15 Scholars with the Shortest Distance

主题指定型		主题互补型		主题相似型	
邱均平	0.608052543	邱均平	0.563794142	邱均平	2.213863535
马海群	0.687248141	马海群	0.660913500	杨建林	2.713306821
张志强	0.855339886	苏新宁	0.796374331	苏新宁	2.756499512
苏新宁	0.856922118	化柏林	0.836864179	赵蓉英	2.865683963
杨国立	0.868033184	张志强	0.849889477	王昊	3.066955732
化柏林	0.889747914	杨国立	0.849917421	李纲	3.153193471
赵蓉英	0.892540064	赵蓉英	0.864907588	许鑫	3.167215980
杨思洛	0.906203690	杨思洛	0.879693904	武夷山	3.196781168
杨建林	0.948176199	孙建军	0.891554219	杨思洛	3.372362176
李品	0.955670023	李品	0.923641389	朱庆华	3.413235256
李静	0.963567174	李纲	0.931027862	孙建军	3.455217721
孙建军	0.975644464	杨建林	0.949877626	华薇娜	3.484010092
张静	0.976427345	武夷山	0.953146716	马海群	3.553851118
王伟	0.981009595	李静	0.954369038	白云	3.613383794
刘雪立	0.981302643	姜春林	0.961594533	张洋	3.634932468

求下，按主题指定型推荐指标的计算方法，学者邓三鸿与研究描述的距离为 0.798533，说明在案例研究所涉及的主题中，学者邓三鸿属于权威学者，不选择在主题指定型指标、主题互补型指标中排名靠前的学者是可以被接受的。

发现学者邱均平在主题指定的要求下，与研究描述的距离为 0.608052543，其与学者邓三鸿在假设合作的情况下，团队与研究描述的距离为 0.563794142，可知当学者邓三鸿与邱均平达成合作后，团队与研

究描述的距离都有较大的缩短，同理还有学者马海群、苏新宁。因此，从主题擅长、知识储备的角度来看，学者可以将邱均平作为可考虑的合作对象。

在关系网络结构方面，基于学者合作关系网络超图，计算学者邓三鸿与备选学者的合作可能与备选学者的，前 15 名推荐学者结果列表如表 3 所示。

在关系网络方面，学者王昊的推荐得分较高，位于第 4 名。结合学者王昊与学者邓三鸿曾有多次合作的情况，推测学者王昊与邓三鸿研究领域、

表3 前15名关系网络推荐得分最高的学者
Tab. 3 Top 15 Scholars with the Highest Score in Relationship Network Recommendation

学者	结构维度 社会资本 (x_i)	合作可能 (Pro)	关系网络 推荐得分 (Stru)
邱均平	0.116081540	0.377777778	43.8530263
苏新宁	0.043035108	1.000000000	43.03510759
王东波	0.034541336	1.000000000	34.54133635
王昊	0.033975085	1.000000000	33.97508494
杨建林	0.028878822	1.000000000	28.8788222
刘雪立	0.104190260	0.272834967	28.42674631
赵蓉英	0.086636467	0.324074074	28.07663269
吴晨生	0.083238958	0.293869732	24.46141029
方曙	0.068516421	0.323763955	22.18314756
李长玲	0.073046433	0.273308271	19.96419418
王伟	0.018686297	1.000000000	18.68629672
冷伏海	0.052095130	0.345454545	17.99649954
谢靖	0.016421291	1.000000000	16.42129105
许鑫	0.015855040	1.000000000	15.85503964
许海云	0.080973952	0.172152560	13.93987321

主题重合较高，关系较为密切，彼此间已形成稳定的合作关系。学者可从以下角度利用本方案所提供信息：从主题角度出发判断原有合作者与研究目标是否契合、是否有更契合的新合作者、新合作者对团队的增益是否值得打破原有合作关系；再从关系网络结构的角度，判断与该学者合作的可能与否。例如，学者邱均平、马海群、苏新宁在主题上都能够使团队更接近研究目标，在关系与权威方面，学者邱均平与苏新宁的优势大于其他学者。从主题与关系网络两方面综合来看，学者可以将邱均平、苏新宁视作可尝试的合作对象。

4 总结与展望

与以往合作者推荐方法向学者推荐潜在合作者、可能合作者或过往合作中最具价值的合作者的情况不同^[50]，本文所构建的科研合作者推荐方案以学者合作动机为切入点，支持学者为具体研究主题与需求寻找合作者，同时推荐结果中既有曾经的合作者，也有新的推荐合作者。学者既可以了解新的、尚未合作过的合作者，拓展合作范围，又可以横向

对比曾经的合作者，以决定是否仍需维持原有合作。

在研究主题方面，本文应用最新主题模型 BERTopic 识别领域内的学者的研究主题，该方法对文本语义理解程度更加深入，对主题信息表示更加准确。在关系网络结构方面，基于充分反映学者实际合作情况的超图结构构建学者合作关系网络超图，并从学术关系网络的结构维度社会资本视角对其进行挖掘。方案创新性地将前两部分内容以学者按需提供的权重线性结合，使得方案能够为学者提供高度个性化的推荐服务，同时过程清晰且可解释性较强。实验结果表明，BERTopic 模型能够较为准确地识别领域内情报学相关主题，同时关系网络部分准确挖掘出学者关系网络可达范围内社会资本强度更高的备选合作者，证实本文所构建的模型具有一定的可实践性。

从跨领域学者到权威学者，本文所提推荐方案都可以提供个性化推荐服务。通过自主选指标、定权重等环节个性化推荐方案框架的设计，过程简洁、可操作性强，一定程度上实现了为已有明确研究目标的科研工作者提供高自由度的推荐服务的目标，为科研合作者推荐方法的个性化探索提供可能的解决思路。与此同时，本文尚存在一定不足。本文仅针对情报学领域展开实践，事实上情报学领域与图书馆学领域的期刊重合度较高，分类号 G25 与 G35 在内容上有一定重合，并且部分学者更倾向于发表外刊，仅基于中国知网中的中文社会科学引文索引 (CSSCI) 来源期刊使得实验的数据来源不够全面，这可能对学者结构维度社会资本的测量准确性造成一定影响。但是，对能够获取全来源数据的数据服务商而言，可对本文所提推荐方案进行一定尝试。

参 考 文 献

- [1] 张雪, 张志强, 陈秀娟. 基于期刊论文的作者合作特征及其对科研产出的影响——以国际医学信息学领域高产作者为例 [J]. 情报学报, 2019, 38 (1): 29-37.
- [2] 宋艳辉, 魏新星, 邱均平. 融合 Altmetrics 指标的领域高产学者综合影响力研究——以 Nature 生物科技子刊为例 [J]. 现代情报, 2024, 44 (1): 153-167.
- [3] 吴小兰, 章成志. 国家自然科学基金视角下学科跨学科性演变研究 [J]. 科技情报研究, 2022, 4 (3): 20-32.
- [4] 杨瑞仙, 李贤. 科学合作与论文影响力之间的相关性研究 [J].

- 现代情报, 2019, 39 (4): 125-133.
- [5] 邱均平, 温芬芳. 作者合作程度与科研产出的相关性分析——基于“图书情报档案学”高产作者的计量分析 [J]. 科技进步与对策, 2011, 28 (5): 1-5.
- [6] 胡译文, 崔静静, 任萍. 国内科研合作研究进展综述 [J]. 现代情报, 2022, 42 (6): 148-159.
- [7] 刘云, 朱东华, 许海力, 等. 基础学科国际科学合作的重要模式 [J]. 科学学研究, 1996, 27 (1): 37-42.
- [8] 谢彩霞. 科学合作动力因素的科学计量学分析 [J]. 情报杂志, 2009, 28 (2): 167-171.
- [9] Beaver D D. Reflections on Scientific Collaboration (and its Study): Past, Present, and Future [J]. *Scientometrics*, 2001, 52: 365-377.
- [10] Melin G. Pragmatism and Self-organization: Research Collaboration on the Individual Level [J]. *Research Policy*, 2000, 29 (1): 31-40.
- [11] Birnholtz J P. When Do Researchers Collaborate? Toward a Model of Collaboration Propensity [J]. *Journal of the Association for Information Science and Technology*, 2007, 58 (14): 2226-2239.
- [12] Garcia A, López-López W, Acevedo-Triana C A, et al. Cooperation in the Latin American Behavioral Sciences: Motivation, Evaluation and Difficulties [J]. *Suma Psicológica*, 2016, 23 (2): 125-132.
- [13] 马凤, 武夷山. 中国科技期刊研究界科研合作动机及相关问题研究 [J]. 科技管理研究, 2009, 29 (8): 572-575.
- [14] 赵君, 廖建桥. 科研合作研究综述 [J]. 科学管理研究, 2013, 31 (2): 117-120.
- [15] Georghiou L. Global Cooperation in Research [J]. *Research Policy*, 1998, 27 (6): 611-626.
- [16] Dusdal J, Powell J J W. Benefits, Motivations, and Challenges of International Collaborative Research: a Sociology of Science Case Study [J]. *Science and Public Policy*, 2021, 48 (2): 235-245.
- [17] Maglaughlin K L, Sonnenwald D. Factors that Impact Interdisciplinary Natural Science Research Collaboration in Academia [C] // *Proceedings of the ISSI*, 2005: 24-25.
- [18] Yang C, Ma J, Sun J S, et al. A Weighted Topic Model Enhanced Approach for Complementary Collaborator Recommendation [C] // *18th Pacific Asia Conference on Information Systems, Association for Information Systems*, 2014.
- [19] 刘萍, 郑凯伦, 邹德安. 基于 LDA 模型的科研合作推荐研究 [J]. 情报理论与实践, 2015, 38 (9): 79-85.
- [20] 熊回香, 李跃艳. 基于 Word2vec 的学者推荐与跨语言论文推荐模型研究 [J]. 情报科学, 2019, 37 (12): 19-26.
- [21] 杨梦婷, 熊回香, 肖兵, 等. 基于动态特征的学者推荐研究 [J]. 情报理论与实践, 2022, 45 (4): 120-127.
- [22] Kong X J, Jiang H Z, Wang W, et al. Exploring Dynamic Research Interest and Academic Influence for Scientific Collaborator Recommendation [J]. *Scientometrics*, 2017, 113 (1): 369-385.
- [23] Bukowski M, Geisler S, Schmitz-Rode T, et al. Feasibility of Activity-Based Expert Profiling Using Text Mining of Scientific Publications and Patents [J]. *Scientometrics*, 2020, 123 (2): 579-620.
- [24] Wang W, Xu B, Liu J Y, et al. CSTeller: Forecasting Scientific Collaboration Sustainability Based on Extreme Gradient Boosting [J]. *World Wide Web*, 2019, 22 (6): 2749-2770.
- [25] Lorrain F, White H C. Structural Equivalence of Individuals in Social Networks [J]. *Social Networks*, 1971, 1 (1): 67-98.
- [26] Zhou X, Ding L X, Li Z K, et al. Collaborator Recommendation in Heterogeneous Bibliographic Networks Using Random Walks [J]. *Information Retrieval Journal*, 2017, 20 (4): 317-337.
- [27] Chen Y, Liu T T, Chen X H, et al. HNRWalker: Recommending Academic Collaborators with Dynamic Transition Probabilities in Heterogeneous Networks [J]. *Scientometrics*, 2020, 123 (1): 429-449.
- [28] Clauset A, Moore C, Newman M E. Hierarchical Structure and the Prediction of Missing Links in Networks [J]. *Nature*, 2008, 453: 98-101.
- [29] Chi K, Yin G S, Dong Y X, et al. Link Prediction in Dynamic Networks Based on the Attraction Force Between Nodes [J]. *Knowledge-Based Systems*, 2019, 183: 105073.
- [30] Chen J, Wang X, Zhao S, et al. Content-Enhanced Network Embedding for Academic Collaborator Recommendation [J]. *Complexity*, 2021: 1-12.
- [31] Liu X Y, Wu K, Liu B, et al. HNERec: Scientific Collaborator Recommendation Model Based on Heterogeneous Network Embedding [J]. *Information Processing & Management*, 2023, 60 (2): 103253.
- [32] Xi X W, Wei J Q, Guo Y, et al. Academic Collaborations: a Recommender Framework Spanning Research Interests and Network Topology [J]. *Scientometrics*, 2022, 127 (11): 6787-6808.
- [33] Li P, Li T C, Wang X, et al. Scholar Recommendation Based on High-Order Propagation of Knowledge Graphs [J]. *International Journal on Semantic Web & Information Systems*, 2022, 18 (1): 1-19.
- [34] 董文慧, 熊回香, 杜瑾, 等. 基于学者画像的科研合作者推荐研究 [J]. 数据分析和知识发现, 2022, 6 (10): 20-34.
- [35] Grootendorst M. BERTopic: Neural Topic Modeling with a Class-based TF-IDF Procedure [J]. *ArXiv Preprint ArXiv*: 2203. 05794.
- [36] Angelov D. Top2vec: Distributed Representations of Topics [J]. *ArXiv Preprint ArXiv*: 2008. 09470.

[37] Lee S, Bozeman B. The Impact of Research Collaboration on Scientific Productivity [J]. *Social Studies of Science*, 2005, 35 (5): 673-702.

[38] 曾粤亮. 跨学科科研合作行为影响因素研究 [D]. 武汉: 武汉大学, 2019.

[39] Liao C H. How to Improve Research Quality? Examining the Impacts of Collaboration Intensity and Member Diversity in Collaboration Networks [J]. *Scientometrics*, 2011, 86 (3): 747-761.

[40] 宋志红, 武天兰, 李冬梅. 合著网络、社会资本与科研影响力 [J]. *情报学报*, 2015, 34 (11): 1123-1131.

[41] 蔡小静. 科学基金的资助效益研究 [D]. 杭州: 浙江大学, 2021.

[42] Liu C H. Network Position and Cooperation Partners Selection Strategies for Research Productivity [J]. *Management Decision*, 2015, 53 (3): 494-511.

[43] 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.

[44] 解峰, 陈轩瑾, 窦天芳. 学术合作网络与研究影响——基于我国高校科学家的微观证据 [J]. *科学学研究*, 2021, 39 (5): 777-785.

[45] Newman M E J. Modularity and Community Structure in Networks [J]. *Proc Natl Acad Sci USA*, 2006, 103 (23): 8577-8582.

[46] Battiston F, Cencetti G, Iacopini I, et al. Networks Beyond Pairwise Interactions: Structure and Dynamics [J]. *Physics Reports*, 2020, 874: 1-92.

[47] Gallagher S R, Goldberg D S. Clustering Coefficients in Protein Interaction Hypernetworks [C] //BCB'13: Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics, 2013: 552-560.

[48] Cui Y M, Che W X, Liu T, et al. Pre-Training with Whole Word Masking for Chinese BERT [J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2021, 29: 3504-3514.

[49] 邓三鸿, 杨杰, 王昊, 等. 多源异构数据视角下的学术评价: 内涵、进展与展望 [J]. *科技情报研究*, 2023, 5 (4): 42-56.

[50] Kong X, Jiang H, Yang Z, et al. Exploiting Publication Contents and Collaboration Networks for Collaborator Recommendation [J]. *PLoS One*, 2016, 11 (2): 0148492.

(责任编辑: 郭沫含)

(上接第 27 页)

[42] 苏竣. 公共科技政策导论 [M]. 北京: 科学出版社, 2014.

[43] 中国信通院. 数据要素流通视角下数据安全保障研究报告 (2022 年) [EB/OL]. http://www.caict.ac.cn/kxyj/qwfb/ztbg/202301/t20230104_413655.htm, 2024-04-19.

[44] 孙瑞英, 陈宜泓. 基于 PMC 指数模型的我国公共数据开放政策评价研究 [J]. *情报理论与实践*, 2023, 46 (8): 33-42.

[45] 张永安, 郑海拓. “大众创业、万众创新”政策量化评价研究——以 2017 的 10 项双创政策情报为例 [J]. *情报杂志*, 2018, 37 (3): 158-164, 186.

[46] 熊焯. 政策工具视角下的医疗卫生体制改革: 回顾与前瞻——基于 1978—2015 年医疗卫生政策的文本分析 [J]. *社会保障研究*, 2016, (3): 51-60.

[47] 周海炜, 张蔚怡, 王腾, 等. 政策工具视角下我国水安全政策文本分析三维框架 [J]. *情报杂志*, 2022, 41 (9): 62-70.

[48] Howlett M, Ramesh M. Studying Public Policy: Policy Cycles and Policy Subsystems [J]. *American Political Science Association*, 2009, 91 (2): 548-580.

[49] 何林莹, 马海群. 地方公共数据政策量化评价研究——基于 PMC 指数模型 [J]. *现代情报*, 2023, 43 (8): 14-26.

[50] 周海炜, 陈青青. 大数据发展政策的量化评价及优化路径探究——基于 PMC 指数模型 [J]. *管理现代化*, 2020, 40 (4): 74-78.

[51] 卜令通, 张嘉伟. 基于 PMC 指数模型的数字经济政策量化评价 [J]. *统计与决策*, 2023, 39 (7): 22-27.

[52] 梅傲, 陈子文. 政府数据开放中的数据安全隐忧及其纾解 [J]. *情报杂志*, 2023, 42 (5): 76-85.

[53] 复旦大学数字与移动治理实验室. 中国地方政府数据开放报告——省域指数(2022 年度) [R] 上海: 复旦大学数字与移动治理实验室, 2022.

[54] 国家法律法规数据库. 山东省大数据发展促进条例 [EB/OL]. <https://flk.npc.gov.cn/detail2.html?ZmY4MDgxODE3Y2JhZGZhZTAxN2NjYjAyMDYwZTRlNzI%3D>, 2024-04-19.

[55] 国家法律法规数据库. 深圳经济特区数据条例 [EB/OL]. <https://flk.npc.gov.cn/detail2.html?ZmY4MDgxODE3YTlhMzU1NzAxN2FhM2UxMzQ4NTJlYjY%3D>, 2024-04-19.

[56] 国家法律法规数据库. 贵阳市大数据安全管理条例 [EB/OL]. <https://flk.npc.gov.cn/detail2.html?MmM5MGU1YmE2NGZhODIwNjAxNjVjMTc3M2YzZDEzZDU%3D>, 2024-04-19.

[57] 国家互联网信息办公室. 关于《网络数据安全条例(征求意见稿)》公开征求意见的通知 [EB/OL]. http://www.cac.gov.cn/2021-11/14/c_1638501991577898.htm, 2024-04-19.

[58] 朱春奎. 政策网络与政策工具: 理论基础与中国实践 [M]. 上海: 复旦大学出版社, 2011.

(责任编辑: 郭沫含)