基于模块化引文网络的国内数字人文领域知识扩散研究*

刘桂锋[1]　马明睿[1]　韩牧哲[1]　刘琼[2]
（1. 江苏大学科技信息研究所，镇江 212013；2. 江苏大学图书馆，镇江 212013）

摘要：在数字人文研究不断发展的过程中，知识扩散呈现出跨学科交互与动态演进的特征，引文网络为相关特征的识别提供了分析工具。基于国内数字人文领域发表论文的引文数据，利用Gephi软件构建引文网络进行模块化分析，基于知识净扩散量、知识扩散相对速度、知识扩散相对广度和知识扩散相对强度4个指标对模块化子网的知识扩散水平进行测度，并结合模块间引用关系揭示知识流动路径。研究识别出国内数字人文领域的5个主题模块，并且揭示各模块之间存在明显的知识流向，整体扩散过程呈现跨学科、多路径和动态化特征，展现了数字人文领域的研究格局与演进路径，对理解学科结构和推动未来发展具有参考价值。

关键词：数字人文；知识扩散；引文网络；知识流动；模块化分析
中图分类号：G250；G353.1　DOI：10.3772/j.issn.1673-2286.2025.10.006
引文格式：刘桂锋，马明睿，韩牧哲，等. 基于模块化引文网络的国内数字人文领域知识扩散研究[J]. 数字图书馆论坛，2025，21（10）：54-64.

知识扩散是知识在主体间通过传播、吸收与转化实现共享创新的过程，对揭示学科规律与优化资源配置具有重要意义。学界已围绕学科知识扩散开展诸多研究：张以迪等[1]融合正式与非正式载体产生的交流数据，构建学术论文知识扩散效果评价模型，探究影响学术论文知识扩散效果的影响因素，并提出促进学术论文知识扩散的优化策略；郑碧丽等[2]以信息资源管理学科核心期刊论文为对象，构建多层级引文网络，识别传播者、触发者等论文角色，深化了对知识扩散机制的认知。在知识扩散指标体系演进方面，相关研究持续完善：Rowlands等[3]提出期刊扩散指数（JDI）；Rousseau[4]提出知识扩散速度（KDS），后续进一步提出知识扩散广度（KDB）和知识扩散强度（KDI）[5]；邱均平等[6]基于期刊引证关系重构知识扩散广度与知识扩散强度的计算逻辑，解决了不同规模研究对象间可比性不足的问题；韩牧哲等[7]基于模块化引文网络提出知识净扩散量和相对速度指标。尽管当前已涌现出多种计量指标，但现有指标（如速度）侧重效率，难以全面反映交叉学科知识传播的广度（跨领域范围）和强度（外溢比例）。

数字人文作为融合人文学科理论与数字技术方法的新兴交叉领域，通过计算手段挖掘人文数据价值、重构研究范式，2022年被确立为本科专业[8]，2024年明确为信息资源管理二级学科[9]，形成贯通式学科格局。作为多学科交叉领域，数字人文的知识扩散呈不规则的非线性特征，这对知识获取、创新等至关重要。

因此，本文选取数字人文领域，研究其知识扩散水平，通过构建模块化引文网络，克服传统单一指标难以刻画交叉学科复杂知识流动的局限性，从范围与效能两个维度更全面地揭示学科内在结构与演进路径，为推动国内数字人文的均衡发展提供量化支撑与决策参考。

收稿日期：2025-09-14
*本研究得到国家社会科学基金青年项目“出土文献的多模态知识组织与融合研究”（编号：23CTQ038）资助。

1　理论基础与测度方法

1.1　引文网络模块化

模块化网络理论源于复杂系统科学对网络社区结构的研究，Newman[10]提出的模块度Q是量化网络社区划分效果的核心指标，Q的计算方法如式（1）所示，其核心逻辑是对比网络实际社区内部边数与随机网络中相同规模社区的期望边数，Q值越大，社区划分越清晰，内部连接的聚集性越强。

$$Q=\frac{1}{2m}\sum_{i,j}\left(A_{ij}-\frac{k_ik_j}{2m}\right)\cdot\delta\left(c_i,c_j\right) \tag{1}$$

式中：m为网络总边数；A为网络邻接矩阵，A_{ij}值为1代表存在从节点i指向节点j的边，反之A_{ij}值为0；k_i、k_j分别为节点i、j的度数（节点的入度与出度之和）；c_i为节点i所在的社区，$\delta\left(c_i,c_j\right)$=1时两节点属同一模块，值为0则相反。当$Q$值为0.3~0.7时，网络社区结构具有统计学意义，划分结果可有效反映领域内的主题聚类[10-11]。

引文网络是一种天然适合观察知识扩散的自组织网络[12]，节点代表文献，有向边代表引用关系，知识沿着边的方向从被引文献流向施引文献。对引文网络进行模块化分析，可识别出内部连接紧密、外部连接稀疏的节点子集（模块）[7]，每个模块通常对应一个特定的研究主题或知识子领域，便于将庞大的学科文献体系解构为若干主题模块，揭示学科内部知识结构。

本文借鉴王亮[13]的研究思路，在传统引文网络基础上定义引用延时并赋予边权重，量化知识传播的时间维度差异，借助Gephi 0.9.3软件构建网络并利用其内置社区探测算法完成模块化划分。同时，在邱均平等[6]研究基础上，针对模块化网络特点改进并新增两个指标：①知识扩散相对广度，衡量单位年份知识传播范围，体现影响的持久性与覆盖面；②知识扩散相对强度，衡量模块对外输出知识占总产出的比例，反映知识生产效率。结合知识净扩散量和知识扩散相对速度，可从范围与效能双维度全面刻画数字人文学科的知识扩散图景。

1.2　知识净扩散量

知识溢出量指的是某一领域被其他领域论文引用的论文数量，知识吸纳量为该领域引用其他领域论文的数量，知识净扩散量用于衡量某一领域知识的溢出量与吸纳量的差值，反映其发展成熟度，记为Q_M（M与模块属性关联，以区分各领域的知识净扩散量指标）[7]。该指标仅统计模块间的知识流动，不计模块内部引用，其基本关系如式（2）所示。

$$Q_M=Q_M^{\text{out}}-Q_M^{\text{in}} \tag{2}$$

式中：Q_M^{out}代表模块知识的溢出量；Q_M^{in}代表模块对外部知识的吸纳量。

为适应模块化引文网络的特点并增强指标的可比性，本研究对该指标进一步具体定义：知识净扩散量Q_M是刻画模块知识流动效率差异的核心指标，定义为模块溢出效率与吸纳效率的差值。为消除不同模块规模差异带来的影响，同时考虑到知识扩散的效率与引用延时（边权重ω）成反比，需要对指标进行标准化处理：将溢出与吸纳的节点数分别除以其所在模块的论文总数N_M，乘以相应边权重均值的倒数（抵消时间成本影响），再乘以放大系数100以便于观察分析，Q_M具体计算如式（3）所示。

$$Q_M=\left(\frac{N_M^{\text{out}\prime}\cdot L_M^{\text{out}\prime}}{N_M\cdot\sum\omega^{\text{out}\prime}}-\frac{N_M^{\text{in}\prime}\cdot L_M^{\text{in}\prime}}{N_M\cdot\sum\omega^{\text{in}\prime}}\right)\times 100 \tag{3}$$

式中：$N_M^{\text{out}\prime}$与$N_M^{\text{in}\prime}$分别代表模块内被其他模块引用的论文数量（溢出节点数）和引用外部模块论文的论文数量（吸纳节点数）；$L_M^{\text{out}\prime}$与$L_M^{\text{in}\prime}$分别代表模块向外溢出的引用边数和从外部吸纳的引用边数；$\sum\omega^{\text{out}\prime}$与$\sum\omega^{\text{in}\prime}$分别代表模块溢出边权重和与吸纳边权重和，即相应引用延时的总和。式（3）前项为溢出效率，由溢出节点率（被其他模块引用的节点比例）与平均溢出延时（知识溢出的时间成本）的比值构成；后项为吸纳效率（模块吸收外部知识的效率），计算逻辑同理。

需特别说明，本研究定义的Q_M是衡量知识流动效率的相对指标。数量维度上，模块间溢出边总数必然等于吸纳边总数，但知识流动的效率则不受此约束。后文Q_M全为负值的结果，揭示了特定发展阶段下知识吸纳效率普遍高于溢出效率的学科特征。

1.3　知识扩散相对速度

知识扩散速度指单位时间内某领域的知识扩散量[4]，知识扩散速度S_M的计算方法如式（4）所示。

$$S_M=\frac{L_M^{\text{out}2}}{N_M^{\text{out}}\cdot\sum\omega^{\text{out}}} \tag{4}$$

式中：L_M^{out} 为模块总扩散量（内部边数与溢出边数之和）；N_M^{out} 为模块内出度不小于1的节点数量；$\sum\omega^{\text{out}}$ 为总扩散引用延时总和。式（4）由单位时间知识扩散量（总扩散量与有引用行为节点数的比值）与平均延时（总扩散引用延时与总扩散量的比值）的比值化简所得，用以衡量单位时间内知识的传播效率。

为更直观地进行模块间比较，需要将绝对速度转换为知识扩散相对速度S_M^{R}。该指数以整体网络速度S为基准，通过模块的知识扩散速度与整体网络速度的比值减1得到速度差，再乘以放大系数10以突出差异。知识扩散相对速度S_M^{R}计算方法如式（5）所示。当$S_M^{\text{R}}>0$时，模块的知识扩散效率高于整体水平；当$S_M^{\text{R}}<0$时，其效率低于整体水平。

$$S_M^{\text{R}}=\left(\frac{S_M}{S}-1\right)\times 10 \tag{5}$$

1.4 知识扩散相对广度

在邱均平等[6]提出的知识扩散广度指标基础上，结合模块化网络分析需求进行改进，提出知识扩散相对广度B_M^{R}，将其定义为模块的总知识扩散量与其文献发表时间跨度之比，以衡量单位年份内的知识传播范围。知识扩散相对广度B_M^{R}的计算方法如式（6）所示。

$$B_M^{\text{R}}=\frac{L_M+L_M^{\text{out}\prime}}{Y_M^{\text{pub}}} \tag{6}$$

式中：L_M为模块内部引用边数；$L_M+L_M^{\text{out}\prime}$ 为模块总知识扩散量；Y_M^{pub}为模块统计年份数（最晚与最早论文发表年之差）。

1.5 知识扩散相对强度

邱均平等[6]将知识扩散强度定义为被引学科数与总被引频次的比值。为适配模块化网络，提出知识扩散相对强度I_M^{R}，将其定义为模块对外知识溢出量占其总知识扩散量的比例，聚焦知识输出的结构性特征，衡量模块产出的知识中突破自身边界、贡献给其他主题模块的知识比例，以此评估模块作为知识生产者的外溢效能。知识扩散相对强度I_M^{R}的计算方法如式（7）所示。

$$I_M^{\text{R}}=\frac{L_M^{\text{out}\prime}}{L_M+L_M^{\text{out}\prime}} \tag{7}$$

2 引文网络构建与模块化

2.1 数据采集与清洗

在中国知网（CNKI）中文期刊数据库中以SU%=（‘人文计算’or‘数字人文’）为检索式进行检索，为保证年度数据的完整性、全面性与时效性，文献发表日期的上限统一为2024年12月31日，检索工作分阶段于2024年12月至2025年7月完成，最终检索日期为2025年7月22日。为保证数据质量与权威性，初始文献来源为《中文核心期刊要目总览》、中文社会科学引文索引（CSSCI）、中国科学引文数据库（CSCD）和《中国人文社会科学期刊AMI综合评价报告》收录期刊，共检索到期刊论文1 862篇；为扩大数据覆盖面，补充检索CNKI中相关硕博学位论文399篇，及《数字人文》《数字人文研究》两种专业期刊自创刊至2024年12月31日的文献343篇，初始样本总量达2 604篇。

为保证数据的有效性，数据清洗采用自动化脚本初筛和人工深度核对的组合流程，分三步完成：①初步筛选，剔除会议纪要、新闻报道、书评、刊讯等非研究性文献385篇，自动匹配准确率约为95%，歧义记录均通过人工核查原文予以修正；②人工复核，逐一排查题录信息，剔除与研究主题明显不相关的文献172篇；③数据去重，合并同一文献在不同来源的重复记录131篇。对解析歧义严重或元数据残缺的复杂记录，进行第二轮深度人工复核与原文追溯，最终得到有效文献1 916篇，构成引文网络构建的核心数据样本。

2.2 引文网络构建与模块化

通过CNKI数据库对采集到的1 916篇文献的被引文献和施引文献及相关论文的题录信息进行收集（含期刊名称、标题、发表年份等字段），为便于数据处理和网络构建，对原始文献集的每一篇文献进行唯一编码（规则为“年份+JA+时间序号”，如2005JA001），并以编码替换原文献标题。构建网络时，仅保留1 916篇文献之间的相互引用关系，将编码替换后的数据集导入Gephi 0.9.3软件，剔除孤立节点，最终构建出一个包

含1 635个节点（论文）的引文网络，包括有效被引记录6 833条、有效施引记录7 429条。

采用软件内置的Fruchterman Reingold算法对引文网络进行球形布局，设置区值为2 000、重力为25.0、速度为1.0，以入度为节点参数对节点大小和标签文字尺寸进行排序，节点及标签文字大小与论文被引频次呈正相关。运用社区探测算法进行模块化统计，分辨率设定为1.0且不考虑边权重时，网络模块度为0.478，共识别出10个模块。通过分辨率参数敏感性测试验证，在分辨率设定为0.9与1.1时，网络核心的五大主题模块在成员构成与规模上保持高度稳定，模块度Q值稳定在0.47左右，处于Newman[10]定义的社区结构显著区间（0.3~0.7），证明划分结果具有较强的鲁棒性，Louvain算法适用性良好[11]。

为确保主题模块的代表性与分析效度，对初始模块进行了人工审查与优化合并：剔除无法构成有效研究主题的极小模块，将主题相近且规模较小的模块，依据其引用关系和核心语义归并到主题更明确的较大模块中，最终形成国内数字人文领域引文网络模块化图（见图1）。图1中包括5个主题明确、结构稳健的模块化子网（编号为A、B、C、D、E），涵盖1 574个节点，网络内部总边数为7 260条（模块内部边4 840条，模块间引用边2 420条）。结合所涵盖论文的研究主题分别对这5个模块进行命名，命名及参数如表1所示。

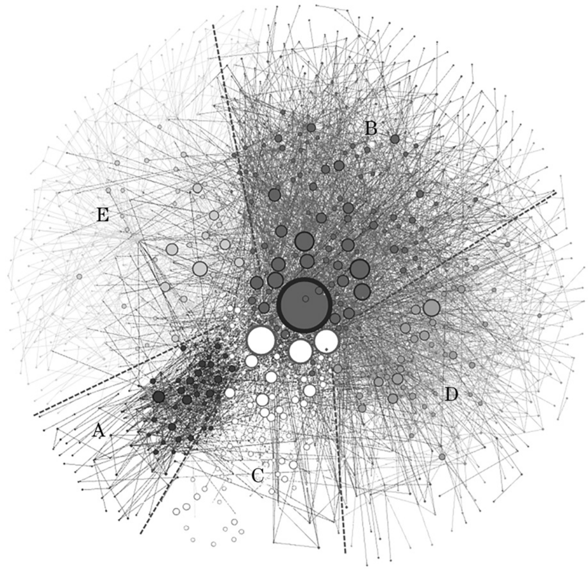

图1　数字人文领域引文网络模块化图

表1　数字人文领域引文网络模块命名及参数

模块编号	网络名称	节点数/个	边数/条
A	数字人文教育	149	370
B	数字人文领域的技术方法	532	1 740
C	数字人文与信息资源管理	256	1 139
D	数字人文档案	288	943
E	数字人文方法的跨学科应用	349	648

2.3　引文网络模块分析

为揭示各模块具体研究情况，对引文网络模块化得到的5个主题模块逐一进行详细阐述，包括时序发展、研究热点以及典型案例，为深入理解国内数字人文领域的学科结构与发展脉络提供微观视角。

（1）模块A：数字人文教育。模块A围绕数字人文教育实践、研究生培养模式与课程体系建设等维度，为构建中国数字人文教育体系提供理论与实践基础。该模块论文发表与引用情况如图2所示，呈现前期蓄力、后期影响力爆发的发展路径。2014—2015年的起步阶段研究成果尚少，2024年相关论文数量增至28篇，体现出学界关注度的持续上升。该模块影响力分别在2019年和2022年出现高峰。其中，2022年年度总引用次数达到240次，年度被引次数为151次，且被引次数明显高于同期的施引次数。

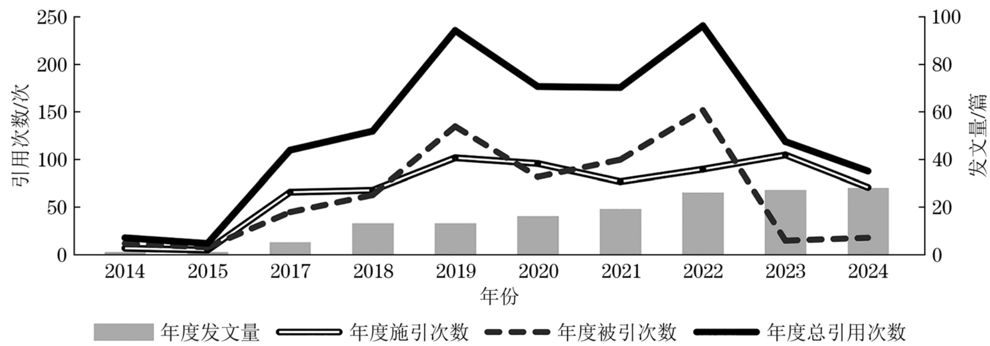

图2　模块A论文发表与引用情况

结合模块A高被引文献的内容分析发现，学界逐渐从单纯的经验借鉴开始转向更深层次的理论建构与本土化实践探索，早期以引介和借鉴英美等国成熟的、分层分类的人才培养模式为主[14-15]，后来出现对“计算X”教育范式的探讨[16-17]和“理论-方法-实践”三段式教育路径的验证[18]，产出了一批质量较高并受到广泛认可的研究成果，标志着该领域从知识输入阶段迈向能够有效进行教育赋能的较成熟阶段。

（2）模块B：数字人文领域的技术方法。模块B专注数字人文技术方法，包括文本数字化、语义标注、大规模数据挖掘与可视化等。该模块论文发表与引用情况如图3所示，呈现出起步缓慢、快速爆发的发展路径。2012—2015年为研究起步期，发文量稀少，自2016年起进入快速上升期，并在2017—2019年迎来关键的理论与技术突破。其学术影响力于2021年达到顶峰，该年度总引用次数达1 360次。

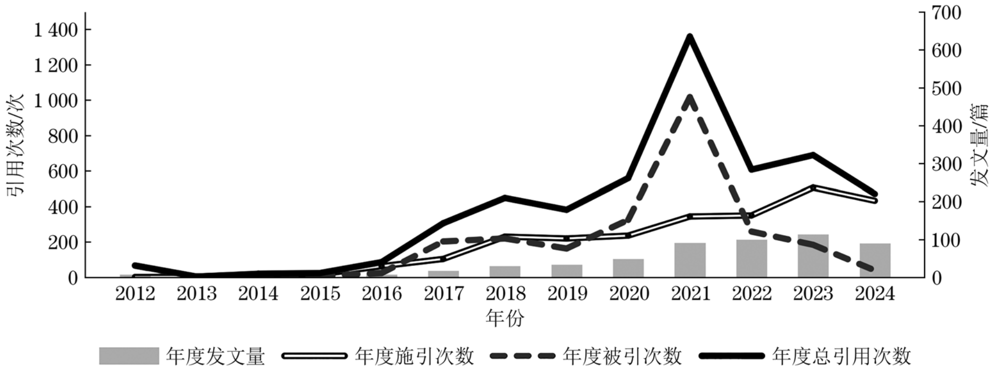

图3　模块B论文发表与引用情况

结合核心研究的时序推进轨迹，模块B的研究内容沿数字化、数据化、智慧化的清晰演进路径，涵盖从资源数字化、语义处理到知识挖掘的完整技术栈[19-22]。2017年刘炜等[19]对技术体系与理论结构的探讨，2018年曾蕾等[23]对智慧数据角色的分析，以及2019年陈涛等[20]对知识图谱应用的总结，共同构建了数字人文技术方法的理论基础。数字人文方法体系正从基础设施建设阶段向认知范式升级阶段推进，轻量语义栈、分布式知识图谱与多模态AI将推动数字人文向情境驱动的生成式人文科学发展。

（3）模块C：数字人文与信息资源管理。模块C聚焦数字人文的学科属性、理论范式与学术伦理研究，核心探讨信息资源管理与数字人文的交叉融合路径。该模块论文发表与引用情况如图4所示，呈现出核心议题热度爆发式增长后趋于稳定的发展路径。2005—2015年为早期探索期，发文量极少，2015年后进入快速发展期，2018年和2021年达到阶段性峰值40篇。2016—2019年为影响力高峰期，年均被引次数超过300次，2016年被引397次，反映出数字人文学科定位与方法论基础问题在该阶段受到学界持续关注。2018年后被引次数下降，表明该模块热度趋于稳定。

模块C影响力高峰期涌现了一批奠基性研究。柯平等[24]在2016年对演化路径与热点领域的分析为该阶段研究拉开了序幕，随后沈振萍等[25]在2017年提出三位一体融合模式，邓君等[26]在2019年的计量研究进一步用数据证实相关研究发展趋势，共同为信息资源

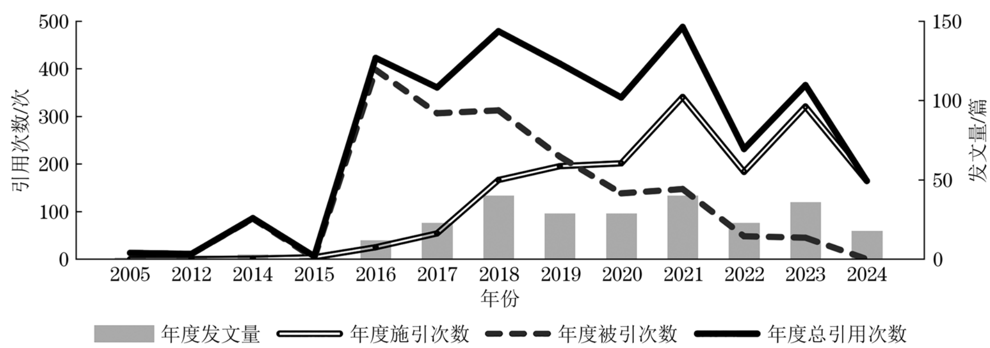

图4　模块C论文发表与引用情况

管理在数字人文中的角色定位与双向赋能路径提供了理论与数据支撑。2019年之后，随着学科定位逐渐明晰，引用热度趋于稳定，研究方向开始向更具体的应用深化，如基于实践项目的中国数字人文学术体系构建[27]，以及对生成式AI通过数字人文项目渗透至信息资源管理主题领域、形成双向赋能的学科演进路径的分析[28]。信息资源管理既是数字人文数据与方法论的基石，也促使数字人文的计算转向与跨学科合作，孕育出新的学科生长点，从而推动数智时代知识生态系统构建。

（4）模块D：数字人文档案。模块D主要关注数字人文视角下档案资源的数字化整理、开发与传播，以及档案学与数字人文的学科融合趋势。该模块论文发表与引用情况如图5所示，研究热度呈现持续增长态势。2015年仅有2篇研究案例，2017年后发文量稳步增长，2021年跃升至57篇，2024年达到70篇的阶段性高峰。模块D影响力在2021年达到峰值，年度被引418次，总引用次数为717次。2022—2024年被引次数虽略有回落，但施引次数持续增长，2023年施引次数达380次，由此可以看出该模块研究持续活跃。

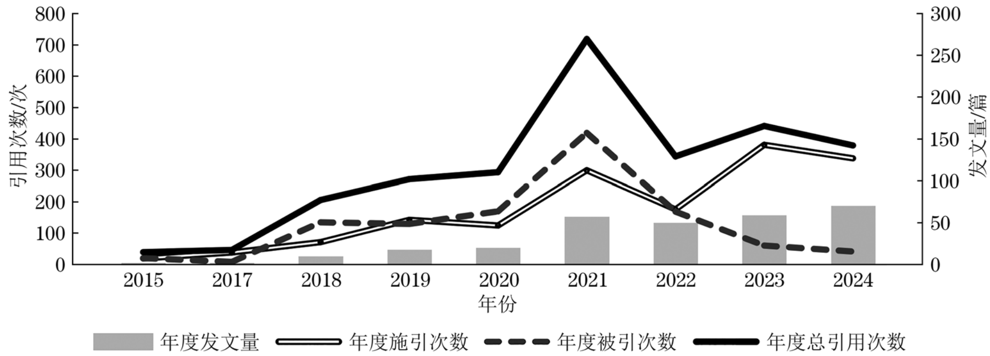

图5　模块D论文发表与引用情况

模块D影响力峰值的出现，与该模块在理论方法上呈现体系化演进密切相关。正是在2021年，牛力等[29]提出以发现、重构、故事化为主线的知识服务链条，为数字人文档案研究构建了体系化的理论路径。后续的代表性研究，无论是对红色档案数字化叙事的实践探索[30]，还是对文化遗产数字叙事信任模型的构建[31]，都可以看作是在这一理论框架下的深化与拓展。这些研究共同推动数字人文理念在档案领域的落地实施，并为档案工作向社会化服务转型提供了理论基础与实践路径。

（5）模块E：数字人文方法的跨学科应用。模块E专注于数字人文方法在其他学科中的多维应用，包括国家级数字人文平台建设、数据标准体系的构建、跨机构协作机制的建立等。该模块论文发表与引用情况如图6所示，研究呈现出持续活跃、逐步提速的发展路径。2014年仅有1篇研究案例，2016年后发文量逐步增长，2020年起增速加快，2023年发文量高达75篇。早期阶段施引次数主导整体引用表现，自2018年起被引次数连续5年超过施引次数，2020年该模块研究成果被引用319次，为历年最高，印证了该模块研究在数字人文领域从早期的知识借鉴阶段，过渡到能够产出高影响力原创成果的成熟阶段。

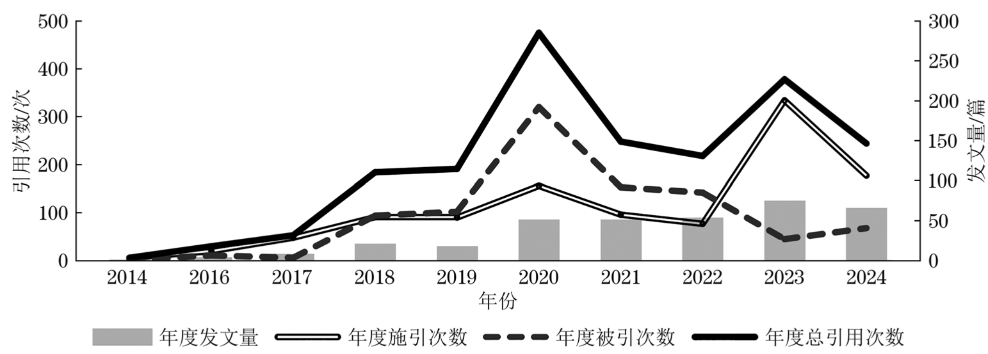

图6　模块E论文发表与引用情况

2020年模块E影响力峰值时期涌现出了一系列高质量范式转型研究。例如，胡开宝等[32]与王兆鹏等[33]的代表性研究分别将语料库技术和地理信息系统（GIS）方法深度融入翻译学与古代文学研究，并催生出新的研究范式。相较于陈刚[34]在2014年对历史地理信息化的早期探索，近期研究正在经历从数据建模到界面批判的跃迁[35]，技术元素已不再仅作为工具存在，而是成为研究对象本身的一部分，数字人文从技术借用阶段迈向了理论、方法、技术深度融合的系统性重构阶段。

3 知识扩散水平的计量与分析

3.1 知识扩散指标测算结果

（1）知识净扩散量。根据式（2）和式（3）对5个模块的知识净扩散量进行计算，结果如表2所示。

表2 知识净扩散量及相关参数计算结果

模块编号	N_M	$N_M^{out\prime}$	$L_M^{out\prime}$	$\sum\omega^{out\prime}$	$N_M^{in\prime}$	$L_M^{in\prime}$	$\sum\omega^{in\prime}$	Q_M
A	149	75	307	844	173	297	880	−20.88
B	532	203	881	2 732	267	633	1 931	−4.14
C	256	143	636	2 426	211	510	1 383	−15.75
D	288	105	247	562	233	537	1 703	−9.39
E	349	91	349	961	247	443	1 628	−9.79

5个模块的知识净扩散量均为负值，并非源于计算偏差，而是由本研究效率差非守恒的指标定义，与国内数字人文的“技术引进-本土化适配”发展阶段特征共同导致的，反映出国内数字人文领域当前仍处于知识吸纳效率高于溢出效率的发展阶段。其中：模块A的知识净扩散量（−20.88）最低，是最显著的知识“净消费者”，对外部知识的吸纳需求远高于自身输出；模块B的知识净扩散量（−4.14）最接近平衡，是网络中的核心知识生产者，其溢出效率与吸纳效率的差距最小；模块C、D、E的知识净扩散量处于中间水平，均呈现适度吸纳、缓慢溢出的特征，符合交叉学科从早期积累到转化的发展规律。

各模块需要通过互动获取互补知识（如模块B向模块D传递数字化技术，模块C向模块A输出数据治理经验），且需要结合自身场景适配吸纳的知识（如模块D将模块B的通用技术改造为档案专用处理工具，模块A将模块C的治理经验融入课程设计）。模块吸纳其他模块的现成知识无需额外适配成本，效率较高，而模块向其他模块溢出知识前需要先完成场景化改造，消耗更多时间成本，效率较低，这种吸纳快、溢出慢的特征普遍存在，最终使所有模块的吸纳效率均高于溢出效率。以模块B为例，其向外溢出的引用边数与从外部吸纳的引用边数的净差值为248，是全网络最大知识输出方，但因溢出平均延时（3.10年）略长于吸纳平均延时（3.05年），溢出效率（12.32%）低于吸纳效率（16.46%），知识净扩散量为−4.14。这正是新兴学科“打基础”阶段的量化体现，即优先通过内化外部知识构建自身基础，再逐步实现原创输出，与毋江波等[36]在知识生态视角下所观察到的新兴领域吸收强度大于溢出强度的规律相印证。

（2）知识扩散相对速度。根据式（4）和式（5）测算整体网络（编号为S）相关参数及5个模块的知识扩散相对速度，结果如表3所示。

整体网络的知识扩散速度为2.368，为计算各模块相对速度的参照依据。相对速度以0值为基准，其为负值表示该模块的知识扩散效率低于整体网络平均水

表3 知识扩散相对速度及相关参数计算结果

模块编号	N_M^{out}	L_M	$L_M^{out\prime}$	L_M^{out}	$\sum\omega^{out}$	S_M	S_M^R
A	100	370	307	677	1 786	2.566	0.84
B	332	1 740	881	2 621	7 923	2.612	1.03
C	204	1 139	636	1 775	5 742	2.690	1.36
D	182	943	247	1 190	3 129	2.487	0.50
E	208	648	349	997	3 118	1.533	−3.53
S	1 026	4 840	2 420	7 260	21 698	2.368	

平。模块A、B、C、D的相对速度均为正值，说明其知识扩散效率高于整体水平，仅模块E的相对速度为负值（−3.53），且该模块的知识扩散速度（1.533）显著低于其他模块，这与跨学科应用的知识转化周期长、适配成本高密切相关。模块C的相对速度（1.36）最高，体现出其内部知识迭代速度快的特征；模块B的相对速度（1.03）次之，印证了技术方法作为核心引擎的高效传播属性。

（3）知识扩散相对广度。根据式（6）测算整体网络的相关参数及各模块的知识扩散相对广度，结果如表4所示。

知识扩散相对广度衡量单位年份内的知识传播范围。模块B的相对广度（218.42）显著高于其他模块，说明其知识传播的覆盖效率最高，影响力的持久性最高、覆盖面最广，这与其技术方法的基础性、普适性密切相关；模块D的相对广度（132.22）次之，反映出其虽以内部循环为主，但单位时间内的传播规模仍较大；模块A的相对广度（67.70）最低，与其研究聚焦教育领域、应用场景相对集中有关。

（4）知识扩散相对强度。根据式（7）测算整体网络相关参数和5个模块的知识扩散相对强度，结果如表5所示。

知识扩散相对强度反映模块知识外溢的比例。模块A的相对强度（0.45）最高，说明其45%的知识产出能够突破自身边界，对外赋能其他模块，体现出高效转化的特征；模块D的相对强度（0.21）最低，仅21%的知识实现外部溢出，印证了其内部循环的属性；模块B、C、E的相对强度处于0.34~0.36，属于中等外溢水平，其中模块B虽外溢比例不及模块A，但因其总扩散量最大，仍是对外输出知识最多的核心模块。

表4　知识扩散相对广度及相关参数计算结果

模块编号	N_M	L_M	L_M^{out}	Y_M^{pub}	B_M^{R}
A	149	370	677	10	67.70
B	532	1 740	2 621	12	218.42
C	256	1 139	1 775	19	93.42
D	288	943	1 190	9	132.22
E	349	648	997	10	99.70
S	1 574	4 840	7 260		

表5　知识扩散相对强度及相关参数计算结果

模块编号	N_M	L_M	$L_M^{\mathrm{out}\prime}$	L_M^{out}	I_M^{R}
A	149	370	307	677	0.45
B	532	1 740	881	2 621	0.34
C	256	1 139	636	1 775	0.36
D	288	943	247	1 190	0.21
E	349	648	349	997	0.35
S	1 574	4 840	2 420	7 260	

3.2　模块间知识引用关系

上述指标从不同维度刻画了各模块的知识扩散特征。为进一步揭示模块间知识流动的具体路径与相互依赖关系，统计了模块间的引用数据，详见表6。表6展示了任意两个模块之间的双向知识流动情况，包括吸纳节点数、溢出节点数、吸纳边数和溢出边数。由表6汇总可知，模块吸纳边总数与溢出边总数均等于模块间引用边数（2 420条），遵循封闭网络的数量守恒定律。需要说明的是，表6中吸纳节点与溢出节点的并集总数为1 156个，与网络总节点数（1 574个）的差异在于有418个节点仅在模块内部存在引用关系，而未参与模块间的知识流动。结合表6数据，对模块间的知识互动模式进行深入分析。

从引用流向来看，技术方法（模块B）与信息资源管理（模块C）构成了网络中互动最频繁的交互轴心。模块C从模块B吸纳302条引用边，主要集中于文本挖掘、图像识别及可视化等前沿工具的应用；与此同时，模块C向模块B反向溢出237条边，这些溢出主要体现在文献语义组织规范、知识图谱架构等方法论层面，为

表6　模块间引用数据

模块编号		吸纳节点数/个	溢出节点数/个	吸纳边数/条	溢出边数/条
A	B	62	46	106	87
	C	66	43	125	88
	D	25	31	32	64
	E	25	44	34	68
B	A	46	62	87	106
	C	80	102	237	302
	D	79	111	138	270
	E	67	101	171	203
C	A	43	66	88	125
	B	102	80	302	237
	D	33	56	43	136
	E	38	74	77	138
D	A	31	25	64	32
	B	111	79	270	138
	C	56	33	136	43
	E	40	33	67	34
E	A	44	25	68	34
	B	101	67	203	171
	C	74	38	138	77
	D	33	40	34	67

模块B的技术工具提供了人文场景适配的反馈与完善方向。这种高强度的双向知识流动证明了我国数字人文发展并非单向的技术驱动，而是呈现技术开发与管理理论之间的双向赋能与协同演进。

相比之下，数字人文档案（模块D）和跨学科应用（模块E）表现出对核心驱动模块显著的单向知识依赖。模块D分别从模块B、C吸纳了270条和136条引用边，反映出档案学领域在应对海量数字资源整理、长期保存及开发利用时，高度依赖外部成熟的方法论支撑；然而，其向外部模块的溢出边数普遍较低，这种不对称的知识流动特征反映出档案领域目前仍处于深度的场景内化与技术吸收阶段。跨学科应用（模块E）对模块B（吸纳203条边）和模块C（吸纳138条边）的强烈需求，也侧面印证在推动传统人文学科范式重构时，需要承担高昂的外部知识引入成本。

此外，数字人文教育（模块A）的知识流向呈现出独特的反哺特征。虽然从知识净扩散量（−20.88）来看，其仍属于网络中的知识吸纳方，但模块A分别向模块B、C溢出87条边和88条边，说明我国数字人文教育在实践探索中所沉淀的课程模型、人才培养范式以及教学实验结论，已经开始反向影响技术工具的迭代与管理理论的修正。跨学科应用（模块E）向模块A仅溢出34条边，揭示了跨学科融合成果转化为通用的教学资源仍需较长的转化周期。这些微观互动数据与3.1节的指标测算结果形成互证，印证了各主题领域的角色分工，为核心集聚、应用分散的整体格局提供实证支撑。

4　结论与展望

本研究通过构建引文网络并开展模块化分析，识别并系统刻画了国内数字人文领域5个核心主题的知识扩散特征。在此基础上，通过数量（涵盖扩散总量与相对广度）以及效率（涵盖净扩散量、相对速度与相对强度）维度，系统整合测度指标，弥补了单一计量指标在交叉学科研究中的局限性。分析结果显示，各模块知识净扩散量均为负值，这并非数字人文领域独有的特征，而是新兴交叉学科早期快速积累期的共性信号。此类学科的子模块需要优先通过高效吸纳同一学科内其他子模块的互补知识构建基础，再结合本土需求完成知识转化与原创输出。这种吸纳效率暂时高于输出效率的特征，符合新兴交叉学科从内部子领域知识整合到跨领域原创引领的正常发展规律[37]。借鉴叶佳鑫等[38]对知识流动的阶段划分，内部子领域知识整合即学科早期致力于吸收外部知识（知识流入），并进行本土化改造以构建内部范式（概念创新）的阶段[37]；跨领域原创引领指学科发展成熟后，其形成的独特理论或方法开始反向溢出（知识反流）[38]，对其他学科产生引领性影响，这种不同学科交叉点上产生的知识融合与创新，被认为是形成新的科学前沿和实现重大科学突破的关键驱动力[39]。

基于上述学科阶段的判定，进一步观察发现我国数字人文领域形成核心驱动、中介传导、应用吸纳的三级结构体系，呈现出技术与方法论支撑广泛辐射、应用与跨域融合多元转化的非均衡发展格局。技术方法与信息资源管理构成了知识生产与扩散的核心枢纽，前者作为核心引擎提供底层支撑，后者则发挥枢纽桥梁作用，二者通过高频双向引用实现技术演进与理论完善的协同发展，且在当前阶段展现出较高的知识迭代效率。作为中介传导层的数字人文教育在外围扮演着高效知识转化器的角色，其外溢强度居全领域之首，能将实践中沉淀的教学经验迅速反哺给技术与管理领域。处于应用层面的数字人文档案与跨学科应用则分别承担着场景内化者与深度整合者的角色，前者侧重于内部的知

识吸收与场景迭代，后者则因跨学科深度融合中高昂的转化成本而表现出较长的知识流转周期。这一结构特征不仅印证了核心层集聚与应用层分散的现状，也为未来缩短知识的“吸纳-转化-溢出”周期提供了实证依据。

为推动我国数字人文健康均衡发展，基于上述发现提出以下针对性实践建议。在政策与资助层面，鉴于跨学科应用的缓慢特性，设立长周期资助项目，鼓励“慢科学”；在学科建设层面，加强模块间的对话机制，如举办专题研讨会，鼓励档案与跨学科应用的研究者合作，推动档案知识“破圈”；在研究者层面，技术方法研究者结合应用领域，开发低门槛、高可用的工具，应用领域研究者凝练可迁移方法论，缩短知识的流转周期。后续研究可在广度上通过与国际数字人文领域开展横向比较，揭示国内外知识扩散的共性与差异，在深度上结合自然语言处理技术对引文数据的引用动机进行细粒度划分，以克服目前大规模自动化语义标注工具不完善的局限，更全面地评估知识扩散效能。

参考文献

[1] 张以迪，魏绪秋. 学术论文知识扩散效果评价与影响因素研究：以生物芯片领域的高被引论文为例[J]. 知识管理论坛（网络版），2025，10（3）：241-255.

[2] 郑碧丽，侯剑华. 引文网络中的角色识别及其对知识扩散的影响研究：以信息资源管理学科为例[J]. 现代情报，2025，45（9）：130-139.

[3] ROWLANDS I，NICHOLAS D，HUNTINGTON P，et al. Journal diffusion factors：a new approach to measuring research influence[J]. Aslib Proceedings，2002，54（2）：77-84.

[4] ROUSSEAU R. Robert Fairthorne and the empirical power laws[J]. Journal of Documentation，2005，61（2）：194-202.

[5] LIU Y X，ROUSSEAU R. Knowledge diffusion through publications and citations：a case study using ESI-fields as unit of diffusion[J]. Journal of the American Society for Information Science and Technology，2010，61（2）：340-351.

[6] 邱均平，瞿辉，罗力. 基于期刊引证关系的学科知识扩散计量研究：以我国“图书馆、情报、档案学”为例[J]. 情报科学，2012，30（4）：481-485，491.

[7] 韩牧哲，李秀霞，张艺蔓. 基于模块化引文网络的网络计量学知识扩散水平测度[J]. 情报杂志，2015，34（7）：189-194，178.

[8] 教育部. 教育部关于公布2022年度普通高等学校本科专业备案和审批结果的通知[EB/OL]. [2025-02-15]. https://www.gov.cn/zhengce/zhengceku/2023-04/20/content_5752327.htm.

[9] 中国学位与研究生教育学会官网. 研究生教育学科专业简介及其学位基本要求[EB/OL]. [2025-09-12]. https://www.acge.org.cn/encyclopediaFront/enterEncyclopediaIndex.

[10] NEWMAN M E J. Modularity and community structure in networks[J]. Proceedings of the National Academy of Sciences of the United States of America，2006，103（23）：8577-8582.

[11] 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（10）：P10008.

[12] 邱均平，李小涛. 基于引文网络挖掘和时序分析的知识扩散研究[J]. 情报理论与实践，2014，37（7）：5-10.

[13] 王亮. 基于SCI引文网络的知识扩散研究[D]. 哈尔滨：哈尔滨工业大学，2014.

[14] 徐孝娟，侯莹，赵宇翔，等. 国外数字人文课程透视：兼议我国数字人文课程设置及人才培养[J]. 图书馆论坛，2018，38（7）：1-11.

[15] 吴加琪，董梅香，赵子菲. 国外数字人文专业研究生教育调查[J]. 图书馆论坛，2018，38（6）：42-48.

[16] 黄水清，刘浏，王东波. 计算人文教育的回顾和探讨[J]. 情报理论与实践，2024，47（7）：1-9.

[17] 景晓悦. 全球化视野下的数字人文教育：来自欧美的实践启示[J]. 山西档案，2024（5）：100-102.

[18] 王涛. 数字人文的本科教育实践：总结与反思[J]. 图书馆论坛，2018，38（6）：37-41.

[19] 刘炜，叶鹰. 数字人文的技术体系与理论结构探讨[J]. 中国图书馆学报，2017，43（5）：32-41.

[20] 陈涛，刘炜，单蓉蓉，等. 知识图谱在数字人文中的应用研究[J]. 中国图书馆学报，2019，45（6）：34-49.

[21] 欧阳剑. 面向数字人文研究的大规模古籍文本可视化分析与挖掘[J]. 中国图书馆学报，2016，42（2）：66-80.

[22] 李明杰，郑翔，关思雨. 数字人文在古籍整理领域的应用路径及发展策略[J]. 中国图书馆学报，2024，50（4）：21-38.

[23] 曾蕾，王晓光，范炜. 图档博领域的智慧数据及其在数字人文研究中的角色[J]. 中国图书馆学报，2018，44（1）：17-34.

[24] 柯平，宫平. 数字人文研究演化路径与热点领域分析[J]. 中国图书馆学报，2016，42（6）：13-30.

[25] 沈振萍，黄水清. 我国数字人文研究脉络及其在图书馆学情报学领域的典型应用[J]. 大学图书馆学报，2017，35（6）：56-68.

[26] 邓君，宋先智，钟楚依. 我国数字人文领域研究热点及前沿探析[J]. 现代情报，2019，39（10）：154-164.

[27] 夏翠娟，祁天娇，徐碧姗．中国数字人文学术体系构建考察：基于实践项目的内容分析和文献研究[J]．数字人文研究，2023，3（4）：3-36.

[28] 张海，段荟，王东波．信息资源管理领域生成式人工智能：研究现状、研究主题及研究展望[J]．情报杂志，2024，43（10）：150-156.

[29] 牛力，高晨翔，张宇锋，等. 发现、重构与故事化：数字人文视角下档案研究的路径与方法[J]. 中国图书馆学报，2021，47（1）：88-107.

[30] 陈艳红，陈晶晶．数字人文视域下档案馆红色档案资源开发的时代价值与路径选择[J]. 档案学研究，2022（3）：68-75.

[31] 王晓光，赵珂. 文化遗产数字叙事信任模型：概念与框架[J]. 中国图书馆学报，2024，50（2）：30-41.

[32] 胡开宝，黑黟. 数字人文视域下翻译研究：特征、领域与意义[J]. 中国翻译，2020，41（2）：5-15，187.

[33] 王兆鹏，邵大为．数字人文在古代文学研究中的初步实践及学术意义[J]. 中国社会科学，2020（8）：108-129，206-207.

[34] 陈刚．“数字人文”与历史地理信息化研究[J]. 南京社会科学，2014（3）：136-142.

[35] 夏翠娟．数字人文的理论沿革和范式转向：从“人文研究的数字方法”到“后数字时代的人文研究”[J]．中国图书馆学报，2024，50（5）：83-101.

[36] 毋江波，司玥．知识生态视角下数字人文领域学科知识的演化态势及其扩散逻辑分析[J]．图书情报工作，2025，69（20）：134-152.

[37] 张雪，张志强．学科交叉研究系统综述[J]．图书情报工作，2020，64（14）：112-125.

[38] 叶佳鑫，熊回香，陈勇跃. 知识流动视角下的跨学科知识分类研究[J]. 情报理论与实践，2024，47（9）：156-165.

[39] 张琳，孙蓓蓓，黄颖. 跨学科合作模式下的交叉科学测度研究：以ESI社会科学领域高被引学者为例[J]．情报学报，2018，37（3）：231-242.

作者简介

刘桂锋，男，博士，教授，研究方向：科学数据管理、大数据分析，E-mail：liuguifeng29@163.com。
马明睿，女，硕士研究生，研究方向：数字人文。
韩牧哲，男，博士，讲师，研究方向：知识组织、知识管理。
刘琼，女，硕士，副研究馆员，研究方向：科学数据管理。

Knowledge Diffusion in the Field of Digital Humanities in China Based on Modularity Citation Networks

LIU GuiFeng[1] MA MingRui[1] HAN MuZhe[1] LIU Qiong[2]
(1. Institute of Science and Technology Information, Jiangsu University, Zhenjiang 212013, P. R. China;
2. Jiangsu University Library, Zhenjiang 212013, P. R. China)

Abstract: With the continuous development of digital humanities research, knowledge diffusion is characterized by interdisciplinary interaction and dynamic evolution, and citation networks serve as an analytical tool for identifying such features. Based on citation data from papers published in China related to digital humanities, this study constructs a citation network using Gephi software and conducts modularity analysis. It measures the knowledge diffusion level of modular subnets through four indicators: net knowledge diffusion volume, relative speed of knowledge diffusion, relative breadth of knowledge diffusion, and relative intensity of knowledge diffusion. It further reveals knowledge flow paths by examining citation relationships between modules. The research identifies five thematic modules in China's digital humanities field, with obvious knowledge flow directions between each module. The overall diffusion process exhibits interdisciplinary, multi-path, and dynamic characteristics, which can reveal the research pattern and evolution path of the digital humanities field. This is of reference value for understanding the disciplinary structure and promoting future development.

Keywords: Digital Humanities; Knowledge Diffusion; Citation Network; Knowledge Flow; Modularity Analysis

（责任编辑：王玮）