

基于非对称视角的作者相似性探测

席崇俊 刘志辉 张均胜

(中国科学技术信息研究所, 北京 100038)

摘要: [目的/意义] 针对传统作者相似性探测中存在的问题, 完善作者相似度算法, 在梳理领域知识结构、挖掘作者合作关系以及科研社区发现等方面具有重要意义。[方法/过程] 从非对称视角, 将作者发文量、关键词使用频率、行文习惯等因素纳入考虑, 提出一种融合作者研究强度和词关联度的作者相似度算法, 先用关键词数据进行实证分析, 然后将其应用于引文数据, 进行计算并对比分析。[结果/结论] 实验结果表明, 本文提出的非对称视角下融入研究强度后的作者相似度算法可以更准确地反映作者之间的相似性, 提高作者推荐精度; 融入词关联度以及基于引文数据的计算可以放大作者间的相似度, 对挖掘作者潜在合作关系具有一定帮助。

关键词: 作者耦合关系 研究强度 词关联度 非对称视角

分类号: G350

DOI: 10.31193/SSAP.J.ISSN.2096-6695.2022.02.03

0 引言

随着科学技术的不断发展, 学术文献的数量迅猛增长, 文献作者的规模也日益庞大, 基于学术论文探测作者相似性一直是图书情报领域的热点研究问题之一^[1], 在研究领域知识结构^[2-4]、挖掘作者合作关系^[5,6]以及科研社区发现^[7]等方面具有深远的意义, 提高作者相似性的识别精度是研究所致力的主要方向。

目前, 关于作者相似性的计算多基于作者之间的耦合关系进行研究, 根据耦合内容可分为基于关键词的作者关键词耦合分析、基于引文的作者文献耦合分析以及基于署名的作者合著分析。然而, 以上方法在进行作者相似度计算时, 未考虑作者使用关键词或引文数据的频次, 且多在一种对称视角下进行研究, 无法真实准确地反映作者的关系。因此, 为了对作者相似性进行更加细致和深入的分析, 本文考虑利用作者使用的关键词和引文数据反映作者研究兴趣, 并以作者使

[作者简介] 席崇俊 (ORCID: 0000-0003-0334-7109), 男, 硕士研究生, 研究方向为技术竞争情报, Email: xicj7465@163.com; 刘志辉 (ORCID: 0000-0002-1850-4325), 男, 博士, 研究员, 研究方向为竞争情报, Email: liuzhihui@istic.ac.cn; 张均胜 (ORCID: 0000-0001-8740-2851), 男, 博士, 研究员, 研究方向为大数据挖掘与情报分析, Email: zhangjs@istic.ac.cn。

用关键词和引文数据的频次反映作者在其研究兴趣上的研究强度,从非对称视角计算作者的相似度,以弥补传统算法下相似度相同的作者无法进行二次排序的不足。

1 相关研究概述

作者耦合指的是两个或两个以上的论文作者通过某种现象而建立起的耦合关系^[8],目前对作者耦合关系的研究路径主要有三种:基于文献引用的“作者文献耦合分析(Author Bibliographic Coupling Analysis, ABCA)”、基于关键词的“作者关键词耦合分析(Author Keyword Coupling Analysis, AKCA)”和基于文献署名的“作者合著分析(Author Co-author Analysis, ACA)”。

作者合著分析是利用作者之间的合著论文数量来表征其相似性,合著数量越多,则研究内容越为相似^[9]。但是,这一研究路径只考虑作者之间是否存在合著关系,难以分辨不同作者在合作关系中的主次程度,而且,作者合著关系中经常夹杂着作者之间的社会关系,不能真实可靠地反映作者之间的学术关系^[10],这些因素都会影响作者相似度辨识的准确性。出于这一考虑,本文将主要基于作者文献耦合分析与作者关键词耦合分析方法进行探讨。

作者文献耦合分析是以作者对参考文献的引用行为为研究对象,如果两个或两个以上的作者引用相同参考文献的数量越多,则他们的研究内容越为相似,其中,他们引用相同参考文献的数量称为“作者文献耦合强度”。目前已有的研究,Zhao Dangzhi等首先提出了作者文献耦合分析并利用其对国内外图书情报学领域的知识结构演化趋势进行了分析^[11];陈远和王菲菲对作者文献耦合分析进行了实证研究,探索了图书情报学领域的研究前沿与研究热点^[12];随后,邱均平在此基础上对作者文献耦合分析与作者共被引分析进行了对比研究^[13];与此同时,李国俊和肖明等构建了作者文献耦合的可视化系统,直观地展示了作者文献耦合的隐含含义^[14]。

既然引文可以表征作者的关联关系,那么作为文献的内部单元——关键词也能将作者关联起来。基于这一思想,刘志辉等首次提出了作者关键词耦合分析:两个或两个以上的作者使用相同关键词的数量越多,则他们的研究越为相似^[15];随后,邱均平分别与王菲菲、陈木配等人合作,利用文献关键词耦合分析挖掘了竞争情报领域和计量学领域的作者潜在合作关系^[16,17];宋艳辉和武夷山以 *Scientometrics* 期刊为例,对比研究了作者关键词耦合分析与作者文献耦合分析的异同^[18];丁敬达将作者关键词耦合与作者文献耦合分析方法相结合,拓宽了这两种方法的应用范围^[19];陈卫静和郑颖综合考虑了关键词频次、作者发文量等因素对作者相似度的影响,对作者关键词耦合强度的计算方法进行了改进^[9]。

以上基于关键词和引文的作者耦合关系研究都可用于探测作者相似性,但其中又存在一些不足:

首先,这些研究多在一种对称视角下进行考察,即在作者A的视角下考察作者A与作者B的相似度,与从作者B的视角下考察的相似度一致。但在实际情况中,由于不同作者研究兴趣的广泛程度不同,他们的相似性在不同视角下往往有所差异。

其次,以上作者耦合分析中只考虑了作者使用的关键词或引文的种类,并没有考虑他们使用的频次。由于作者的发文量不同,即使他们的研究兴趣相同,其研究强度也有所差异。应该将这

种研究强度也纳入作者的相似度计算中。

此外, 受作者写作风格等诸方面因素的影响, 不同作者可能会使用不同的关键词或引文来表征同一含义, 或是对同一事物从不同角度选取关键词或引文进行表达, 即有些关键词或引文作者虽然没有直接使用, 但可能间接使用了, 将这种情况纳入考虑更能真实准确地反映作者之间的相似情况。

针对以上不足, 本文首先以作者使用的关键词数据为依据, 基于关键词共现矩阵得到词关联度矩阵, 进一步推算出作者使用所有关键词的可能频次, 从而得到广义上的作者-关键词共现矩阵; 然后基于该矩阵, 从非对称视角构建融合作者兴趣研究强度的广义作者相似度算法, 并探讨该算法与传统算法的优劣; 进一步将该算法应用到作者使用的引文数据上, 对比基于关键词和基于引文的作者相似度计算有何异同。

2 作者相似度算法

由于作者文献耦合分析与作者关键词耦合分析的基本原理相同, 只是选取的研究单元不同, 因此, 本文以关键词数据为例进行算法构建的说明。

2.1 对称视角下的作者相似度算法

传统的基于关键词耦合计算作者相似度的算法原理是, 通过统计两位作者共同使用的关键词种类数占他们使用的所有关键词种类数的比重来表征他们的相似度。假设待研究作者集合为 $Z = \{A_1, A_2, \dots, A_n\}$, 他们对应的关键词集合为 $G = \{K_1, K_2, \dots, K_n\}$, 则作者 A_i 与作者 A_j 的相似度 λ_{ij} 可定义为:

$$\lambda_{ij} = \frac{K_i \cap K_j}{K_i \cup K_j} (i, j = 1, 2, \dots, n) \quad (\text{公式 1})$$

2.2 非对称视角下的作者相似度算法

2.2.1 非对称视角

公式 1 从对称视角下将作者 A_i 与作者 A_j 所使用的关键词不加区分地进行集合, 但在实际研究中, 由于作者研究兴趣的广泛程度不同, 共同的关键词在他们各自研究中所占的比重是不同的。比如, 作者 A_i 可能使用了 100 种关键词, 其中有 20 种与作者 A_j 使用的关键词相同, 而作者 A_j 只使用了 20 种关键词, 且全部包含在作者 A_i 使用的关键词中。很明显, 作者 A_i 的研究兴趣要比作者 A_j 更广泛, 二者的相似度存在很大不同, 运用公式 1 进行作者相似度计算难以真实准确地反映作者之间的相似情况。

因此, 本文考虑从一种非对称视角下计算作者的相似度: 即在作者 A_i 视角下, 他与作者 A_j 的相似度由他们共同使用的关键词种类数占作者 A_i 使用的所有关键词种类数的比重决定, 于是在作者 A_i 的视角下, 作者 A_i 与作者 A_j 的相似度 $\lambda_{i \rightarrow j}$ 可定义为:

$$\lambda_{i \rightarrow j} = \frac{K_i \cap K_j}{K_i} (i, j = 1, 2, \dots, n) \quad (\text{公式 2})$$

2.2.2 作者兴趣研究强度

公式2虽然从非对称视角下计算了作者的相似度,但是仅考虑作者使用的关键词种类数,并未考虑关键词被使用的频次。受作者发文量的影响,假设作者 A_i 和作者 A_j 共同使用了两种关键词,但作者 A_i 使用这两种关键词的频次分别为50次和40次,而作者 A_j 使用这两种关键词的频次仅为3次和2次,虽然二者对这一关键词均有研究兴趣,但是研究强度却颇为不同。为此,本文将关键词的使用频次也纳入作者的相似度计算中。

假设作者 A_i 与作者 A_j 使用的共同关键词集合 $K_{ij} = \{k_1, k_2, \dots, k_s\}$,其中,作者 A_i 使用这些共同关键词的频次依次为 $q_i = \{q_{i1}, q_{i2}, \dots, q_{is}\}$,而作者 A_j 使用这些共同关键词的频次依次为 $q_j = \{q_{j1}, q_{j2}, \dots, q_{js}\}$ 。因此,考虑关键词频次的影响,作者 A_i 对他们共同兴趣的研究强度 δ_i 可定义为:

$$\delta_i = \frac{\sum_{t=1}^s q_{it}}{\sum_{t=1}^s (q_{it} + q_{jt})} (i, j = 1, 2, \dots, n) \quad (\text{公式3})$$

2.2.3 词关联性

公式1~公式3仅考虑了作者直接使用的关键词数据,由于词与词之间存在关联性,受作者写作风格等诸方面因素的影响,不同的作者可能使用了不同的关键词来表征同一含义,或是对同一事物从不同角度选取了关键词进行表达,将关键词之间的关联性纳入考虑更能真实准确地反映作者之间的相似度。

为此,本文考虑基于作者使用的关键词共现矩阵构建关键词关联度矩阵。记所有作者使用的关键词组成的集合为 $K = K_1 \cup K_2 \cup \dots \cup K_n = \{k_1, k_2, \dots, k_m\}$,其中,关键词 k_p 出现的频次为 f_p ,关键词 k_p 与关键词 k_q 共现的频次为 f_{pq} , $f_{pq} = f_{qp}$ ($p, q = 1, 2, \dots, m$),则从关键词 k_p 视角下考虑关键词 k_p 与 k_q 的关联度 $\eta_{p \rightarrow q}$ 可定义为:

$$\eta_{p \rightarrow q} = \frac{f_{pq}}{f_p} \quad (\text{公式4})$$

记关键词与关键词的关联矩阵为 η ,则 η 为一个 $m \times m$ 的非对称矩阵;记作者-关键词共现矩阵为 θ , θ 为一个 $n \times m$ 的矩阵,行数为作者总数,列数为关键词总数,元素为作者使用的关键词频次,定义矩阵 $\varphi = \theta \times \eta$,则 φ 为融入词关联性的广义作者-关键词共现矩阵。

综上,在广义作者-关键词共现矩阵中,假设作者 A_i 使用的关键词集合为 K'_i ,作者 A_j 使用的关键词集合为 K'_j ,他们所使用的共同关键词集合为 $K'_{ij} = \{K'_1, K'_2, \dots, K'_s\}$,其中,作者 A_i 使用这些共同关键词的频次依次为 $q'_i = \{q'_{i1}, q'_{i2}, \dots, q'_{is}\}$,作者 A_j 使用这些共同关键词的频次依次为 $q'_j = \{q'_{j1}, q'_{j2}, \dots, q'_{js}\}$ 。根据广义作者-关键词共现矩阵 φ ,结合公式2与公式3,可计算出非对称视角下融入作者研究强度的广义作者相似度。在作者 A_i 的视角下,融合了作者研究强度的作者 A_i 与 A_j 的广义相似度 $\lambda'_{i \rightarrow j}$ 可定义为:

$$\lambda'_{i \rightarrow j} = \frac{\sum_{t=1}^s q'_{it}}{\sum_{t=1}^s (q'_{it} + q'_{jt})} * \frac{K'_i \cap K'_j}{K'_i} (i, j = 1, 2, \dots, n) \quad (\text{公式5})$$

同理, 将关键词数据替换成引文数据, 可以构造作者-引文共现矩阵。假设作者 A_i 使用的引文集合为 K_i'' , 作者 A_j 使用的引文集合为 K_j'' , 他们所使用共同引文集合为 K_{ij}'' , 其中, 作者 A_i 使用这些共同引文的频次依次为 q_{it}'' , 作者 A_j 使用这些共同引文的频次依次为 q_{jt}'' , 那么, 在作者 A_i 的视角下, 基于引文数据的融入作者兴趣研究强度的广义作者相似度可定义为:

$$\lambda_{i \rightarrow j}'' = \frac{\sum_{t=1}^s q_{it}''}{\sum_{t=1}^s (q_{it}'' + q_{jt}'')} * \frac{K_i'' \cap K_j''}{K_i''} (i, j = 1, 2, \dots, n) \quad (\text{公式 6})$$

3 实证分析

3.1 数据集构建

本文的数据集来源于 Web of Science 核心合集中信息科学领域 8 种具有代表性的国外情报学期刊^[20] (*Journal of the Association for Information Science and Technology*、*Scientometrics*、*Library Journal*、*Information Processing Management*、*Journal of Academic Librarianship*、*Electronic Library*、*Journal of Information Science*、*Library Trends*), 发文时间限定为 2015 ~ 2021 年, 文献类型限定为 Article, 检索到 5832 篇文献, 共计 4084 位第一作者, 11411 个关键词, 1622674 篇引文。对上述文献导出题录数据后利用 TDA (Thomson Data Analyzer) 软件对关键词大小写、缩写以及作者重名等现象进行数据清洗。考虑到实验效果, 本文选择发文量排名前 30 的作者进行研究。

3.2 结果分析

假设有 n 位作者, 不考虑作者与自身的相似度, 则在非对称视角下, n 位作者之间共有 $n(n-1)$ 种相似度; 而在对称视角下, n 位作者之间共有 $n(n-1)/2$ 种相似度。因此, 为了简化研究, 在进行算法结果的对比分析时, 本文考虑以发文量排名第一的作者为例, 分析在该作者视角下与其他作者的相似度, 以及在其他作者视角下与该作者的相似度变化情况。同时, 从非对称视角、融入作者兴趣研究强度、基于广义作者-关键词共现矩阵三个方面依次进行对比, 以验证本文构建的作者相似度算法的有效性。最后, 将此算法应用到作者使用的引文数据, 对比基于关键词数据和基于引文数据计算出的作者相似度结果的差异。

3.2.1 非对称视角下的作者相似度

通过图 1 可以看出, 相同的两位作者在非对称视角下的相似度均高于在对称视角下的相似度, 以作者 1 和作者 2 为例, 他们在对称视角下的相似度最低, 在作者 1 视角下的相似度最高, 在作者 2 视角下的相似度略低于在作者 1 视角下的相似度。从发文量较低的作者视角看他与发文量高的作者的相似度要明显高于从发文量高的作者视角下和对称视角下他们的相似度, 这是由于, 发文量低的作者的研究兴趣一般较为单一, 而发文量高的作者的研究兴趣则相对较为宽泛, 前者的研究兴趣大都包含在后者的研究兴趣中, 所以在发文量低的作者视角下, 他们的相似度比较高。此外, 对于同一位作者, 在不同视角下与其相似度最高的作者也不同, 例如, 对作者 1 来说, 在对称视角下, 作者 13 与其相似度最高; 但在作者 1 视角下, 作者 2 与其相似度最高; 在其他作者视角下, 作者 22 与其相似度最高。可见, 传统对称视角下的作者相似度并不能很好地

反映作者之间的相似情况，从非对称视角计算作者相似度更符合客观事实，尤其是在作者相似推荐中，从目标作者视角下寻找与其相似的作者往往结果更精准。

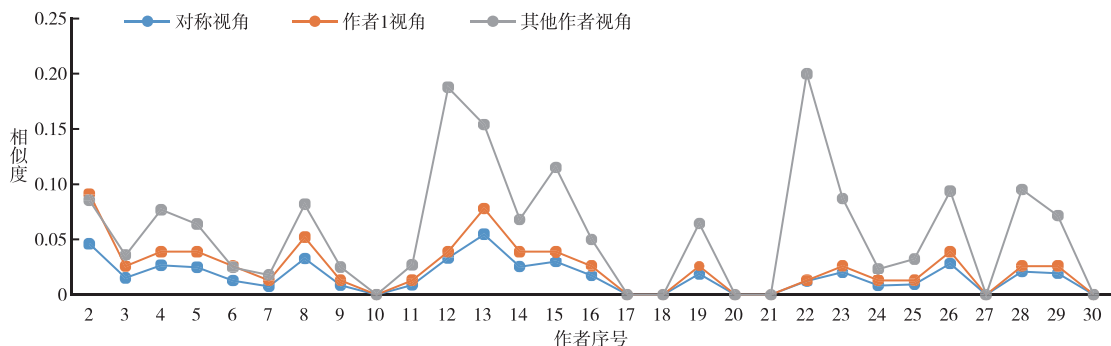


图1 非对称视角下的作者相似度

此外，结合表1和表2所有作者之间的相似度情况也可以看出，相同的两位作者，在不同作者视角下看他们的相似度均要高于在对称视角下看他们的相似度。如，Thelwall,M与Bornmann,L在对称视角下的相似度为0.046；但是在非对称视角下，从Thelwall,M视角看他与Bornmann,L的相似度为0.091，而从Bornmann,L视角看他与Thelwall,M的相似度为0.085；这说明Bornmann,L的研究兴趣要比Thelwall,M的略为宽泛。在对称视角下，除Chang,Y W与Huang,M H、Bu,Y与Huang,Y这两对作者之间的相似度在0.1以上（分别为0.127和0.200），其他作者之间的相似度均在0.1以下；但在非对称视角下，有多位作者之间的相似度在0.1以上，其中在Bu,Y视角下，他与Huang,Y的相似度达到了0.438；在Kousha,K视角下，他与Walters,W H的相似度也达到了0.400；而在对称视角下，Kousha,K与Walters,W H的相似度仅为0.057。

3.2.2 融入研究强度后的作者相似度

任意两位作者对他们共同兴趣的研究强度存在着较大差异（见表3），假设这种差异性主要体现在发文量上，那么融入研究强度后的作者相似度有所下降（见图2）。例如，未考虑研究强度时，在作者1的视角下，作者1与作者12的相似度要高于考虑研究强度时作者1与作者12的相似度，在作者12视角下同样如此。尤其是从发文量低的作者视角来看，由于其研究强度低于发文量高的作者，因此融合了研究强度后的相似度降低明显。

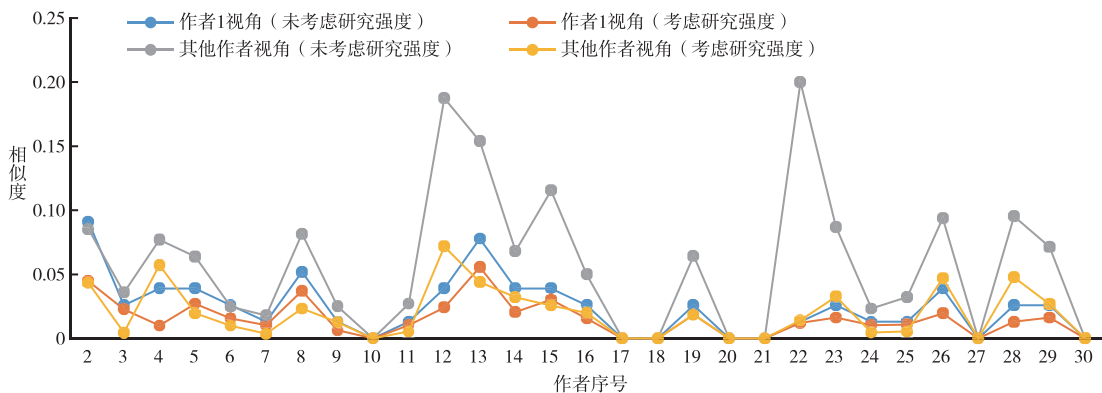


图2 融入研究强度的作者相似度

表 1 对称视角下的作者相似度 (基于关键词数据)

相似度	T.M	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	B.D
TheWall,M	1.000	0.046	0.015	0.027	0.025	0.013	0.068	0.033	0.009	0.000	0.009	0.033	0.055	0.025	0.030	0.017	0.000	0.000	0.019	0.000	0.000	0.000	0.012	0.020	0.008	0.009	0.028	0.000	0.021	0.019	0.000	
Borrmann,L	0.046	1.000	0.030	0.025	0.008	0.032	0.000	0.023	0.008	0.017	0.008	0.032	0.043	0.024	0.029	0.008	0.009	0.000	0.027	0.000	0.000	0.012	0.029	0.016	0.027	0.018	0.019	0.000	0.028	0.000		
Leydesdorf,L	0.030	0.030	1.000	0.067	0.010	0.022	0.018	0.029	0.032	0.000	0.011	0.014	0.011	0.042	0.012	0.021	0.000	0.000	0.024	0.000	0.000	0.017	0.000	0.031	0.012	0.035	0.000	0.013	0.024	0.000		
Abramo,G	0.027	0.025	0.067	1.000	0.024	0.008	0.022	0.086	0.013	0.028	0.013	0.078	0.068	0.092	0.066	0.026	0.000	0.017	0.045	0.000	0.000	0.000	0.051	0.025	0.029	0.060	0.031	0.017	0.047	0.020		
Kim,J	0.025	0.008	0.010	0.024	1.000	0.000	0.010	0.021	0.012	0.012	0.024	0.033	0.012	0.022	0.028	0.000	0.013	0.015	0.013	0.000	0.000	0.000	0.014	0.011	0.000	0.013	0.000	0.000	0.014	0.000		
Silva,J	0.013	0.032	0.022	0.008	0.000	1.000	0.015	0.008	0.008	0.000	0.000	0.000	0.043	0.016	0.009	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.025	0.037	0.018	0.009	0.000	0.019	0.000			
Garcia,JA	0.008	0.000	0.018	0.022	0.010	0.015	1.000	0.019	0.011	0.000	0.011	0.029	0.022	0.020	0.025	0.011	0.000	0.013	0.000	0.015	0.015	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.015		
Zhang,L	0.033	0.023	0.029	0.086	0.021	0.008	0.019	1.000	0.023	0.012	0.012	0.048	0.035	0.045	0.056	0.011	0.000	0.015	0.039	0.000	0.000	0.014	0.022	0.013	0.025	0.000	0.029	0.041	0.000			
Chang,YW	0.009	0.008	0.032	0.013	0.012	0.008	0.011	0.023	1.000	0.000	0.000	0.018	0.000	0.012	0.000	0.000	0.000	0.000	0.127	0.000	0.000	0.000	0.000	0.000	0.029	0.059	0.000	0.034	0.030	0.000		
Ronda,GA	0.000	0.017	0.000	0.028	0.012	0.000	0.000	0.012	0.000	1.000	0.000	0.041	0.000	0.000	0.034	0.000	0.000	0.000	0.015	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
Zhang,Y	0.009	0.008	0.011	0.013	0.024	0.000	0.011	0.012	0.000	0.000	1.000	0.019	0.013	0.025	0.033	0.013	0.030	0.000	0.015	0.000	0.000	0.017	0.000	0.039	0.000	0.000	0.000	0.000	0.016	0.000		
Bu,Y	0.033	0.032	0.014	0.078	0.033	0.000	0.029	0.048	0.018	0.041	0.019	1.000	0.058	0.034	0.200	0.018	0.000	0.000	0.044	0.000	0.000	0.050	0.054	0.035	0.022	0.000	0.028	0.048	0.000			
Orduna,E	0.055	0.043	0.011	0.068	0.012	0.043	0.022	0.035	0.000	0.000	0.013	0.058	1.000	0.025	0.066	0.000	0.000	0.000	0.029	0.000	0.000	0.048	0.051	0.025	0.029	0.000	0.015	0.017	0.063	0.000		
Prathap,G	0.025	0.024	0.042	0.092	0.022	0.016	0.020	0.045	0.012	0.000	0.025	0.034	0.025	1.000	0.029	0.024	0.000	0.000	0.027	0.000	0.000	0.031	0.048	0.027	0.056	0.029	0.000	0.043	0.018			
Huang,Y	0.030	0.029	0.012	0.066	0.028	0.009	0.025	0.056	0.000	0.034	0.033	0.200	0.066	0.029	1.000	0.000	0.000	0.000	0.018	0.000	0.000	0.033	0.021	0.015	0.000	0.000	0.000	0.019	0.000			
Chan,HF	0.017	0.008	0.021	0.026	0.000	0.017	0.011	0.011	0.000	0.000	0.013	0.018	0.000	0.024	0.000	1.000	0.000	0.017	0.000	0.000	0.000	0.000	0.016	0.025	0.014	0.014	0.000	0.000	0.030	0.019		
Chen,L	0.000	0.009	0.000	0.000	0.013	0.000	0.000	0.000	0.000	0.000	0.030	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
Zhang,J	0.000	0.000	0.000	0.017	0.015	0.000	0.013	0.015	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	1.000	0.000	0.000	0.016	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
Huang,MH	0.019	0.027	0.024	0.045	0.013	0.000	0.000	0.039	0.127	0.015	0.015	0.044	0.029	0.027	0.018	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.080	0.014	0.033	0.050	0.018	0.020	0.035	0.000		
Hu,X	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.026	0.000		
Wu,D	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.037	0.000	0.000	0.000	0.000		
Kousha,K	0.012	0.012	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.050	0.048	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.057	0.000	0.000	0.000	0.000		
Li,J	0.020	0.029	0.000	0.051	0.014	0.000	0.000	0.014	0.000	0.000	0.017	0.054	0.051	0.031	0.021	0.016	0.000	0.000	0.080	0.000	0.000	0.000	1.000	0.048	0.000	0.019	0.000	0.000	0.063	0.000		
Moed,HF	0.008	0.016	0.031	0.025	0.011	0.025	0.021	0.022	0.000	0.000	0.039	0.035	0.025	0.048	0.015	0.025	0.000	0.016	0.014	0.014	0.000	0.000	0.048	1.000	0.014	0.027	0.000	0.000	0.060	0.000		
Sotudeh,H	0.009	0.027	0.012	0.029	0.000	0.037	0.000	0.013	0.029	0.000	0.000	0.022	0.029	0.027	0.000	0.014	0.000	0.000	0.033	0.000	0.000	0.000	0.000	0.014	1.000	0.033	0.000	0.020	0.017	0.023		
Walters,WH	0.028	0.018	0.035	0.060	0.013	0.018	0.000	0.025	0.059	0.000	0.000	0.000	0.000	0.056	0.000	0.014	0.000	0.000	0.050	0.000	0.000	0.037	0.057	0.019	0.027	1.000	0.054	0.019	0.034	0.023		
Yuret,T	0.000	0.019	0.000	0.031	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.029	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.000	0.020	0.019	0.000	1.000	0.000			
Deng,SL	0.021	0.000	0.013	0.017	0.000	0.000	0.000	0.029	0.034	0.000	0.000	0.028	0.017	0.000	0.000	0.000	0.000	0.000	0.020	0.000	0.000	0.000	0.000	0.000	0.020	0.019	0.000	1.000	0.000			
Chinehilla,Z	0.019	0.028	0.024	0.047	0.014	0.019	0.000	0.041	0.030	0.000	0.016	0.048	0.063	0.043	0.019	0.030	0.000	0.000	0.035	0.026	0.000	0.000	0.063	0.060	0.017	0.034	0.019	0.000	1.000	0.000		
Bawden,D	0.000	0.000	0.000	0.020	0.000	0.000	0.015	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023	0.023	0.000	0.000	0.000	1.000		

表2 非对称视角下的作者相似度(基于关键词数据)

Table with columns for similarity (相似度) and authors (作者) T, M, B, L, L, A, G, K, J, S, J, A, G, J, A, Z, L, C, Y, W, R, G, A, Z, Y, B, Y, O, E, P, G, H, Y, C, H, F, C, L, Z, J, H, M, H, X, W, D, K, K, L, J, M, H, F, S, H, W, W, H, Y, T, D, S, L, C, Z, B, D. Rows list authors like Thelwall, M, Bornmann, L, Leydesdorff, L, etc.

表 3 非对称视角下的作者研究强度 (基于关键词数据)

研究强度	T.M	B.L	L.L	A.G	K.J	S.JA	G.JA	Z.L	C.YW	R.CA	Z.Y	B.Y	O.E	P.C	H.Y	C.HF	C.L	Z.J	H.MH	H.X	W.D	K.K	L.J	M.HF	S.H	W.WH	Y.T	D.SL	C.Z	B.D
TheWall,M	1.000	0.492	0.875	0.257	0.692	0.600	0.800	0.714	0.500	0.000	0.800	0.618	0.714	0.529	0.778	0.600	0.000	0.000	0.714	0.000	0.000	0.929	0.625	0.800	0.833	0.500	0.000	0.500	0.625	0.000
Bormann,L	0.508	1.000	0.500	0.478	0.909	0.526	0.000	0.793	0.500	0.500	0.952	0.727	0.811	0.733	0.857	0.333	0.500	0.000	0.880	0.000	0.000	0.750	0.846	0.913	0.583	0.500	0.600	0.000	0.815	0.000
Leydesdorff,L	0.125	0.500	1.000	0.684	0.500	0.417	0.400	0.563	0.714	0.000	0.500	0.167	0.500	0.600	0.500	0.571	0.000	0.000	0.600	0.000	0.000	0.500	0.000	0.556	0.500	0.786	0.000	0.667	0.692	0.000
Abraham,G	0.743	0.522	0.316	1.000	0.850	0.167	0.667	0.667	0.500	0.667	0.933	0.655	0.775	0.629	0.731	0.400	0.000	0.500	0.889	0.000	0.000	0.000	0.857	0.882	0.333	0.500	0.500	0.500	0.828	0.500
Kim,J	0.308	0.091	0.500	0.150	1.000	0.000	0.500	0.600	0.500	0.600	0.600	0.273	0.400	0.333	0.375	0.000	0.500	0.500	0.667	0.000	0.000	0.000	0.500	0.667	0.000	0.500	0.000	0.000	0.500	0.000
Silva,J	0.400	0.474	0.583	0.833	0.000	1.000	0.182	0.556	0.500	0.000	0.000	0.000	0.600	0.667	0.750	0.600	0.000	0.000	0.000	0.000	0.000	0.000	0.667	0.500	0.667	0.500	0.667	0.500	0.692	0.000
Garcia,JA	0.200	0.000	0.600	0.333	0.500	0.818	1.000	0.500	0.750	0.000	0.667	0.250	0.400	0.500	0.333	0.500	0.000	0.500	0.000	0.500	0.000	0.000	0.000	0.600	0.000	0.000	0.000	0.000	0.000	0.500
Zhang,L	0.286	0.207	0.438	0.333	0.400	0.444	0.500	1.000	0.500	0.333	0.500	0.250	0.333	0.444	0.364	0.667	0.000	0.500	0.429	0.000	0.000	0.000	0.333	0.714	0.571	0.714	0.000	0.500	0.545	0.000
Chang,YW	0.500	0.500	0.286	0.500	0.500	0.500	0.250	0.500	1.000	0.000	0.000	0.857	0.000	0.500	0.000	0.000	0.000	0.000	0.625	0.000	0.000	0.000	0.000	0.000	0.778	0.600	0.000	0.778	0.500	0.000
Ronda,GA	0.000	0.500	0.000	0.333	0.500	0.000	0.000	0.667	0.000	1.000	0.000	0.500	0.000	0.000	0.000	0.500	0.000	0.000	0.667	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Zhang,Y	0.200	0.048	0.500	0.067	0.400	0.000	0.333	0.500	0.000	0.000	1.000	0.250	0.250	0.222	0.400	0.500	0.500	0.000	0.500	0.000	0.000	0.333	0.500	0.000	0.000	0.000	0.000	0.333	0.000	0.000
Bu,Y	0.382	0.273	0.833	0.345	0.727	0.000	0.750	0.750	0.143	0.500	0.750	1.000	0.600	0.571	0.643	0.500	0.000	0.000	0.667	0.000	0.000	0.833	0.571	0.667	0.500	0.000	0.000	0.500	0.400	0.000
Orduna,E	0.286	0.189	0.500	0.225	0.600	0.400	0.600	0.667	0.000	0.000	0.750	0.400	1.000	0.400	0.583	0.000	0.000	0.000	0.667	0.000	0.000	0.750	0.556	0.625	0.333	0.000	0.500	0.500	0.467	0.000
Prathap,G	0.471	0.267	0.400	0.371	0.667	0.333	0.500	0.556	0.500	0.000	0.778	0.429	0.600	1.000	0.545	0.571	0.000	0.000	0.750	0.000	0.000	0.000	0.667	0.643	0.429	0.600	0.500	0.000	0.615	0.500
Huang,Y	0.222	0.143	0.500	0.269	0.625	0.250	0.667	0.636	0.000	0.500	0.600	0.357	0.417	0.455	1.000	0.000	0.000	0.000	0.667	0.000	0.000	0.500	0.500	0.667	0.000	0.000	0.000	0.500	0.000	0.000
Chan,HF	0.400	0.667	0.429	0.600	0.000	0.400	0.500	0.333	0.000	0.000	0.500	0.500	0.000	0.429	0.000	1.000	0.000	0.500	0.000	0.000	0.000	0.000	0.500	0.600	0.400	0.667	0.000	0.000	0.333	0.667
Chen,L	0.000	0.500	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Zhang,J	0.000	0.000	0.000	0.500	0.500	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000
Huang,MH	0.286	0.120	0.400	0.111	0.333	0.000	0.000	0.571	0.375	0.333	0.500	0.333	0.333	0.2250	0.333	0.000	0.000	0.000	1.000	0.000	0.000	0.444	0.500	0.500	0.000	0.500	0.500	0.500	0.400	0.000
Hu,X	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Wu,D	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.400	0.000	0.000	0.000	0.000
Kousha,K	0.071	0.250	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.167	0.250	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000
Li,J	0.375	0.154	0.000	0.143	0.500	0.000	0.667	0.000	0.000	0.000	0.667	0.429	0.444	0.333	0.500	0.500	0.000	0.000	0.556	0.000	0.000	0.000	1.000	0.571	0.000	0.500	0.000	0.000	0.364	0.000
Moedt,HF	0.200	0.087	0.444	0.118	0.333	0.333	0.400	0.286	0.000	0.000	0.500	0.333	0.375	0.357	0.333	0.400	0.000	0.500	0.500	0.000	0.000	0.000	0.429	1.000	0.250	0.500	0.000	0.000	0.333	0.000
Sotudeh,H	0.167	0.417	0.500	0.667	0.000	0.500	0.000	0.429	0.222	0.000	0.000	0.500	0.667	0.571	0.000	0.600	0.000	0.000	0.500	0.000	0.000	0.000	0.750	1.000	0.000	0.667	0.000	0.500	0.600	0.500
Walters,WH	0.500	0.500	0.214	0.500	0.500	0.333	0.000	0.286	0.400	0.000	0.000	0.000	0.400	0.000	0.333	0.000	0.333	0.000	0.500	0.000	0.600	0.500	0.500	0.500	0.333	1.000	0.571	0.500	0.333	0.500
Yuret,T	0.000	0.400	0.000	0.500	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.500	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.429	1.000	0.000	0.500	0.000
Deng,SL	0.500	0.000	0.333	0.500	0.000	0.000	0.000	0.500	0.222	0.000	0.000	0.500	0.500	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.500	0.500	0.000	1.000	0.000	0.000
Chinchilla,Z	0.375	0.185	0.308	0.172	0.500	0.308	0.000	0.455	0.500	0.000	0.667	0.600	0.533	0.385	0.500	0.667	0.000	0.000	0.600	0.500	0.000	0.000	0.636	0.667	0.400	0.667	0.500	0.000	1.000	0.000
Bawden,D	0.000	0.000	0.000	0.500	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.333	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.500	0.000	0.000	1.000	0.000

另外,在作者相似度计算中考虑研究强度也可以避免由于作者之间发文量悬殊导致相似度结果出现极端值。例如,未考虑研究强度时,在作者22视角下和作者12视角下,他们与作者1的相似度远远高于在其他作者视角下与作者1的相似度;但是考虑研究强度后,他们与作者1的相似度与其他作者与作者1的相似度差距不太大,其中作者12仍与作者1的相似度最高,但作者22与作者1的相似度却低于大多数作者与作者1的相似度。

此外,若两位作者的研究兴趣完全重合,即两种视角下他们的相似度都为1,此时运用融合研究强度的相似度算法则可以对他们进行二次区分,可以更加深入细致地分析作者之间的相似情况,对于预测作者的潜在合作具有一定帮助。如果融入研究强度后的相似度高,说明作者之前的发文量高,后续与该作者合作的可能性更大。

结合表3和表4也可以看出,虽然Bu,Y和Huang,Y的相似度在众多作者之中最高,但是他们对于共同兴趣的研究强度却不同,Bu,Y为0.643,Huang,Y为0.357;Bornmann,L与Zhang,Y对他们共同兴趣的研究强度差距最大,Bornmann,L为0.952,Zhang,Y仅为0.048;此外,Bornmann,L与Kim,J、Bornmann,L与Moed,HF、Thelwall,M与Kousha,K、Abramo,G与Zhang,Y对于二者共同兴趣的研究强度差距也很大,分别为0.909和0.091、0.913和0.087、0.929和0.071、0.933和0.067。因此,在表4中,融合了研究强度的作者相似度放大了不同视角下的相似度差距。如:未考虑研究强度时,Bornmann,L和Zhang,Y在各自视角下的相似度分别为0.012和0.027(见表2),而考虑了研究强度后的相似度分别为0.012和0.001(见表4)。其他几位作者之间同样如此。

3.2.3 融入词关联度的广义作者相似度

表1~表4中,近45%的作者之间的相似度为0。然而,本文已限定于图书情报学领域,同一领域内作者之间的研究大多具有一定的关联性,因而这一结果与客观事实相违背。究其原因,仅凭作者直接使用的关键词很难真实地反映其相似度,将词关联度纳入考虑可以有效弥补这一不足。

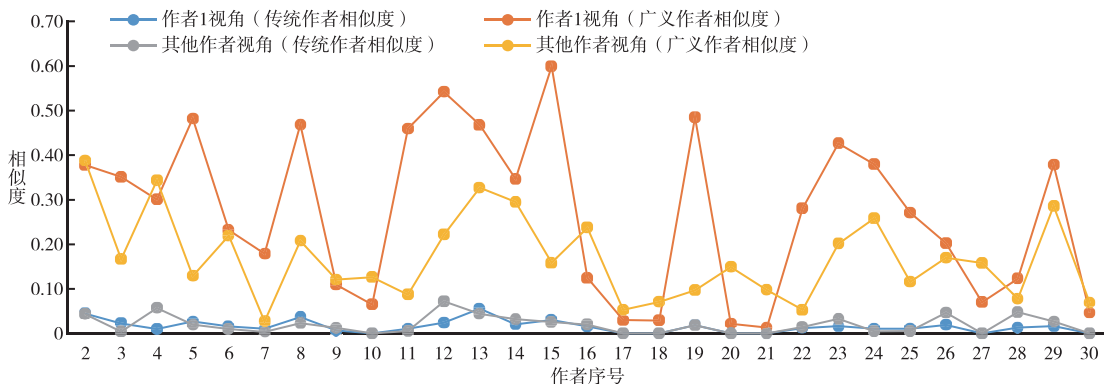


图3 广义作者相似度

表 4 非对称视角下融入研究强度的作者相似度 (基于关键词数据)

相似度	T,M	B,L	L,L	A,G	K,J	S,JA	G,JA	Z,L	C,Y,W	R,C,A	Z,Y	B,Y	O,E	P,C	H,Y	C,H,F	C,L	Z,J	H,M,H	H,X	W,D	K,K	L,J	M,H,F	S,H	W,WH	Y,T	D,SL	C,Z	B,D
TheWall,M	1.000	0.045	0.023	0.010	0.027	0.016	0.010	0.037	0.006	0.000	0.010	0.024	0.056	0.021	0.030	0.016	0.000	0.000	0.019	0.000	0.000	0.012	0.016	0.010	0.011	0.019	0.000	0.013	0.016	0.000
Bormann,L	0.043	1.000	0.024	0.017	0.011	0.032	0.000	0.029	0.006	0.012	0.012	0.027	0.049	0.027	0.031	0.004	0.006	0.000	0.032	0.000	0.000	0.009	0.031	0.022	0.021	0.012	0.015	0.000	0.030	0.000
Leydesdorff,L	0.004	0.036	1.000	0.073	0.009	0.022	0.014	0.030	0.038	0.000	0.009	0.003	0.009	0.043	0.009	0.020	0.000	0.000	0.021	0.000	0.000	0.009	0.000	0.030	0.009	0.042	0.000	0.012	0.025	0.000
Abraimo,G	0.057	0.040	0.049	1.000	0.044	0.004	0.034	0.120	0.013	0.034	0.024	0.067	0.099	0.113	0.075	0.021	0.000	0.013	0.068	0.000	0.000	0.066	0.045	0.017	0.051	0.026	0.013	0.064	0.013	
Kim,J	0.020	0.002	0.011	0.006	1.000	0.000	0.011	0.026	0.011	0.011	0.026	0.012	0.009	0.014	0.016	0.000	0.011	0.011	0.014	0.000	0.000	0.011	0.014	0.000	0.011	0.011	0.000	0.000	0.011	0.000
Silva,J	0.010	0.029	0.022	0.010	0.000	1.000	0.004	0.007	0.006	0.000	0.000	0.000	0.037	0.016	0.009	0.015	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.025	0.025	0.016	0.006	0.000	0.017	0.000
Garcia,JA	0.004	0.000	0.021	0.012	0.009	0.029	1.000	0.018	0.013	0.000	0.012	0.009	0.014	0.018	0.012	0.009	0.000	0.009	0.000	0.009	0.000	0.000	0.000	0.021	0.000	0.000	0.000	0.000	0.000	0.009
Zhang,L	0.023	0.013	0.027	0.048	0.016	0.009	0.020	1.000	0.020	0.007	0.010	0.015	0.020	0.036	0.030	0.014	0.000	0.010	0.026	0.000	0.000	0.007	0.029	0.012	0.029	0.060	0.000	0.020	0.033	0.000
Chang,YW	0.013	0.013	0.021	0.013	0.013	0.013	0.006	0.025	1.000	0.000	0.000	0.021	0.000	0.013	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.039	0.060	0.000	0.039	0.025	0.000
Ronda,GA	0.000	0.029	0.000	0.019	0.014	0.000	0.000	0.019	0.000	0.000	1.000	0.000	0.029	0.000	0.000	0.029	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Zhang,Y	0.005	0.001	0.014	0.002	0.022	0.000	0.009	0.014	0.000	0.000	1.000	0.007	0.007	0.012	0.022	0.014	0.027	0.000	0.014	0.000	0.014	0.000	0.009	0.041	0.000	0.000	0.000	0.000	0.009	0.000
Bu,Y	0.072	0.051	0.052	0.086	0.091	0.000	0.094	0.141	0.009	0.063	0.047	1.000	0.113	0.071	0.281	0.031	0.000	0.000	0.083	0.000	0.000	0.052	0.071	0.083	0.031	0.000	0.000	0.031	0.050	0.000
Orduna,E	0.044	0.024	0.013	0.029	0.015	0.051	0.031	0.051	0.000	0.000	0.019	0.031	1.000	0.021	0.060	0.000	0.000	0.000	0.034	0.000	0.000	0.038	0.043	0.032	0.017	0.000	0.013	0.013	0.048	0.000
Prathap,G	0.032	0.018	0.036	0.059	0.030	0.015	0.023	0.051	0.011	0.000	0.035	0.019	0.027	1.000	0.025	0.026	0.000	0.000	0.034	0.000	0.000	0.030	0.058	0.019	0.055	0.023	0.000	0.042	0.011	
Huang,Y	0.026	0.016	0.019	0.041	0.048	0.010	0.051	0.098	0.000	0.038	0.046	0.096	0.064	0.035	1.000	0.000	0.000	0.026	0.000	0.000	0.000	0.019	0.019	0.026	0.000	0.000	0.000	0.019	0.000	
Chan,HF	0.020	0.017	0.021	0.030	0.000	0.020	0.013	0.008	0.000	0.000	0.013	0.013	0.000	0.021	0.000	1.000	0.000	0.013	0.000	0.000	0.000	0.013	0.030	0.010	0.017	0.000	0.000	0.017	0.017	
Chen,L	0.000	0.016	0.000	0.000	0.016	0.000	0.000	0.000	0.000	0.000	0.031	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
Zhang,J	0.000	0.000	0.000	0.025	0.025	0.000	0.025	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.025	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.025	0.000	0.000	0.000	0.000	0.000	
Huang,MH	0.018	0.012	0.026	0.011	0.011	0.000	0.000	0.055	0.097	0.011	0.016	0.022	0.022	0.016	0.011	0.000	0.000	0.000	1.000	0.000	0.000	0.057	0.016	0.032	0.048	0.016	0.016	0.026	0.000	
Hu,X	0.000	0.000	0.000	0.000	0.000	0.042	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.042	0.000	
Wu,D	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.021	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.033	0.000	0.000	0.000	
Kousha,K	0.014	0.050	0.100	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.033	0.100	0.000	0.100	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.200	0.000	0.000	0.000	0.000	
Li,J	0.033	0.020	0.000	0.019	0.022	0.000	0.000	0.029	0.000	0.000	0.029	0.037	0.058	0.029	0.022	0.022	0.000	0.000	0.097	0.000	0.000	0.000	1.000	0.075	0.000	0.022	0.000	0.000	0.047	0.000
Moed,HF	0.005	0.004	0.031	0.005	0.008	0.023	0.019	0.013	0.000	0.000	0.035	0.016	0.017	0.033	0.008	0.019	0.000	0.012	0.012	0.000	0.000	0.030	1.000	0.006	0.023	0.000	0.000	0.031	0.000	
Sotudeh,H	0.005	0.040	0.016	0.043	0.000	0.065	0.000	0.014	0.014	0.000	0.000	0.016	0.043	0.037	0.000	0.019	0.000	0.000	0.032	0.000	0.000	0.000	0.024	1.000	0.043	0.000	0.016	0.019	0.016	
Walters,WH	0.047	0.031	0.020	0.063	0.016	0.021	0.000	0.018	0.050	0.000	0.000	0.000	0.000	0.050	0.000	0.010	0.000	0.047	0.000	0.000	0.038	0.031	0.016	0.031	0.021	1.000	0.054	0.016	0.021	0.016
Yuret,T	0.000	0.030	0.000	0.037	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.048	1.000	0.000	0.019	0.000	
Deng,SL	0.048	0.000	0.016	0.024	0.000	0.000	0.000	0.048	0.021	0.000	0.000	0.024	0.024	0.000	0.000	0.000	0.000	0.024	0.000	0.000	0.000	0.000	0.000	0.000	0.024	0.000	1.000	0.000	0.000	
Chinchilla,Z	0.027	0.020	0.022	0.018	0.018	0.022	0.000	0.049	0.036	0.000	0.024	0.043	0.076	0.041	0.018	0.048	0.000	0.000	0.043	0.018	0.000	0.000	0.068	0.095	0.014	0.048	0.018	0.000	1.000	0.000
Bawden,D	0.000	0.000	0.000	0.038	0.000	0.000	0.038	0.000	0.000	0.000	0.000	0.000	0.000	0.038	0.000	0.026	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.038	0.000	0.000	0.000	1.000	0.000

图3和表5是同时考虑了作者研究强度和词关联度的广义作者相似度计算结果。可以看出,这一算法丰富了作者使用关键词的种类及频次,作者相似度明显提高,所有作者之间均存在一定的关联,且原先存在关联的作者之间的相似度变大,由弱关联变为强关联。例如,在作者1视角下,与其他29位作者的广义相似度均高于传统相似度,且作者1与其他作者的传统相似度均低于0.1,但广义相似度均值在0.3左右,最高达到了0.6。

同时由于作者发文量的差距也加大了不同视角下的作者相似度结果的差距,即发文量高的作者在考虑词关联度后,可能使用了更多的关键词,因此从他的视角下看与发文量低的作者相似度会继续降低。例如在作者1视角下,他与作者15的广义相似度与在作者15视角下的差距最大(0.569)。基于这种考虑了词关联度的广义作者相似度可以在一定程度上降低受作者写作风格以及发文量较低等因素而导致的零关联假象,对于挖掘作者的潜在合作对象具有一定帮助。

3.2.4 基于引文数据的广义作者相似度

由于一篇文献中的关键词数量远远少于引文数量,对于发文量少的作者来说,仅仅基于关键词数据较难全面地反映其研究情况,同时,引文数据是作者对前人研究成果的吸收与消化,可以间接反映作者的研究情况。因此,本文继续考虑基于引文数据的作者相似度计算,并对比二者结果有何差异。

通过图4可以看出,基于引文数据,作者之间的关联程度被进一步放大。在发文量高的作者1视角下,基于引文数据的广义作者相似度要明显高于基于关键词数据的广义作者相似度,这是由于在同样一份数据集中,引文样本数要远高于关键词样本数。但也正是因为引文样本数过高而导致了引文之间的关联度要低于关键词之间的关联度,在部分发文量低的作者视角下,相似度可能会降低。

此外,结合表5和表6可以看出,基于关键词数据,在Ronda,GA、Chen,L、Zhang,J、Hu,X、Wu,D、Bawden,D的视角下,他们与其他作者的相似度均很低,而在引文数据下,只有在Hu,X、Wu,D、Bawden,D的视角下与其他作者的相似度很低。通过对相似度值从小到大按从红到绿的颜色进行填色,可以看出表6中的红色格数较表5明显减少,尤其是在Ronda,GA、Chen,L、Zhang,J视角下的相似度均由红转绿,即相似度值提高了,这说明在发文量少或者关键词使用种类数较少的情况下,基于引文数据的广义作者相似度比基于关键词数据的广义作者相似度更能表达作者之间的相似性。

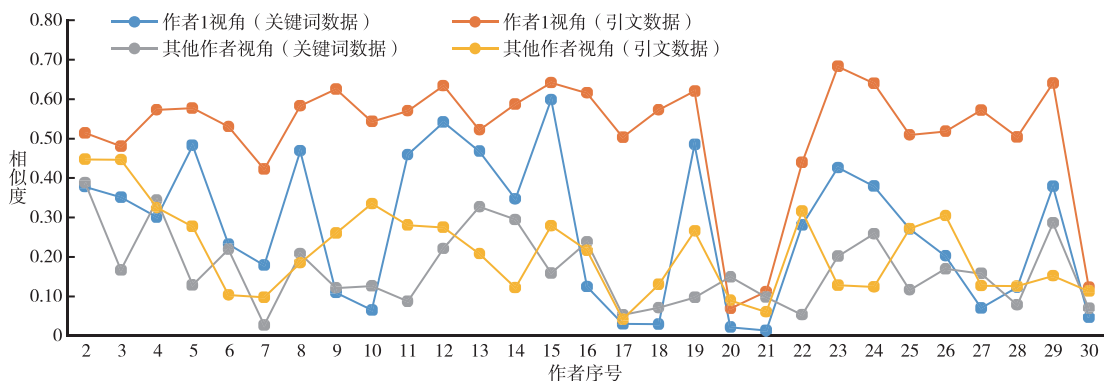


图4 基于关键词数据和基于引文数据的广义作者相似度

表5 非对称视角下融入词关联度的广义作者相似度(基于关键词数据)

相似度	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
	T,M	B,L	L,L	A,G	K,J	S,J	G,A	Z,L	C,Y	R,G	Z,Y	B,Y	O,E	P,G	H,Y	C,H	F,C	L,Z	H,M	H,X	W,D	K,K	L,J	M,H	S,H	W,W	Y,T	D,S	C,Z	B,D
TheWall,M	1.000	0.378	0.351	0.300	0.482	0.232	0.179	0.469	0.109	0.065	0.459	0.542	0.468	0.347	0.599	0.125	0.030	0.029	0.485	0.022	0.013	0.281	0.426	0.379	0.271	0.202	0.070	0.123	0.378	0.047
Bornmann,L	0.388	1.000	0.333	0.406	0.435	0.231	0.140	0.499	0.116	0.052	0.454	0.485	0.500	0.472	0.516	0.182	0.034	0.030	0.462	0.026	0.022	0.214	0.421	0.447	0.296	0.256	0.117	0.101	0.470	0.051
Leydesdorff,L	0.167	0.282	1.000	0.347	0.240	0.188	0.146	0.398	0.228	0.043	0.309	0.255	0.275	0.423	0.306	0.216	0.008	0.030	0.311	0.033	0.021	0.205	0.244	0.307	0.268	0.332	0.062	0.142	0.348	0.053
Abramo,G	0.343	0.345	0.292	1.000	0.471	0.182	0.202	0.560	0.165	0.065	0.481	0.415	0.352	0.481	0.460	0.179	0.017	0.035	0.515	0.028	0.022	0.130	0.432	0.440	0.228	0.303	0.119	0.119	0.482	0.056
Kim,J	0.129	0.094	0.094	0.117	1.000	0.039	0.155	0.292	0.058	0.044	0.392	0.234	0.128	0.144	0.324	0.066	0.018	0.026	0.420	0.022	0.007	0.099	0.292	0.202	0.096	0.087	0.043	0.062	0.172	0.031
Silva,J	0.219	0.341	0.385	0.306	0.222	1.000	0.126	0.354	0.179	0.016	0.280	0.213	0.346	0.381	0.278	0.248	0.016	0.037	0.221	0.048	0.025	0.140	0.209	0.386	0.387	0.321	0.072	0.072	0.399	0.062
Garcia,JA	0.027	0.060	0.236	0.088	0.126	0.276	1.000	0.191	0.240	0.024	0.325	0.115	0.148	0.171	0.108	0.143	0.006	0.031	0.188	0.070	0.006	0.113	0.123	0.227	0.135	0.122	0.028	0.093	0.142	0.059
Zhang,L	0.208	0.225	0.218	0.317	0.337	0.144	0.145	1.000	0.143	0.031	0.385	0.246	0.205	0.334	0.314	0.175	0.016	0.036	0.312	0.023	0.019	0.097	0.288	0.318	0.211	0.264	0.065	0.113	0.343	0.046
Chang,YW	0.121	0.135	0.184	0.170	0.229	0.089	0.100	0.288	1.000	0.067	0.290	0.258	0.150	0.225	0.200	0.186	0.009	0.014	0.389	0.043	0.018	0.065	0.215	0.259	0.305	0.249	0.069	0.256	0.178	0.034
Ronda,GA	0.126	0.314	0.119	0.209	0.408	0.047	0.225	0.377	0.092	1.000	0.339	0.380	0.184	0.228	0.472	0.037	0.012	0.013	0.417	0.012	0.000	0.138	0.249	0.287	0.100	0.112	0.075	0.176	0.237	0.040
Zhang,Y	0.087	0.056	0.055	0.072	0.358	0.035	0.038	0.194	0.035	0.026	1.000	0.134	0.091	0.091	0.182	0.062	0.021	0.013	0.314	0.016	0.028	0.063	0.204	0.187	0.079	0.064	0.039	0.046	0.117	0.029
Bu,Y	0.222	0.200	0.235	0.217	0.491	0.108	0.181	0.425	0.102	0.050	0.500	1.000	0.303	0.256	0.535	0.106	0.015	0.017	0.477	0.026	0.009	0.280	0.422	0.332	0.203	0.144	0.060	0.139	0.272	0.041
Orduna,E	0.327	0.277	0.277	0.294	0.415	0.210	0.127	0.422	0.093	0.053	0.450	0.435	1.000	0.357	0.492	0.187	0.020	0.012	0.449	0.025	0.010	0.270	0.390	0.331	0.193	0.177	0.092	0.092	0.380	0.028
Prathap,G	0.295	0.289	0.248	0.408	0.430	0.156	0.161	0.474	0.130	0.056	0.448	0.365	0.330	1.000	0.415	0.213	0.014	0.029	0.453	0.025	0.028	0.113	0.409	0.341	0.205	0.265	0.111	0.090	0.407	0.055
Huang,Y	0.159	0.143	0.168	0.154	0.413	0.080	0.171	0.328	0.068	0.039	0.480	0.364	0.257	0.176	1.000	0.073	0.015	0.014	0.450	0.019	0.005	0.260	0.339	0.241	0.121	0.098	0.043	0.087	0.207	0.033
Chan,HF	0.237	0.316	0.340	0.404	0.234	0.285	0.147	0.393	0.089	0.008	0.266	0.230	0.179	0.371	0.244	1.000	0.006	0.066	0.197	0.038	0.031	0.044	0.204	0.378	0.307	0.388	0.071	0.058	0.371	0.150
Chen,L	0.053	0.059	0.013	0.026	0.288	0.069	0.023	0.139	0.011	0.004	0.297	0.061	0.112	0.043	0.067	0.019	1.000	0.064	0.116	0.011	0.000	0.040	0.087	0.067	0.096	0.021	0.007	0.020	0.065	0.015
Zhang,J	0.071	0.026	0.077	0.052	0.198	0.057	0.099	0.164	0.018	0.003	0.074	0.039	0.024	0.048	0.062	0.034	0.051	1.000	0.070	0.026	0.017	0.011	0.044	0.224	0.094	0.134	0.019	0.028	0.062	0.045
Huang,MH	0.097	0.068	0.081	0.105	0.259	0.046	0.081	0.367	0.170	0.016	0.342	0.251	0.101	0.132	0.213	0.102	0.006	0.009	1.000	0.015	0.020	0.085	0.277	0.175	0.173	0.092	0.073	0.127	0.187	0.027
Hu,X	0.149	0.048	0.080	0.059	0.204	0.139	0.192	0.103	0.220	0.007	0.288	0.139	0.092	0.084	0.219	0.131	0.018	0.036	0.256	1.000	0.021	0.043	0.163	0.084	0.155	0.183	0.027	0.059	0.143	0.040
Wu,D	0.098	0.151	0.071	0.169	0.087	0.042	0.049	0.062	0.028	0.000	0.138	0.064	0.068	0.103	0.091	0.079	0.000	0.009	0.123	0.010	1.000	0.023	0.132	0.107	0.072	0.104	0.205	0.054	0.086	0.087
Kouhsa,K	0.053	0.048	0.194	0.040	0.138	0.058	0.093	0.125	0.015	0.017	0.209	0.085	0.077	0.063	0.157	0.046	0.011	0.006	0.163	0.008	0.001	1.000	0.131	0.099	0.079	0.064	0.018	0.041	0.077	0.022
Li,J	0.201	0.129	0.118	0.194	0.469	0.096	0.146	0.340	0.073	0.040	0.552	0.348	0.251	0.227	0.399	0.128	0.014	0.016	0.543	0.031	0.023	0.123	1.000	0.320	0.136	0.111	0.089	0.075	0.265	0.047
Moed,HF	0.258	0.258	0.311	0.315	0.382	0.215	0.120	0.422	0.108	0.018	0.475	0.300	0.357	0.470	0.373	0.234	0.009	0.029	0.406	0.030	0.018	0.104	0.386	1.000	0.210	0.244	0.053	0.073	0.439	0.053
Soutdeh,H	0.116	0.279	0.325	0.312	0.173	0.279	0.074	0.336	0.080	0.016	0.223	0.148	0.311	0.342	0.171	0.261	0.017	0.040	0.170	0.032	0.029	0.138	0.145	0.309	1.000	0.367	0.073	0.124	0.314	0.053
Walters,WB	0.170	0.285	0.269	0.325	0.229	0.175	0.111	0.285	0.184	0.038	0.290	0.185	0.195	0.338	0.237	0.212	0.010	0.035	0.331	0.030	0.066	0.095	0.240	0.321	0.295	1.000	0.154	0.138	0.317	0.075
Yurei,T	0.158	0.129	0.117	0.170	0.215	0.070	0.066	0.217	0.036	0.037	0.217	0.169	0.128	0.162	0.218	0.128	0.008	0.015	0.227	0.015	0.074	0.052	0.190	0.210	0.106	0.184	1.000	0.068	0.233	0.067
Deng,SL	0.078	0.067	0.145	0.075	0.150	0.047	0.058	0.136	0.149	0.066	0.213	0.172	0.142	0.092	0.165	0.061	0.013	0.012	0.170	0.017	0.011	0.175	0.129	0.137	0.213	0.202	0.044	1.000	0.118	0.026
Chinehilla,Z	0.286	0.254	0.248	0.314	0.424	0.174	0.129	0.445	0.167	0.029	0.482	0.382	0.323	0.387	0.423	0.239	0.013	0.031	0.461	0.038	0.018	0.122	0.426	0.420	0.239	0.275	0.090	0.092	1.000	0.053
Bawden,D	0.069	0.048	0.134	0.152	0.116	0.083	0.264	0.108	0.076	0.008	0.132	0.062	0.060	0.146	0.093	0.111	0.006	0.049	0.109	0.019	0.078	0.062	0.091	0.114	0.323	0.249	0.122	0.042	0.120	1.000

表6 非对称视角下融入兴趣研究强度的广义作者相似度 (基于引文数据)

相似度	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
Thelwall, M	1.000	0.514	0.480	0.573	0.577	0.531	0.422	0.583	0.625	0.543	0.570	0.634	0.522	0.587	0.642	0.615	0.503	0.573	0.620	0.069	0.112	0.439	0.683	0.640	0.509	0.518	0.572	0.504	0.641	0.122	
Bornmann, L	0.447	1.000	0.466	0.551	0.540	0.507	0.433	0.565	0.610	0.532	0.558	0.623	0.502	0.604	0.618	0.601	0.512	0.561	0.591	0.065	0.099	0.407	0.680	0.636	0.510	0.510	0.571	0.492	0.648	0.101	
Leydesdorff, L	0.447	0.496	1.000	0.556	0.586	0.512	0.448	0.610	0.645	0.554	0.596	0.644	0.513	0.647	0.648	0.606	0.547	0.600	0.623	0.062	0.093	0.373	0.706	0.653	0.485	0.521	0.593	0.526	0.655	0.134	
Abramo, G	0.325	0.346	0.329	1.000	0.484	0.505	0.376	0.517	0.524	0.428	0.474	0.507	0.402	0.548	0.510	0.557	0.517	0.517	0.540	0.053	0.071	0.336	0.650	0.575	0.412	0.427	0.577	0.442	0.609	0.114	
Kim, J	0.277	0.303	0.256	0.362	1.000	0.456	0.357	0.414	0.463	0.356	0.418	0.438	0.333	0.434	0.367	0.426	0.447	0.410	0.415	0.067	0.083	0.222	0.607	0.495	0.381	0.270	0.517	0.384	0.523	0.112	
Silva, J	0.103	0.166	0.152	0.168	0.167	1.000	0.380	0.206	0.199	0.102	0.161	0.207	0.239	0.244	0.272	0.156	0.435	0.269	0.196	0.031	0.085	0.107	0.281	0.268	0.252	0.123	0.236	0.323	0.402	0.052	
Garcia, JA	0.097	0.097	0.044	0.153	0.167	0.219	1.000	0.122	0.225	0.096	0.058	0.097	0.112	0.156	0.078	0.132	0.308	0.134	0.095	0.023	0.030	0.118	0.148	0.181	0.118	0.139	0.142	0.169	0.293	0.058	
Zhang, L	0.185	0.209	0.188	0.277	0.404	0.438	0.435	1.000	0.413	0.254	0.279	0.321	0.338	0.491	0.340	0.328	0.512	0.451	0.352	0.065	0.073	0.145	0.493	0.490	0.232	0.375	0.424	0.432	0.447	0.132	
Chang, YW	0.260	0.267	0.231	0.372	0.382	0.470	0.373	0.415	1.000	0.366	0.380	0.435	0.324	0.463	0.508	0.507	0.436	0.049	0.076	0.206	0.556	0.513	0.319	0.349	0.502	0.398	0.495	0.502	0.398	0.495	0.103
Ronda, GA	0.335	0.362	0.337	0.460	0.489	0.534	0.428	0.543	0.516	1.000	0.497	0.531	0.457	0.618	0.527	0.553	0.537	0.562	0.518	0.038	0.098	0.313	0.673	0.599	0.441	0.420	0.608	0.453	0.588	0.088	
Zhang, Y	0.280	0.292	0.278	0.373	0.420	0.491	0.422	0.505	0.480	0.381	1.000	0.464	0.342	0.579	0.439	0.411	0.537	0.558	0.456	0.069	0.108	0.245	0.628	0.543	0.342	0.343	0.534	0.449	0.512	0.103	
Bu, Y	0.275	0.298	0.285	0.397	0.434	0.459	0.421	0.501	0.480	0.401	0.438	1.000	0.350	0.561	0.461	0.479	0.510	0.538	0.456	0.063	0.079	0.242	0.611	0.541	0.379	0.359	0.547	0.463	0.546	0.095	
Orduna, E	0.208	0.241	0.246	0.372	0.416	0.504	0.419	0.419	0.447	0.254	0.394	0.439	1.000	0.486	0.425	0.393	0.527	0.499	0.468	0.067	0.115	0.272	0.583	0.553	0.328	0.373	0.419	0.450	0.497	0.100	
Prathap, G	0.122	0.119	0.112	0.216	0.311	0.440	0.411	0.288	0.262	0.174	0.230	0.232	0.230	1.000	0.222	0.255	0.549	0.390	0.231	0.039	0.065	0.137	0.407	0.369	0.183	0.170	0.382	0.292	0.332	0.060	
Huang, Y	0.279	0.310	0.300	0.393	0.492	0.430	0.444	0.508	0.483	0.385	0.433	0.477	0.385	0.563	1.000	0.466	0.522	0.531	0.473	0.064	0.089	0.261	0.611	0.563	0.339	0.362	0.524	0.449	0.541	0.129	
Chan, HF	0.216	0.234	0.220	0.341	0.382	0.515	0.393	0.445	0.403	0.327	0.406	0.387	0.362	0.500	0.396	1.000	0.509	0.485	0.438	0.038	0.062	0.258	0.577	0.501	0.328	0.321	0.549	0.363	0.509	0.097	
Chen, L	0.042	0.046	0.037	0.075	0.175	0.295	0.262	0.128	0.108	0.057	0.083	0.084	0.109	0.208	0.085	0.091	1.000	0.202	0.126	0.055	0.047	0.054	0.221	0.179	0.073	0.076	0.189	0.140	0.182	0.036	
Zhang, J	0.130	0.129	0.116	0.220	0.335	0.408	0.417	0.307	0.263	0.189	0.243	0.230	0.215	0.424	0.226	0.264	0.521	1.000	0.265	0.060	0.079	0.119	0.460	0.361	0.244	0.194	0.391	0.334	0.397	0.096	
Huang, MH	0.265	0.284	0.275	0.336	0.443	0.486	0.430	0.480	0.472	0.374	0.424	0.451	0.329	0.555	0.451	0.425	0.518	0.522	1.000	0.060	0.085	0.219	0.606	0.559	0.289	0.319	0.506	0.407	0.473	0.113	
Hu, X	0.090	0.046	0.087	0.042	0.331	0.099	0.278	0.212	0.201	0.111	0.141	0.106	0.061	0.172	0.084	0.156	0.162	0.205	0.153	1.000	0.228	0.040	0.221	0.174	0.055	0.139	0.177	0.367	0.222	0.110	
Wu, D	0.060	0.092	0.138	0.168	0.208	0.264	0.277	0.226	0.167	0.069	0.119	0.194	0.105	0.191	0.174	0.209	0.291	0.319	0.201	0.160	1.000	0.099	0.246	0.250	0.113	0.152	0.219	0.300	0.273	0.083	
Kousha, K	0.316	0.354	0.380	0.443	0.516	0.542	0.414	0.529	0.565	0.416	0.493	0.550	0.486	0.502	0.533	0.534	0.492	0.542	0.573	0.080	0.119	1.000	0.651	0.625	0.513	0.537	0.503	0.489	0.591	0.097	
Li, J	0.128	0.136	0.131	0.212	0.232	0.421	0.409	0.332	0.297	0.208	0.244	0.261	0.215	0.412	0.246	0.287	0.464	0.357	0.259	0.043	0.068	0.139	1.000	0.394	0.252	0.199	0.428	0.283	0.398	0.095	
Moed, HF	0.124	0.137	0.142	0.216	0.261	0.482	0.408	0.308	0.313	0.204	0.233	0.278	0.261	0.418	0.268	0.292	0.516	0.395	0.276	0.059	0.071	0.154	0.445	1.000	0.207	0.245	0.394	0.302	0.356	0.083	
Sotudeh, H	0.272	0.286	0.295	0.404	0.349	0.506	0.447	0.489	0.490	0.322	0.390	0.413	0.447	0.511	0.486	0.472	0.501	0.450	0.518	0.056	0.105	0.228	0.500	0.595	1.000	0.427	0.418	0.468	0.447	0.075	
Walters, WH	0.305	0.307	0.298	0.417	0.532	0.554	0.400	0.395	0.512	0.421	0.477	0.499	0.381	0.606	0.509	0.529	0.574	0.599	0.566	0.068	0.109	0.244	0.631	0.561	0.359	1.000	0.599	0.415	0.541	0.103	
Yuret, T	0.127	0.125	0.118	0.197	0.242	0.445	0.398	0.303	0.275	0.189	0.213	0.226	0.283	0.381	0.224	0.270	0.489	0.372	0.278	0.032	0.058	0.180	0.421	0.368	0.278	0.174	1.000	0.292	0.408	0.107	
Deng, SL	0.126	0.140	0.113	0.233	0.301	0.307	0.359	0.317	0.284	0.210	0.233	0.215	0.229	0.384	0.234	0.302	0.477	0.399	0.279	0.056	0.126	0.119	0.444	0.364	0.171	0.252	0.338	1.000	0.402	0.151	
Chinchilla, Z	0.152	0.162	0.158	0.234	0.273	0.311	0.315	0.329	0.329	0.225	0.273	0.286	0.276	0.415	0.300	0.332	0.455	0.350	0.342	0.050	0.059	0.166	0.442	0.435	0.346	0.275	0.385	0.314	1.000	0.096	
Bawden, D	0.113	0.280	0.173	0.160	0.196	0.262	0.175	0.202	0.222	0.397	0.470	0.385	0.192	0.604	0.094	0.240	0.521	0.390	0.202	0.016	0.039	0.126	0.290	0.206	0.145	0.210	0.170	0.216	0.257	1.000	

由于关键词是深入文献内部单元、由作者给出以代表文章主题的词语, 更能反映作者的研究兴趣, 而引文是文章的延伸, 涵盖了作者的兴趣点及涉猎范围。因而基于关键词数据计算作者相似度相对更准确, 而基于引文数据相对更宽泛、但准确度会下降。

4 结束语

本研究主要针对传统基于作者耦合关系探测作者相似度算法中的不足进行了改进。首先从非对称视角计算了作者相似度, 更能体现由于作者研究的广泛程度不同导致他们相似度的差异性; 其次, 本文将作者使用关键词的频次考虑到作者相似度的计算中, 得到了融合作者研究强度的相似度, 可以对发文量不同的作者做进一步区分。这两种算法均不同程度地展现了作者的深层差异, 符合客观事实, 可以对作者相似度进行更加细致的分析, 有助于提高作者相似度计算的准确性。此外, 针对作者之间无关联现象, 本文运用广义作者-关键词共现矩阵, 将词关联性纳入计算, 可以降低由于作者发文量少或者无直接使用的共同关键词而产生同领域却零关联现象, 更加符合客观事实。最后本文将基于关键词数据的广义作者相似度算法应用到引文数据, 进一步放大了作者之间的关联度, 对于挖掘发文量较少或者使用关键词不多、甚至跨学科领域的作者之间的相似性, 具有一定的实用性。

总体而言, 本文提出的作者相似度算法对进一步提高作者相似推荐的准确性、挖掘作者之间的潜在合作关系具有一定帮助, 但也存在一些不足之处: 首先, 所用案例局限于图书情报领域内的作者相似情况, 未拓展至其他学科领域; 其次, 关于词关联度与引文关联度的计算只是基于共现矩阵, 并没有考虑语义上的关联。这些都有待于后续深入研究。

【参考文献】

- [1] 程结晶, 李秀霞. 基于 J-NMF 算法的作者类团划分研究 [J]. 情报理论与实践, 2020, 43(11): 40-46.
- [2] 马瑞敏, 倪超群. 基于作者同被引分析的我国图书情报学知识结构及其演变研究 [J]. 中国图书馆学报, 2011, 37(6): 17-26.
- [3] 刘爱琴, 吴瑞瑞. 基于引证关系的作者与主题多重共现网络构建 [J]. 数字图书馆论坛, 2019(10): 23-29.
- [4] 何劲, 关鹏, 王曰芬. 作者-主题关联的学科知识网络构建与演化分析 [J]. 情报科学, 2019, 37(1): 56-62, 67.
- [5] 陈卫静, 郑颖. 基于作者关键词耦合的潜在合作关系挖掘 [J]. 情报杂志, 2013, 32(5): 127-131.
- [6] 丁敬达, 郭杰. 融合内容相似度和路径相似性的潜在作者合作关系挖掘 [J]. 情报理论与实践, 2021, 44(1): 124-128, 123.
- [7] 李爱明. 社会网络视角下的国内数字图书馆领域作者合作关系研究 [J]. 情报科学, 2013, 31(11): 57-63.
- [8] 马瑞敏, 倪超群. 作者耦合分析: 一种新学科知识结构发现方法的探索性研究 [J]. 中国图书馆学报, 2012, 38(2): 4-11.
- [9] 陈卫静, 郑颖. 基于作者关键词耦合的潜在合作关系挖掘 [J]. 情报杂志, 2013, 32(5): 127-131.
- [10] 丛敏, 王景周, 陈咏梅. 学术期刊合著论文作者的科研合作特征分析 [J]. 图书情报导刊, 2020, 5(11): 66-72, 79.
- [11] Zhao Dangzhi, Strotmann A. Evolution of research activities and intellectual influences in information science 1996-2005: introducing author bibliographic coupling analysis [J]. Journal of the American Society for Information Science

and Technology,2008,59(13):2070–2086.

[12] 陈远,王菲菲.基于CSSCI的国内情报学领域作者文献耦合分析[J].情报资料工作,2011(5):6–12.

[13] 邱均平,王菲菲.基于引证关系的国内情报学领域作者研究活力与影响力分析[J].图书馆论坛,2011,31(6):51–61.

[14] 李国俊,肖明,邱小花.作者引文耦合分析可视化研究[J].图书情报工作,2012,56(12):81–84.

[15] 刘志辉,张志强.作者关键词耦合分析方法及实证研究[J].情报学报,2010,29(2):268–275.

[16] 邱均平,王菲菲.基于SNA的国内竞争情报领域作者合作关系研究[J].图书馆论坛,2010,30(6):34–40.

[17] 邱均平,陈木配.我国计量学领域作者合作关系研究[J].情报理论与实践,2012,35(11):56–60.

[18] 宋艳辉,武夷山.作者文献耦合分析与关键词耦合分析比较研究:Scientometrics实证分析[J].中国图书馆学报,2014,40(1):25–38.

[19] 丁敬达.创新知识社区内部科学交流的特征和规律——基于国家重点实验室的实证分析[J].情报学报,2011,30(10):1086–1094.

[20] 孟祥保.图书情报学交叉融合与发展——基于国外35种核心期刊的引文分析[J].图书情报知识,2012(5):50–58.

Author Similarity Detection Based on Asymmetric Perspective

Xi Chongjun Liu Zhihui Zhang Junsheng

(Institute of Scientific and Technical Information of China, Beijing 100038, China)

Abstract: [**Purpose/significance**] Aiming at the problems existing in the traditional author similarity detection, improving the author similarity algorithm is of great significance in combing the domain knowledge structure, mining the author cooperation relationship and scientific research community discovery. [**Method/process**] From the perspective of asymmetry, taking into account the author's document quantity, keyword use frequency, writing habits and other factors, an author similarity algorithm integrated with author research intensity and word relevance is proposed. Firstly, the keyword data is used for empirical analysis, and then applied to the citation data for calculation and comparative analysis. [**Results/conclusion**] The experimental results show that the author similarity algorithm integrated with research intensity from the asymmetric perspective proposed in this paper can more accurately reflect the similarity among authors and improve the accuracy of author recommendation. The integration of word relevance and the calculation based on citation data can enlarge the similarity among authors, which is helpful to mine the potential cooperative relationship of authors.

Keywords: Author coupling relationship; Research intensity; Asymmetric perspective of word relevance

(本文责编: 王秀玲)