

开放获取背景和疫情形势下 预印本发表模式的扩散研究

——以bioRxiv为例

The Diffusion of Preprint Publishing in the Contexts of Open Science and COVID-19 Pandemic: A Case Study of bioRxiv

谢维熙¹ 张光耀¹ 郑轩² 王贤文¹
XIE Weixi ZHANG Guangyao ZHENG Xuan WANG Xianwen

(1. 大连理工大学科学学与科技管理研究所暨WISE实验室, 大连, 116024; 2. 大连理工大学经济管理学院, 大连, 116024)

摘要:【目的/意义】在开放获取运动和疫情背景下, 预印本发表模式在科学界得到广泛关注和运用, 探究预印本发表在国家和机构间的扩散趋势对于预印本平台建设以及改善学术交流体系具有重要意义。【研究设计/方法】以生物学领域主流预印本平台 bioRxiv 全部论文 (2013-2021年) 为分析对象, 使用文献计量方法, 从时间、国家和机构维度分析 bioRxiv 发表在国家和机构层面的演化规律, 以及疫情对 bioRxiv 发表扩散的影响。【结论/发现】bioRxiv 发表存在从发达国家向发展中国家, 以及从顶尖机构向普通机构扩散的趋势, 新冠疫情对预印本发表在发展中国家的扩散趋势有负面影响。【创新/价值】从时间、空间等多角度揭示预印本发表的扩散规律, 并分析了疫情对预印本扩散的影响, 为预印本平台建设提供一定启示。

关键词: 预印本发表; 预印本; 扩散趋势; 开放科学; 新冠疫情; bioRxiv

中图分类号: G250 **DOI:** 10.13366/j.dik.2022.03.050

引用本文: 谢维熙, 张光耀, 郑轩, 等. 开放获取背景和疫情形势下预印本发表模式的扩散研究——以 bioRxiv 为例 [J]. 图书情报知识, 2022, 39(3): 50-59. (Xie Weixi, Zhang Guangyao, Zheng Xuan, et al. The Diffusion of Preprint Publishing in the Contexts of Open Science and COVID-19 Pandemic: A Case Study of bioRxiv[J]. Documentation, Information & Knowledge, 2022, 39(3): 50-59.)

Abstract: 【Purpose/Significance】In the context of open access movement and epidemic situation, preprint publishing has been widely concerned and applied in the scientific community. Exploring the diffusion trend of preprint publishing among countries and institutions is of great significance for the construction of preprint platform and the improvement of academic exchange system. 【Design/Methodology】Taking all the preprints of bioRxiv, the mainstream preprint platform in the field of Biology (2013-2021), as the analysis object, and using the bibliometric approach, this paper analyzes the evolution law of preprints in bioRxiv at the national and institutional levels from the temporal perspective and country/institution level, as well as the impact of the epidemic on the publishing spread in bioRxiv. 【Findings/Conclusion】The results show that preprints publishing in bioRxiv spread from developed countries to developing countries and from top institutions to ordinary institutions, and the COVID-19 epidemic has a negative impact on the diffusion trend of preprint publications in developing countries. 【Originality/Value】This work, which reveals the diffusion law of preprint publishing from multiple perspectives such as time and space, and explores the impact of the epidemic on the diffusion of preprint publishing, has implications to provide some enlightenment for the construction of preprint platform.

Keywords: Preprint publishing; Preprint; Diffusion trend; Open science; COVID-19 pandemic; bioRxiv

1 引言

随着互联网的发展和开放科学的不断深入^[1], 科研成果传播和交流的方式也在不断推陈出新。预印本便是一种重要的学术成果交流形式, 研究人员为进行更广泛的学术交流, 将科研成果通过网络发布到预印本平台。相较于正式出版的论文, 预印本具有传播和交流速度快、交流更广泛更开放的特点。研究人员通过预印本向同行和公众及时分享自己的研究成果, 极大

地促进了科学成果的传播和交流。

2019年末爆发的新冠疫情 (COVID-19) 对科学界产生了巨大的影响, 科学家们为阻止病毒的传播以前所未有的速度投入到相关研究中。作为传统以期刊为核心的学术交流模式的替代或补充, 更多科学家选择将研究成果第一时间发布在预印本平台, 预印本的优势在这次疫情中得到了充分体现^[2]。预印本可以实时分享研究结果, 避免同行评审的延误, 在新冠肺炎病毒未得到完全解决以及开放科学不断推进的时代, 预

【基金项目】本文系国家自然科学基金项目“科学文献全景大数据下的研究热点及研究前沿探测”(71974029)的研究成果之一。(This is an outcome of the project "Research Hotspots and Frontier Detection under the Panoramic Big Data of Scientific Literature"(71974029) supported by National Natural Science Foundation of China.)

【通讯作者】王贤文 (ORCID: 0000-0002-7236-9267), 博士, 教授, 研究方向: 科技情报, 科学学与科技管理, Email: xianwenwang@dlut.edu.cn. (Correspondence should be addressed to WANG Xianwen, Email: xianwenwang@dlut.edu.cn, ORCID: 0000-0002-7236-9267)

【作者简介】谢维熙 (ORCID: 0000-0003-2330-7980), 博士研究生, 研究方向: 科学学与科技管理, Email: wsrjcydcfsnt@163.com; 张光耀 (ORCID: 0000-0002-0505-1160), 博士研究生, 研究方向: 科学学, Email: sdgyzhang@163.com; 郑轩 (ORCID: 0000-0002-3093-1514), 本科生, Email: zx1742275572@163.com。

印本在科学交流体系中的关键作用愈发明显^[3]。预印本的扩散趋势是指预印本在一段时间内,通过特定的渠道在某一团体中传播的过程。探究预印本的扩散趋势有着重要的意义,然而以往的研究对此鲜有涉及。因此,本文以bioRxiv预印本为研究对象,探究预印本发表模式的扩散趋势,从国家和机构的角度对其扩散趋势进行分析。

2 相关研究

2.1 开放获取与预印本的发展

开放科学是一种推进科学开放与共享的科学实践,倡导增加科学的透明度,扩大研究成果的访问范围,是全球科技发展的重要趋势,开放获取是组成开放科学的一部分,也是开放科学的核心要素^[4]。为促进科学信息的广泛交流以及提高科学研究的效率,开放获取为作者免费提供在互联网上获取和使用科研文献的机会,预印本平台是开放获取相关实践的重要部分^[5]。预印本arXiv最初于1991年推出,主要为物理研究人员在论文正式发表之前提供分享平台,而后伴随开放科学运动的兴起得到迅速发展,同时,预印本平台也有助于开放科学,能够快速而广泛地分享研究成果。预印本是一种为弥合论文提交和出版之间的时间差而建立的开放获取在线存储库^[6],存储未经正式同行评审的手稿,预印本的筛选过程是粗略和快速的,平台对提交到预印本服务器上的文章的完整性和是否抄袭以及符合法律和道德规范进行检查,大多数预印本服务器为每份提交的文章提供DOI^[7]。继arXiv之后,社会科学、生物学和医学等领域也推出了各自的预印本平台,分别为SSRN、bioRxiv和medRxiv等。

预印本的出现不仅推动了开放科学运动^[8],让读者快速免费地获取科学内容,同时研究者在将论文提交至期刊之前也可以得到及时的反馈并确立其研究的优先权^[9],研究人员通过发布预印本可以即时传达他们的研究结果,从而得到及时的反馈以及促进早期合作。研究表明最终发表的出版物中发布过预印本的文章有更多的关注和引用^[10,11],预印本的推出还降低了论文发表的偏见,尤其是对于职业生涯早期的科研人员^[12]。预印本作为期刊出版物的替代或补充,加速了科学信息的传播,在一定程度上还加速了研究的进程,

同时还为出版生态系统的演变提供了可能性^[13]。

从学术界的态度来看,随着研究进程的加快,研究人员对同行评审的速度以及已发表论文的可获得性愈发不满,尽管预印本的可信度因缺乏严格的评审被诟病^[14-15],然而其及时性对于科学交流至关重要,预印本也得到了学术团体、基金组织甚至传统出版商的认可和支持,生物学出版商鼓励提交预印本^[16],Elife等知名期刊直接接受bioRxiv网站的提交,节省了作者向期刊提交论文的时间^[17]。美国国立卫生研究院(NIH)和欧洲研究理事会等主要研究资助者也表达了对预印本的支持^[16]。

2.2 疫情形势下的预印本

2019年末爆发的新冠疫情引发了科学文献的激增,尤其是与新冠肺炎相关的文献数量大幅增加^[18]。随着新型冠状病毒的传播,与冠状病毒有关的论文出版规范迅速发生变化,传统的论文发表已经无法满足人们对信息的迫切需求,信息的即时性对于决策者和普通民众都至关重要。预印本因其传播速度快而受到广泛使用,其中医学相关预印本论文增长速度尤其快,最受欢迎的两个生物医学领域的预印本服务器是bioRxiv和medRxiv^[19]。

疫情前期,预印本占据冠状病毒相关研究的大部分,相比较于学术出版物,预印本获得了更多的关注和引用^[20],预印本的广泛使用促进了学者之间的快速交流,加速了研究进程,为疫情防控和疫苗研发提供即时信息,也对政策制定起到了至关重要的作用^[21]。然而,预印本的质量控制问题仍令人担忧^[22],预印本服务器上发布的手稿因未经严格的同行评审而饱受争议,研究结果可能会影响临床实践,甚至影响患者的安全^[23]。虽然严格的同行评审过程能够保证科学研究质量和完整性,但是即使在学术期刊已经大幅缩短论文出版时间的情况下,几个月甚至更久的评审过程与疫情下对信息的迫切需求仍然存在巨大的张力^[24]。也有研究表示预印本论文与期刊论文的可靠性相差不大,Brainard对比疫情期间经过同行评审发表在期刊上的文章与发布在预印本平台上的文章后发现,两者在可靠性等方面的差距在变小^[19]。

为了解疫情期间关于冠状病毒预印本累积数量的演化模式,Vasconcelos建立模型发现在疫情爆发后很短的时间内产生了前所未有数量的学术论文,然而这

种增速在几个月后降低,甚至文献数量有下降趋势^[25]。Abdill 基于 bioRxiv 预印本在国家层面展示了全球参与 bioRxiv 的大致情况,国家在 bioRxiv 上的投稿数量以及主导地位分布极不均衡^[26]。相关研究还分析了 bioRxiv 预印本使用量的变化趋势,发现自 2013 年推出以来预印本的提交量和下载量激增,其中首次发布预印本的作者数量也持续上升,此外大部分论文最终经过同行评审发表到期刊上^[24]。

预印本作为一种新兴的学术交流载体已经在许多学科领域得到广泛关注和应用,因此探索预印本的扩散趋势有着重要的意义。目前国内外关于 bioRxiv 以及预印本的研究,问题主要集中于预印本平台的优势、影响与局限性^[27],预印本的提交量和下载量,最终有多少预印本在期刊上发表以及预印本到最终发表的版本发生的变化,预印本研究领域的变化趋势^[17,28],预印本的关注和引用优势^[11,29]等方面,国内也有学者从政策角度对预印本的发展态势进行分析^[30],然而几乎没有研究涉及预印本在时间演化和空间扩散上的趋势。

本文以生物学领域最大的预印本平台 bioRxiv 为研究对象,选取 2013-2021 年 Dimensions 平台上的 bioRxiv 数据集,使用文献计量法从多维度探究 bioRxiv 在国家和机构层面的演化规律,以及疫情对扩散趋势的影响,有助于了解 bioRxiv 的整体格局和发展趋势,从预印本角度总结学术交流演变规律,为预印本平台建设和学术交流体系发展提供参考。本研究主要关注以下几个研究问题:

- (1) 国家层面上, bioRxiv 预印本发表在国家经济发展水平下呈现什么样的扩散趋势?
- (2) 机构层面上, bioRxiv 预印本发表从机构科研水平视角下存在什么样的扩散趋势?
- (3) 新冠疫情形势下, bioRxiv 预印本发表的扩散趋势有何特征?

3 研究数据与方法

Digital Science 公司于 2018 年初发布了一个新型的综合科研信息数据平台——Dimensions (<https://app.dimensions.ai/>), 收录了大量科研信息资源, 包括文献(期刊和会议论文、图书、预印本等)、研究数据集、基金和专利数据、临床试验和跨数据源关联关系

等,为科研人员、出版商以及科研经费资助机构所需的科技情报和决策提供数据支撑^[31]。对本研究而言,虽然数据来源于 bioRxiv,但是需从 Dimensions 采集。这是因为 bioRxiv 网站不提供论文元数据采集的 API,对于论文作者所属的机构和国家信息,从 bioRxiv 下载的原始字段没有经过结构化处理,无法进行本文研究所需要的分析;而 Dimensions 平台提供了 bioRxiv 论文作者的所属机构和国家的结构化字段,从 Dimensions 下载的数据可以方便地进行分析。

预印本服务器 bioRxiv 于 2013 年 11 月由冷泉港实验室(CSHL)及其出版社(CSHLP)推出,每年提交的论文数量稳步增加^[17],现已成为生物科学领域最大的预印本平台,因此在研究疫情相关问题其数据更具有代表性。从研究领域来看, bioRxiv 最初主要包括进化生物学、遗传学和计算生物学等,现在更多的是神经科学、细胞和发育生物学的内容,其他领域的研究者也在 bioRxiv 上发布了大量文章^[13]。

考虑到 bioRxiv 的学科领域和预印本的及时性特点,以及 Dimensions 平台上数据的完整性和可用性,本研究于 2022 年 1 月 15 日在 Dimensions 平台收集了 2013-2021 年 bioRxiv 的文献数据,去除字段内容不完整的数据,得到了 141,731 份预印本作为研究数据,对元数据进行处理后,提取数据中的有用信息,包括预印本的数字对象标识符(DOI)、发布到 bioRxiv 的日期、作者、所属机构和国家等。

4 结果与分析

4.1 bioRxiv 平台上发布文章的国家 and 机构数量变化趋势

从整体角度分析 2013-2021 年在 bioRxiv 平台上发布文章的国家 and 机构数量变化情况。国家和机构数量及其论文量变化在一定程度上反映了 bioRxiv 扩散程度以及未来发展趋势,也从整体上反映了科学界对 bioRxiv 的认可程度。

对论文量进行统计,如图 1 所示, bioRxiv 从 2013 年推出以来论文量稳步增长,其中, 2016 年开始,论文量增长速度明显加快。在作者的国家 and 机构归属方面,由于已有大量研究证实末位作者在研究中有着重要的贡献,尤其是在生物医学领域^[26,32]。因此,进一步

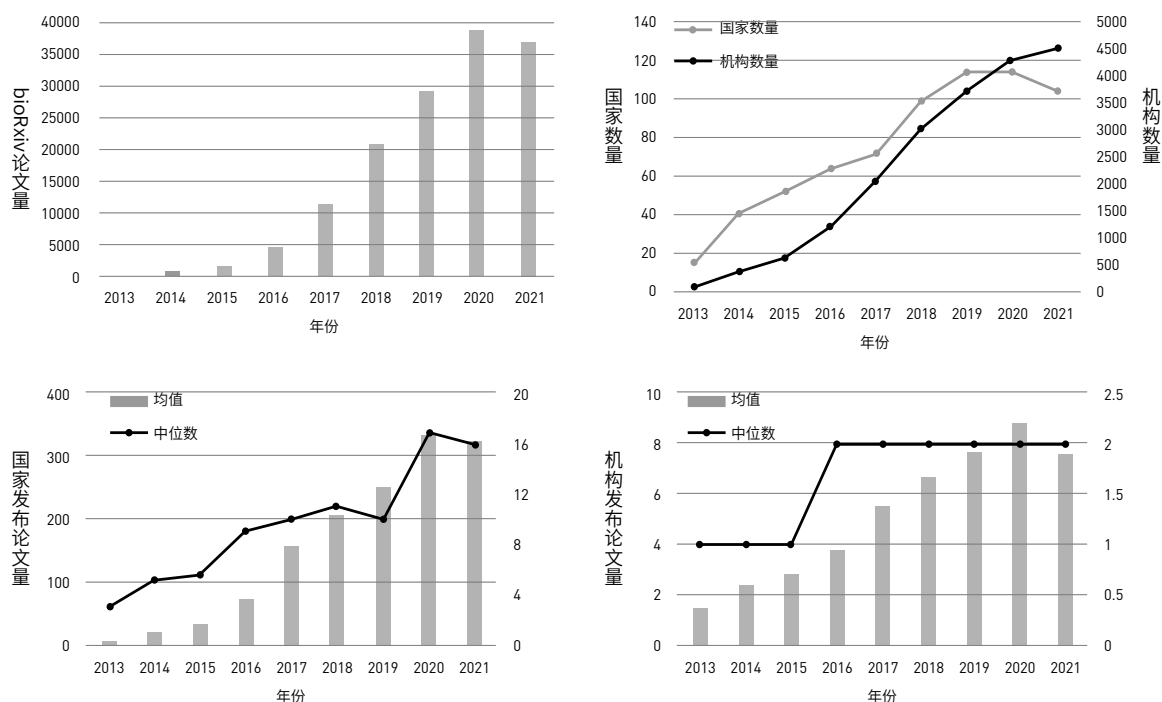


图1 2013-2021年 bioRxiv 论文量和国家 / 机构量及其论文量变化趋势
Fig.1 Changes in the Number of Countries / Institutions and Their Papers on bioRxiv from 2013 to 2021

对每年以末位作者身份在bioRxiv上发文的国家和机构数量及其论文量进行统计,bioRxiv上国家和机构数量及其论文量在2013-2020年呈现明显上升的趋势,但论文量、国家和机构数量都在2021年有小幅下降。2019年以后虽然论文量和机构数量持续增长,但是到2020年论文量和国家数量均有所下降,机构数量增长速度减缓。由此可见,疫情之前,bioRxiv预印本的发文量呈现良好的增长趋势,在疫情爆发后,发文国家数量的增长趋势减缓并在2021年出现回落,发文机构数量的增长趋势也在2021年有所放缓。

4.2 国家经济发展水平视角下的bioRxiv扩散趋势

进一步从国家经济发展水平角度分析bioRxiv在国家层面的扩散趋势。在国家经济发展水平划分上,综合世界银行、国际货币基金组织、联合国开发计划署和美国中央情报局发布的资料,将国家分为发达国家和发展中国家,以分析其数量变化。发达国家和发展中国家数量及其占比变化趋势如图2所示。从国家经济发展水平来看,2013-2018年发展中国家数量呈上升趋势,而发达国家相反。两类国家数量在2018-2020年

基本稳定,然而发展中国家数量在2021年小幅下降。

为进一步探究国家层面的演化规律,取各国家机构数量排名、国家机构数量的对数分别为横纵坐标绘制散点图,如图3所示,2013-2021年结果均呈幂律分布,且幂律分布的长尾分布特征逐渐减弱,2013-2019年发展中国家数量逐渐增多,但2019年之后,发展中国家数量有所减少。总的来说,bioRxiv预印本的发表模式呈现由发达国家向发展中国家扩散的趋势,而疫情使得这一扩散趋势有所逆转。

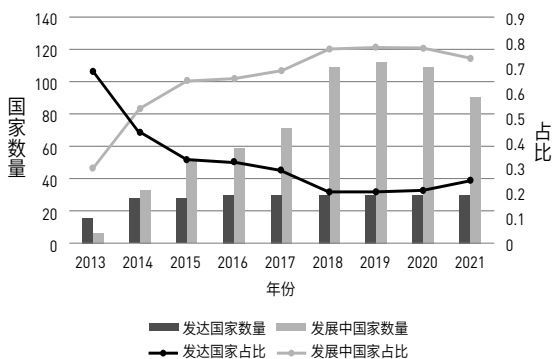


图2 发达国家和发展中国家数量及其占比变化趋势
Fig.2 Changes in the Number and Proportion of Developed and Developing Countries

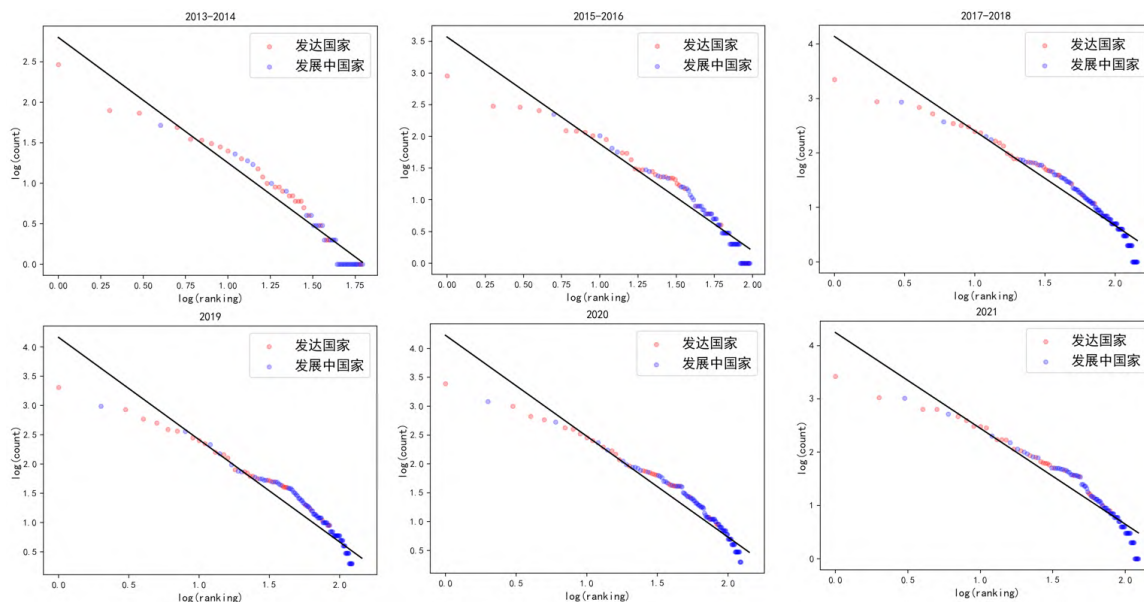


图3 发文机构的幂律分布图

Fig.3 Power Law Distribution of the Institutions Publishing on bioRxiv

4.3 机构科研水平视角下的bioRxiv扩散趋势

bioRxiv 发文机构主要为大学, 也包括研究所和医院等其他机构, 为探究机构层面是否存在由顶尖机构向普通机构扩散的趋势, 本研究只考虑大学的排名, 使用2021软科世界大学学术排名 (Shanghai Ranking's Academic Ranking of World Universities, ARWU) 和U.S. News在生物学和生物化学领域的排名作为大学科研水平的评价标准, 两个排名均是世界范围内公认的较为权威的世界大学排名。对每年发文量排名前50的机构的发文量及其大学的最新软科排名进行统计, 结果如表1所示。为了解发文量Top50机构中高校科研水平排名分布, 考虑到论文发布量还计算了加权平均排名, 结果如图4所示, 加权平均值计算公式见公式1。

$$\bar{X} = \frac{x_1w_1 + x_2w_2 + \cdots + x_nw_n}{w_1 + w_2 + \cdots + w_n} \quad (1)$$

公式1中, \bar{X} 为加权平均排名, x_n 为第n个高校软科排名, w_n 为第n个高校在bioRxiv上发表论文量。

由表1中可以看到, 哈佛大学、斯坦福大学、牛津大学等一些顶尖大学发文量始终位居前列, 2015-2019年有许多软科排名靠后的大学进入发文量前五十, 而2019年之后, 发文量前五十的机构大部分软科排名靠前。除中国科学院大学在2020和2021年分

列26和31位之外, 中国机构很少进入发文量前50。

如图4所示, 无论从软科排名还是U.S. News生物学排名来看, 高校科研水平排名和加权平均排名的变化趋势基本一致, 即2013年和2014年平均排名较高, 2015-2019年平均排名呈上升趋势, 2019年之后, 平均排名有所下降。总的来说, bioRxiv预印本发表模式存在向普通机构扩散的趋势, 但疫情的爆发对该扩散趋势有负向影响。

4.4 新冠疫情对bioRxiv扩散趋势的影响

从整体上对bioRxiv平台上发布文章的国家 and 机构数量变化趋势, 以及从国家和机构层面对bioRxiv扩散趋势的分析, 可知疫情给bioRxiv的扩散趋势造成了负面影响。为进一步验证结论, 对每一年消失的国家和机构 (即前一年在bioRxiv上有论文发布, 但该年没有) 进行统计分析, 结果如表2和表3所示。

2013-2019年即疫情前消失的国家有24个, 主要是亚洲、非洲的发展中国家, 另外还有欧洲的三个发达国家, 分别是卢森堡、斯洛伐克和马耳他; 2019-2021年即疫情期间, 消失的国家有29个且全部是发展中国家, 主要分布在北美洲、亚洲和非洲。

对比疫情前后退出的机构发现, 疫情后消失的机构数量较多, 但相比于疫情前, 疫情后消失的机构大多数是科研水平排名靠后的机构。新冠疫情严重影响

表1 Top50机构的发文量

Table 1 Number of Documents Published by the Top50 Institutions

| 序号 | 高校名称 | 论文数 | 软科排名 | 序号 | 高校名称 | 论文数 | 软科排名 |
|-------------|--------------|-------|-------------|-------------|--------------|-------|--------------|
| 2013-2014 年 | | | | 2015-2016 年 | | | |
| 1 | 哈佛大学 | 61 | 1 | 1 | 哈佛大学 | 366 | 1 |
| 2 | 加利福尼亚大学伯克利分校 | 40 | 5 | 2 | 斯坦福大学 | 233 | 2 |
| 3 | 斯坦福大学 | 36 | 2 | 3 | 牛津大学 | 200 | 7 |
| 4 | 剑桥大学 | 32 | 3 | 4 | 剑桥大学 | 192 | 3 |
| 5 | 加利福尼亚大学戴维斯分校 | 28 | 100 | 5 | 博德研究所 | 183 | - |
| 6 | 华盛顿大学 | 28 | 19 | 6 | 伦敦大学学院 | 160 | 17 |
| 7 | 冷泉港实验室 | 26 | - | 7 | 华盛顿大学 | 157 | 19 |
| 8 | 博德研究所 | 24 | - | 8 | 加利福尼亚大学伯克利分校 | 157 | 5 |
| 9 | 麻省理工学院 | 24 | 4 | 9 | 爱丁堡大学 | 141 | 38 |
| 10 | 多伦多大学 | 23 | 22 | 10 | 麻省理工学院 | 138 | 4 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 50 | 杜克大学 | 8 | 32 | 50 | 宾夕法尼亚州立大学 | 51 | 117 |
| | | | 平均排名 (66) | | | | 平均排名 (39.3) |
| 2017-2018 年 | | | | 2019 年 | | | |
| 1 | 哈佛大学 | 1,537 | 1 | 1 | 哈佛大学 | 1,143 | 1 |
| 2 | 斯坦福大学 | 930 | 2 | 2 | 斯坦福大学 | 724 | 2 |
| 3 | 剑桥大学 | 892 | 3 | 3 | 牛津大学 | 694 | 7 |
| 4 | 牛津大学 | 889 | 7 | 4 | 剑桥大学 | 662 | 3 |
| 5 | 伦敦大学学院 | 726 | 17 | 5 | 伦敦大学学院 | 519 | 17 |
| 6 | 华盛顿大学 | 639 | 19 | 6 | 华盛顿大学 | 509 | 19 |
| 7 | 加利福尼亚大学旧金山分校 | 634 | 20 | 7 | 加利福尼亚大学圣迭戈分校 | 498 | 18 |
| 8 | 加利福尼亚大学圣迭戈分校 | 615 | 18 | 8 | 加利福尼亚大学旧金山分校 | 477 | 20 |
| 9 | 加利福尼亚大学伯克利分校 | 596 | 5 | 9 | 密歇根大学 | 474 | 26 |
| 10 | 约翰斯霍普金斯大学 | 573 | 16 | 10 | 加利福尼亚大学伯克利分校 | 473 | 5 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 50 | 匹兹堡大学 | 245 | 143 | 50 | 贝勒医学院 | 200 | 151 |
| | | | 平均排名 (39) | | | | 平均排名 (42.5) |
| 2020 年 | | | | 2021 年 | | | |
| 1 | 哈佛大学 | 1,418 | 1 | 1 | 哈佛大学 | 1,271 | 1 |
| 2 | 斯坦福大学 | 846 | 2 | 2 | 斯坦福大学 | 706 | 2 |
| 3 | 牛津大学 | 842 | 7 | 3 | 牛津大学 | 685 | 7 |
| 4 | 剑桥大学 | 802 | 3 | 4 | 剑桥大学 | 639 | 3 |
| 5 | 加利福尼亚大学圣迭戈分校 | 692 | 18 | 5 | 华盛顿大学 | 624 | 19 |
| 6 | 加利福尼亚大学旧金山分校 | 647 | 20 | 6 | 康奈尔大学 | 587 | 12 |
| 7 | 华盛顿大学 | 623 | 19 | 7 | 加利福尼亚大学圣迭戈分校 | 565 | 18 |
| 8 | 伦敦大学学院 | 593 | 17 | 8 | 伦敦大学学院 | 545 | 17 |
| 9 | 康奈尔大学 | 592 | 12 | 9 | 加利福尼亚大学旧金山分校 | 542 | 20 |
| 10 | 宾夕法尼亚大学 | 570 | 15 | 10 | 约翰斯霍普金斯大学 | 529 | 16 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 50 | 埃默里大学 | 265 | 105 | 50 | 范德堡大学 | 234 | 65 |
| | | | 平均排名 (36.7) | | | | 平均排名 (36.28) |

开放获取背景和疫情形势下预印本发表模式的扩散研究——以 bioRxiv 为例
The Diffusion of Preprint Publishing in the Contexts of Open Science and COVID-19 Pandemic: A Case Study of bioRxiv

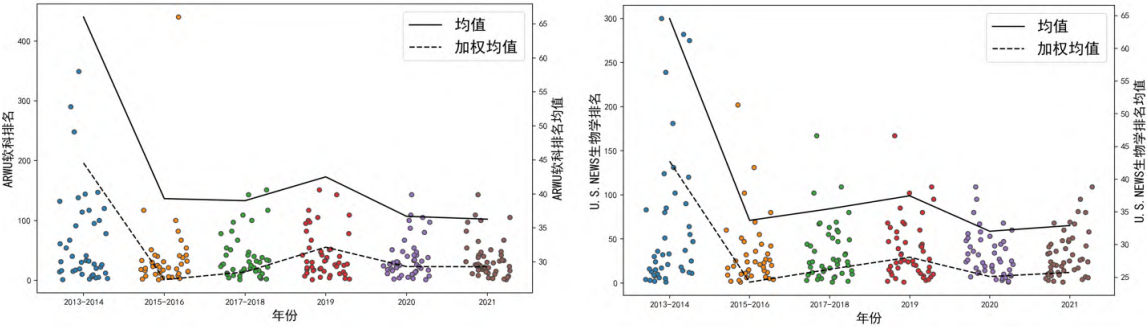


图4 发文量 Top50机构中高校科研水平排名分布图
Fig.4 Ranking Distribution of Research Level of Colleges and Universities Among the Top 50 Institutions

表2 疫情前后退出 bioRxiv 的国家对比
Table 2 Comparison of Countries Withdrawing from bioRxiv Before and After the Epidemic

| 年份 | 国家 | 发达国家 | 所属洲 | 年份 | 国家 | 发达国家 | 所属洲 |
|----------------|----------|------|-----|----------------|--------|------|-----|
| 2013-2019（疫情前） | | | | 2019-2021（疫情后） | | | |
| 2013-2014 | 乌干达 | 否 | 非洲 | 2019-2020 | 巴拉圭 | 否 | 南美洲 |
| 2014-2015 | 贝宁 | 否 | 非洲 | 2019-2020 | 伯利兹 | 否 | 北美洲 |
| 2014-2015 | 科特迪瓦 | 否 | 非洲 | 2019-2020 | 多哥 | 否 | 非洲 |
| 2014-2015 | 卢森堡 | 是 | 欧洲 | 2019-2020 | 海地 | 否 | 北美洲 |
| 2014-2015 | 斯洛伐克 | 是 | 欧洲 | 2019-2020 | 斯威士兰 | 否 | 非洲 |
| 2014-2015 | 乌兹别克斯坦 | 否 | 亚洲 | 2019-2020 | 乌兹别克斯坦 | 否 | 亚洲 |
| 2014-2015 | 亚美尼亚 | 否 | 亚洲 | 2020-2021 | 阿尔巴尼亚 | 否 | 欧洲 |
| 2015-2016 | 巴布亚新几内亚 | 否 | 大洋洲 | 2020-2021 | 阿富汗 | 否 | 亚洲 |
| 2015-2016 | 刚果 | 否 | 非洲 | 2020-2021 | 安哥拉 | 否 | 非洲 |
| 2015-2016 | 古巴 | 否 | 北美洲 | 2020-2021 | 巴巴多斯 | 否 | 北美洲 |
| 2015-2016 | 摩洛哥 | 否 | 非洲 | 2020-2021 | 巴林 | 否 | 亚洲 |
| 2015-2016 | 塞尔维亚 | 否 | 欧洲 | 2020-2021 | 博兹瓦纳 | 否 | 非洲 |
| 2015-2016 | 危地马拉 | 否 | 北美洲 | 2020-2021 | 不丹 | 否 | 亚洲 |
| 2016-2017 | 马里 | 否 | 非洲 | 2020-2021 | 多米尼加 | 否 | 北美洲 |
| 2016-2017 | 蒙古 | 否 | 亚洲 | 2020-2021 | 厄立特里亚 | 否 | 非洲 |
| 2016-2017 | 苏丹 | 否 | 非洲 | 2020-2021 | 斐济 | 否 | 大洋洲 |
| 2017-2018 | 贝宁 | 否 | 非洲 | 2020-2021 | 佛得角 | 否 | 非洲 |
| 2017-2018 | 利比里亚 | 否 | 非洲 | 2020-2021 | 格鲁吉亚 | 否 | 亚洲 |
| 2017-2018 | 马耳他 | 是 | 欧洲 | 2020-2021 | 吉尔吉斯斯坦 | 否 | 亚洲 |
| 2017-2018 | 萨尔瓦多 | 否 | 北美洲 | 2020-2021 | 毛里求斯 | 否 | 非洲 |
| 2018-2019 | 阿拉伯联合酋长国 | 否 | 亚洲 | 2020-2021 | 毛里塔尼亚 | 否 | 非洲 |
| 2018-2019 | 埃塞俄比亚 | 否 | 非洲 | 2020-2021 | 摩纳哥 | 否 | 欧洲 |
| 2018-2019 | 佛得角 | 否 | 非洲 | 2020-2021 | 尼加拉瓜 | 否 | 北美洲 |
| 2018-2019 | 瓦努阿图 | 否 | 大洋洲 | 2020-2021 | 萨摩亚 | 否 | 大洋洲 |
| | | | | 2020-2021 | 塞舌尔 | 否 | 非洲 |
| | | | | 2020-2021 | 瓦努阿图 | 否 | 大洋洲 |
| | | | | 2020-2021 | 文莱 | 否 | 亚洲 |
| | | | | 2020-2021 | 也门 | 否 | 亚洲 |
| | | | | 2020-2021 | 乍得 | 否 | 非洲 |

注：表中内容根据年份排序，同一年份则根据国家名称的拼音字母排序

表3 疫情前后退出 bioRxiv 的机构对比

Table 3 Comparison of Institutions Withdrawing from bioRxiv Before and After the Epidemic

| | 2013-2019 (疫情前) | 2019-2021 (疫情后) |
|---------------|-----------------|-----------------|
| 消失机构数量 | 4,450 | 5,048 |
| 软科排名前1000机构数量 | 159 | 20 |
| 前1000机构排名均值 | 597.11 | 703.1 |
| 软科排名前100机构数量 | 4 | 0 |
| 软科排名前200机构数量 | 14 | 1 |
| 软科排名前500机构数量 | 58 | 3 |

了国家和机构的科研活动，其中对发展中国家和普通机构的影响更大。

5 结论

预印本的出现加速了科学传播，促进了学术交流体系的发展。本文以2013-2021年bioRxiv的预印本为样本，采用文献计量方法，从时间演化、国家经济发展水平以及机构科研水平多角度对bioRxiv在国家和机构层面的扩散趋势进行探索，并探究新冠疫情对bioRxiv扩散趋势的影响。

整体上，越来越多的国家和机构的科研人员将研究成果发布到bioRxiv上，说明该平台作为生物学界最大的预印本平台在全球范围内逐渐得到认可和关注。从国家经济发展水平来看，越来越多的发展中国家在bioRxiv上发布文章，从参与国家数量角度进行分析可知，bioRxiv由发达国家向发展中国家流行，观察疫情前后国家数量变化，发现疫情使得少数发展中国家在bioRxiv上的科学活动受到影响，幂律分布图也证实了该扩散趋势的存在。从机构层面来看，考虑到2013年数据仅包括11月和12月的预印本，论文量较少，且2013-2014年bioRxiv处于发展初期，尚未在全球范围内得到广泛使用，因此2013-2014年大学平均排名较高。2015-2019年bioRxiv发展迅速，越来越多的普通大学发文量进入发文量前五十，然而，2019年后顶尖大学更多占据前列。因此，可以得出结论：bioRxiv有

从顶尖机构向普通机构扩散的趋势，但疫情对普通机构的影响更大，对该扩散有负面影响。

新冠疫情对bioRxiv的扩散趋势有负面影响，整体上虽然bioRxiv论文量在疫情前期持续增长，但在后疫情时代有所下降。分析国家和机构的扩散趋势后发现，疫情对bioRxiv在国家和机构视角的扩散均存在不同程度的负向影响。从每年消失的国家和机构的分析中可以发现，相较于发达国家和顶尖机构，疫情对发展中国家和普通机构的科学工作影响更大，发展中国家和普通机构的科学活动较大程度上受到疫情影响。从中国在bioRxiv的发文情况来看，中国的科研产出能力逐渐增强并仅次于美国，但在机构层面上，单个机构竞争力并不强。

值得注意的是，本文中各种指标的增长趋势在2020年后有所减缓，甚至发生逆转，对于这一现象有必要作出解释。疫情对科学界产生了巨大的影响。虽然疫情的挑战刺激了科学研究数量和速度的大幅提高，但与此同时社会隔离和封锁导致了大量科学研究的中断。首先，在预印本数量急剧增长几个月后，研究人员很难以疫情初期同样的速度持续进行科学研究；其次，研究人员有可能把研究重心从新型冠状病毒转移到新的或者以前的研究方向，尤其是与公共卫生和生物学无关的一些研究人员。此外，对预印本科学可信度的质疑导致预印本筛查审核过程更加严格，这也会影响预印本的提交数量^[25]。

以预印本为媒介的学术交流模式迅速发展，正在改变传统以期刊为主的学术交流模式，很大程度上推动了开放科学的发展，因此对预印本的扩散趋势进行探索是十分有必要的。相比以往的研究，本研究探索了bioRxiv在国家和机构层面的扩散趋势，多层次分析了预印本的扩散趋势，同时从时间、国家经济和机构科研发展水平多角度反映了bioRxiv的演化趋势，为预印本的研究提供了新视角。本研究有助于了解bioRxiv的整体格局及发展趋势，了解各国家和机构在以bioRxiv为载体的学术交流体系中的地位 and 竞争力，为预印本平台建设和学术交流体系研究提供基础。

结合以上分析和结论，本文提出以下建议。首先，对于预印本平台来说，加强预印本在发展中国家和普通科研机构的宣传推广，扩大预印本的使用受众范围。其次，对于所有国家和机构来说，在后疫情时代，应积极采取措施促进科研工作者的科学活动的恢复，尽量

开放获取背景和疫情形势下预印本发表模式的扩散研究——以 bioRxiv 为例
The Diffusion of Preprint Publishing in the Contexts of Open Science and COVID-19 Pandemic: A Case Study of bioRxiv

减轻或者消除疫情对科学活动的影响。再次,对于发展中国家和普通科研机构的科研工作者来说,应积极利用预印本平台,如关注预印本论文、利用预印本平台发布论文等,以预印本为途径提升在学术交流体系中的地位。最后,对于中国来说,目前我国已经开始建设自己的预印本平台ChinaXiv,建设过程中如何吸取国际上成熟的预印本平台如arXiv、bioRxiv、SSRN等的经验,尤其是现阶段完善ChinaXiv平台的基础架构、加

大平台的宣传、提升平台在国内和国际上的影响力等是当务之急。

本研究也存在一定局限。首先,研究数据集仅限于bioRxiv预印本;其次,本文从时间演化、国家经济发展水平和机构科研水平这几个角度分析bioRxiv在国家和机构层面的演化趋势,地理空间等方面的演化规律没有涉及。在之后的研究中,将从更加多源的数据集和更多角度对预印本的扩散趋势进行探究。

作者贡献说明

谢维熙: 设计研究方案,数据收集、处理和分析,论文写作;

张光耀: 论文写作与修改;

郑轩: 数据收集,论文修改;

王贤文: 提出研究思路,设计研究方案,数据收集,论文修改。

支撑数据

支撑数据可开放获取,获取地址为: <https://doi.org/10.6084/m9.figshare.20224638.v1>。

1、王贤文,谢维熙. bioRxiv 历年发文国家数量统计.xlsx. 国家统计。

2、王贤文,谢维熙. bioRxiv 历年发文机构数量统计.xlsx. 机构统计。

参考文献

- [1] 李秋实, 陈传夫. 推进中国开放科学发展的政策环境建设 [J]. 图书情报知识, 2020(3): 11-21. (Li Qiushi, Chen Chuanfu. Promoting the Policy Environment Development of Open Science in China [J]. Documentation, Information & Knowledge, 2020(3): 11-21.)
- [2] 王芳, 安璐, 黄如花, 等. 突发公共卫生事件中的科学应对与思考: 图情专家谈新冠疫情 [J]. 图书情报知识, 2020(2): 4-14. (Wang Fang, An Lu, Huang Ruhua, et al. Scientific Response and Reflection of Public Health Emergencies: An Interview with LIS Experts on COVID-19 [J]. Documentation, Information & Knowledge, 2020(2): 4-14.)
- [3] Nabavi Nouri S, Cohen Y A, Madhavan M V, et al. Preprint Manuscripts and Servers in the Era of Coronavirus Disease 2019 [J]. Journal of Evaluation in Clinical Practice, 2021, 27(1): 16-21.
- [4] Evans J A, Reimer J. Open Access and Global Participation in Science [J]. Science, 2009, 323(5917): 1025.
- [5] 温亮明, 李洋, 郭蕾. 国内外开放科学的实践进展与未来探索 [J]. 图书情报工作, 2021, 65(24): 109-122. (Wen Liangming, Li Yang, Guo Lei. The Practice Progress and Future Exploration of Open Science at Home and Abroad [J]. Library and Information Service, 2021, 65(24): 109-122.)
- [6] Goldschmidt-Clermont L. Communication Patterns in High-energy Physics [J/OL]. High Energy Physics Libraries Webzine (2002-03-20) [2022-04-04]. <https://cds.cern.ch/record/546422/files/sis-2002-163.html>.
- [7] Kirkham J J, Penfold N C, Murphy F, et al. Systematic Examination of Preprint Platforms for Use in the Medical and Biomedical Sciences Setting [J]. BMJ Open, 2020, 10(12): e041849.
- [8] Sever R, Eisen M, Inglis J. Plan U: Universal Access to Scientific and Medical Research via Funder Preprint Mandates [J]. PLoS Biology, 2019, 17(6): e3000273.
- [9] 张智雄. 发展学界自治的预印本交流模式 [J]. 数字图书馆论坛, 2017(10): 1. (Zhang Zhixiong. Developing the Preprint Exchange Mode of Academic Autonomy [J]. Digital Library Forum, 2017(10): 1.)
- [10] Fu D, Hughey J. Releasing a Preprint is Associated with more Attention and Citations for the Peer-reviewed Article [J]. eLife, 2019, 8: e52646.
- [11] Fraser N, Momeni F, Mayr P, et al. The Relationship between bioRxiv Preprints, Citations and Altimetrics [J]. Quantitative Science Studies, 2020, 1(2): 618-638.
- [12] Sarabipour S, Debat H J, Emmott E, et al. On the Value of Preprints: An Early Career Researcher Perspective [J]. PLoS Biology, 2019, 17(2): e3000151.
- [13] Sever R, Roeder T, Hindle S, et al. bioRxiv: The Preprint Server for Biology [OL]. bioRxiv Preprint, bioRxiv: 833400, 2019.
- [14] Fleerackers A, Riedlinger M, Moorhead L, et al. Communicating Scientific Uncertainty in an Age of COVID-19: An Investigation into the Use of Preprints by Digital Media Outlets [J]. Health Communication, 2022, 37(6): 726-738.

