

# 科研论文在社交网络中的关注度分析\*

## ——以PLOS生物医学论文为例

田玥 贺培凤 邵杨芳 吴胜男 张昆 于琦  
(山西医科大学管理学院, 太原 030001)

**摘要:** 运用Altmetrics数据分析生物医学论文在网络中的关注度, 探索影响论文网络关注度的因素。以8 322篇论文连续200天的PLOS ALMs数据和其题录信息为数据, 运用Python自编程序从论文关注过程和关注度差异两个维度进行分析。关注过程主要从论文在不同媒介上的扩散趋势、覆盖率变化趋势和单日访问量进行分析。关注度差异主要是运用特征分数和尺度、词频-逆文档频率方法对关注度差异背后的原因进行探索。从关注过程来看, 生物医学论文在不同媒介上的扩散趋势和覆盖率增长趋势各有不同。论文发表后3天是关注的巅峰期, 发表后7天内是关注的黄金期。从关注度差异来看, 生物医学论文的研究主题、国家/地区、来源期刊均会影响论文在网络中的关注度。研究主题与人类健康迫切相关的论文得到较高的网络关注度, 如传染性病毒。与美国、英国相比, 我国的论文获得网络关注度较低。论文所在期刊的影响因子越高, 论文越有可能获得较高的网络关注度。论文作者的影响力对论文在网络中的关注度不会产生影响。本研究为科研学者展示了生物医学论文在社交网络中的关注过程并分析影响关注度的可能因素, 期望能为科研学者的研究成果更快速地扩散和传播提供依据和参考。

**关键词:** 生物医学论文; 网络关注度; Altmetrics; PLOS ALMs; Counter

**中图分类号:** G353.1

**DOI:** 10.3772/j.issn.1673-2286.2019.04.008

生物医学论文的扩散, 意味着有更多的人获得、认识、理解和受惠于医学知识, 有助于大众健康素养的提高和医学知识及时应用于临床工作。研究生物医学论文在网络中的关注度, 可以为促进生物医学论文扩散、提升论文扩散效率提供参考依据, 拓宽知识扩散渠道, 以充分实现生物医学科研成果的价值。传统的研究方法中, 生物医学论文的出版、订阅和引证等指标常被用来研究论文的扩散。随着数字学术出版革命的兴起和推进、网络媒体的普及, 以及开放获取 (Open Access) 运动的出现, 生物医学论文的扩散渠道呈现多样化趋势, 扩散所用平均时间不断缩短<sup>[1]</sup>。生物医学论文在网络中经电子出版后, 通过各类媒介向外扩散, 不但丰富了扩散的形式和渠道, 也留下了种类繁多、条目细分、数据庞大的网络扩散数据 (即Altmetrics数据), 这些新的知识传播和扩散媒介以及所产生的扩散数据, 为

研究生物医学论文的扩散提供新视角、新途径。

目前, 不乏关于Altmetrics数据的研究。金贞燕等<sup>[2]</sup>分析了Altmetrics数据整合工具的现状特点, 研究可以帮助学者更好地理解、把握Altmetrics数据的应用。Shu等<sup>[3]</sup>研究发现相同时间发表于同一期刊的论文, 被Twitter转发的论文比未被Twitter转发的论文增加大约30%的被引频次。魏绪秋等<sup>[4]</sup>基于动态Altmetrics数据对学术论文持续关注度进行研究, 证实大部分学术论文社交活跃期相对较早, 少部分学术论文社交活跃期相对较晚。Wang等<sup>[5]</sup>追溯了学术论文在社交媒介中的扩散踪迹, 发现科学论文在发表后的很短时间内能够迅速在社交媒体上引起关注, 但是关注度的持续时间较短。王真等<sup>[6-7]</sup>基于38篇生物医学论文的浏览量数据对论文的网络传播特征及规律进行了探讨研究, 并对论文在网络媒体中浏览量的累积规律进行了探索。

\*本研究得到国家自然科学基金项目“基于多元分析的科研文献微观实体评价理论与实证研究——以生物医学为例” (编号: 71573162) 和2018年山西省研究生教育创新项目“网络社交媒体中生物医学学术知识的扩散规律研究” (编号: 2018SY040) 资助。

现有的研究已取得一定的研究成果,可为后续的Altmetrics数据研究提供有益参考。但从上述相关研究的数据量来看,Altmetrics数据的数量级普遍较小,而且所分析的数据基本为横断面静态数据,其研究结论在大样本、持续动态的Altmetrics数据上是否同样适用尚待验证。为此,本研究获取10 903篇论文的面板数据,对论文发表后随时间的关注过程以及论文网络关注度的影响因素进行研究。选取美国科学公共图书馆(Public Library of Science, PLOS)平台7种期刊在2016年12月1日—2017年5月31日发表的论文,使用R语言编写程序,获取每篇论文自发表第2天开始连续200天的ALMs指标。运用自编的Python程序对论文发表后的网络关注过程进行分析,包括论文在不同平台上的扩散趋势、覆盖率变化趋势及单日访问量分析。运用特征分数和尺度(Characteristic Scores and Scales, CSS)及词频-逆文档频率(Term Frequency-Inverse Document Frequency, TF-IDF)方法对论文在网络中的关注度差异进行探索。借鉴论文被引量的影响因素,结合可能会影响论文在网络关注度的因素,综合考虑后,选取4个因素进行分析,包括论文的研究主题、来源国家/地区、期刊的影响因子和作者的影响力。

## 1 数据来源与方法

### 1.1 数据来源

PLOS在网站上提供了一个可以获取论文层面计量指标的资源——ALMs(Article-Level-Metrics),它为每一篇论文提供在线使用、引用、博客、媒体报道、评论活动、社会书签、星级评分、专家推荐等信息<sup>[8-10]</sup>。PLOS记录了论文在22种传播媒介中的扩散数据,包括Twitter、Facebook、Mendeley、PMC、Citeulike、Crossref、Scopus、Figshare、Counter<sup>[11]</sup>等媒介。

编程获取2016年12月1日—2017年5月31日发表于PLOS平台7种期刊上的文献数据。自每一篇文献发表后第2天开始,连续200天追踪获取其每天的ALMs数据。数据下载时间为2016年12月1日—2017年12月18日每天21:00~24:00。由于PLOS平台系统原因,1 423篇文献在其发表初期没有发布和更新ALMs数据(见表1),剩余11 666篇文献的ALM数据被完全记录。为避免由于文献类型引起的偏移,本研究选取论文和综述作为分析对象,共计10 903篇。根据PLOS平台的分类标准,

将10 903篇论文分为生物医学论文(8 322篇)和非生物医学论文(2 581)两类。表2展示了论文在7种期刊中的分布情况。

表1 发表初期没有发布和更新ALMs数据的论文数量

论文发表时间	论文数/篇	论文发表时间	论文数/篇
2016年12月9日	316	2017年3月2日	122
2016年12月13日	78	2017年3月3日	109
2016年12月14日	101	2017年3月6日	93
2017年1月26日	169	2017年4月19日	113
2017年2月7日	60	2017年4月20日	188
2017年2月22日	74		

表2 论文在PLOS期刊上的分布信息表

期刊名称	生物医学论文数/篇	非生物医学论文数/篇	时间/年	2017年影响因子
<i>PLOS One</i>	7 201	2 276	2004	2.766
<i>PLOS Genetics</i>	165	130	2005	5.540
<i>PLOS Pathogens</i>	268	28	2005	6.158
<i>PLOS Computational Biology</i>	157	108	2005	3.955
<i>PLOS Biology</i>	62	29	2003	9.163
<i>PLOS Neglected Tropical Diseases</i>	400	8	2007	4.367
<i>PLOS Medicine</i>	69	2	2004	11.675

本研究选择Web of Science数据库,下载8 322篇生物医学论文的题录信息,包含标题、作者和国家等。选择PubMed数据库,获取全部文献的题录信息。数据下载日期为2018年1月31日。其中,将PubMed数据库中2016—2017年的2 412 042篇论文摘要作为主题分析中TF-IDF指标的语料库。

### 1.2 方法

#### 1.2.1 特征分数和尺度

CSS由Glänzel等<sup>[12]</sup>于1988年提出,是根据论文被引量大小对论文进行分类的一种方法。本文利用该方法按Counter值将8 322篇生物医学论文分为3类,具体步骤如下。

第一步:计算全部论文Counter值的均值,记为M1;

第二步:将Counter值<M1的论文归为第3类,即低关注度论文;

第三步: 对于Counter值 $\geq M1$ 的论文, 计算其Counter值均值, 记为M2;

第四步: 将 $M1 \leq \text{Counter值} < M2$ 的论文归为第2类, 即中关注度论文;

第五步: 将Counter值 $\geq M2$ 的论文归为第1类, 即高关注度论文。

### 1.2.2 词频-逆文档频率算法

TF-IDF是一种统计方法, 可以用来评估一个词汇对于一个文档集或语料库中某个文档的重要程度。字词的重要性随着它在文档中出现的次数成正比增加, 但同时会随着它在语料库中出现的频率成反比下降。本文运用该算法对论文的标题进行分析, 以期获得重要的词代表论文的研究主题。计算公式如下。

$$TF(t) = \frac{\text{词汇}t\text{在文档中出现的次数}}{\text{文档的总次数}} \quad (1)$$

$$IDF(t) = \log_e \frac{\text{语料库文档总数}}{\text{包含词汇}t\text{的文档数}-1} \quad (2)$$

$$TF-IDF = TF(t) \times IDF(t) \quad (3)$$

## 2 生物医学论文网络关注过程分析

生物医学论文在不同类型社交媒介上的扩散从不同侧面体现了论文的社会影响力。本文分别选取Twitter、Mendeley、PMC和Counter 4种媒介研究8 322篇生物医学论文在大众社交媒介、学术社交媒介、使用统计类和网络综合使用类上的扩散趋势及效果(见表3)。

表3 PLOS平台ALMs的4种来源媒介信息

PLOS ALMs	类别名称	含义	PLOS分类
Twitter	大众社交媒介	大众社交媒体提及数据	讨论
Mendeley	学术社交媒体	学术社交媒体中的行为数据	保存
PMC	使用统计类	数据库中所记录的使用数据	浏览和下载
Counter	网络综合使用类	网络在线使用数据	浏览和下载

Twitter是大众社交媒体的代表, 面向学术人员和公众, 该平台上的论文主要通过分享、评论、转发等形式传播。Mendeley是学术社交媒体的代表, 主要面向学术人员, 该平台上的论文主要通过被使用者保存的

形式传播。Mohammadi等<sup>[13]</sup>研究表明, Mendeley值能揭示一些科研论文的隐性影响。PMC是使用统计类指标, 该平台上的论文主要通过浏览和下载的形式传播。Counter在测度论文网络在线使用方面具有权威性<sup>[14]</sup>, 是指论文在该平台上被浏览、被下载等形式传播的次数按一定算法计算后得到的指标, 是用于综合反映一篇论文在社交网络中的扩散效率指数。

### 2.1 不同媒介中生物医学论文的扩散趋势分析

8 322篇生物医学论文和2 581篇非生物医学论文在4种媒介上随论文发表天数增加而累计的访问量曲线如图1所示。显然, 在Counter、Mendeley、Twitter和PMC 4种媒介上, 论文的累计访问量增长呈现出一定的规律性。在Counter、Mendeley和Twitter媒介上, 论文发表后迅速受到众多学者和社会大众的关注, 随着发表时间的增加, 曲线开始呈现趋缓的发展态势。这是因为新发表的论文内容新颖, 很快进入学者、社会公众视野; 但随着时间的推移, 论文的新颖性开始降低, 再加上受新发表论文的冲击, 访问量增加的幅度开始变小。而在PMC上, 由于论文网络出版与数据库收录存在一段时间差, 所以, 前期的访问量为零, 后期医学科研人员根据实际需要检索文献, 累计访问量呈持续快速增长。

在Counter、Mendeley、Twitter 3种媒介上, 生物医学论文获得的访问量低于非生物医学论文。这说明对于没有医学背景的大众而言, 医学论文比较难以理解, 影响了生物医学论文的向外扩散, 导致生物医学论文获得的访问量低于非生物医学论文。而在媒介PMC上, 生物医学论文访问量高于非生物医学论文。这说明, PMC作为生物医学和生命科学期刊全文数据库为主的数据库, 是医学科研人员获取医学知识的首选媒介。

### 2.2 不同媒介中生物医学论文覆盖率分析

计算生物医学论文在发表后的每天分别被Counter、Mendeley、Twitter和PMC 4种媒介关注的论文数占总论文数的比值, 得到发表后不同时间论文在每个平台的覆盖率。如在Mendeley媒介上, 论文发表后第3天的覆盖率是15.61%, 表明在8 322篇论文中有1 299篇论文被Mendeley保存。图2显示了8 322篇论文发表后200天内, 在4种平台上的覆盖率随时间的变化趋势。

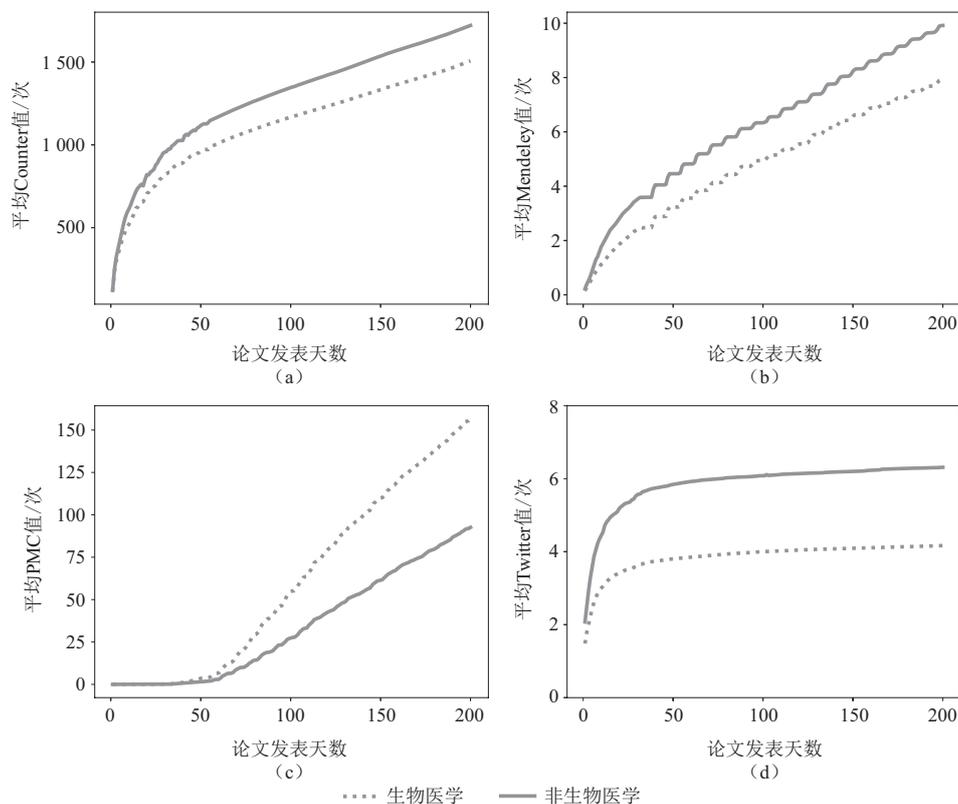


图1 生物医学论文和非生物医学论文在Counter、Mendeley、PMC和Twitter上的扩散趋势

在Mendeley媒介上,论文的覆盖率呈持续增长的趋势。这是由于Mendeley媒介上集聚的用户以科研人员为主,科研人员可能根据科研需求对论文进行阅读,兴趣比较持久,既关注新发表的论文也关注往期论文。刘晓娟等<sup>[15]</sup>也证实了Mendeley对论文的关注是一个持续的过程。在Twitter媒介上,论文的覆盖率初期较高,后期基本保持不变。这是由于Twitter是一种大众化的媒体,关注的用户以大众为主,大众关注的是最新发表的论文。Haustein等<sup>[16]</sup>在2014年的研究中发现, Twitter用户更青睐于提及新发表的文献,而不是往期文献。在Counter媒介上,论文发表后第3天有98.62%的论文被浏览或下载。在PMC上,论文发表后第100天有95.83%的论文被浏览或下载。这说明在网络社交媒体上,论文在短期内得到了较高的关注,这是传统媒体所不能及的。这也提示科研学者,有效地利用现代化网络化的渠道传播新知识,让更多的人受益,让更多的新成果得到应用。

### 2.3 生物医学论文单日访问量分析

由于PLOS ALMs指标数据首次更新日期为论文发表后第2天(不包含论文发表当天),因此首次更新

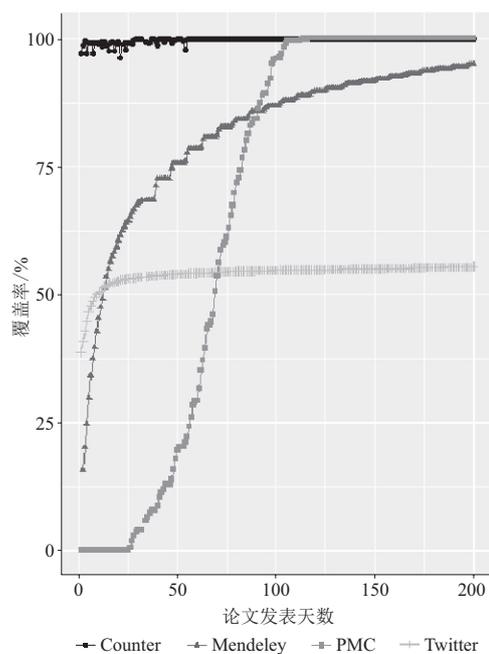


图2 生物医学论文在Counter、Mendeley、Twitter、PMC上的覆盖率随时间的变化趋势

的数据应该是论文发表后第1天和第2天的累计之和。在处理数据时,我们把首次获取的Counter值平均分配至第1天和第2天。如2016年12月1日发表的论文*Do Elite*

and Amateur Soccer Players Outperform Non-Athletes on Neuro cognitive Functioning? A Study Among 8-12 Year Old Children, 2016年12月3日在PLOS ALMs上获得的首次Counter值是38, 则该论文在发表后第1天(12月2日)和第2天(12月3日)的Counter值均为19。

如表4所示, 452篇生物医学论文在发表后第1天或第2天单日访问量达到峰值。1/3(2 814篇)的生物医学论文在发表后第3天单日访问量达到峰值。超过一半(57.28%)论文的单日访问量在发表后7天内达到顶峰。可见, 论文发表后第3天是论文关注的巅峰期, 发表后7天内是论文关注的黄金期。

表4 生物医学论文单日访问量最大时的论文篇数统计表

发表后天数/天	单日访问量最大的论文数/篇	占总数百分比/%	累计百分比/%
1或2	452	5.44	10.88
3	2 814	33.85	44.73
4	390	4.69	49.42
5	268	3.22	52.64
6	172	2.07	54.71
7	214	2.57	57.28
8~15	1 189	14.30	71.58
16~200	2 362	28.42	100.00

### 3 生物医学论文网络关注度差异分析

鉴于生物医学论文关注的巅峰期是发表后的第3天, 本研究首先选用Counter第3天的值作为测度数据, 运用CSS方法将8 322篇生物医学论文分为3类: 高关注度论文, 692篇, Counter值 $\geq 467$ ; 中关注度论文, 2 635篇,  $231 \leq$ Counter值 $< 467$ ; 低关注度论文, 4 995篇, Counter值 $< 231$ 。然后, 从研究主题、来源国家/地区、来源期刊和作者影响力4个维度分析不同关注度论文之间的差异, 探索论文在社交网络中关注度的影响因素。

#### 3.1 不同关注度论文的研究主题分析

本研究选取论文的摘要作为主题分析的对象。对8 322篇生物医学论文的摘要进行切词、归类、移除停用词、取词干后, 计算每个词汇的TF-IDF得分, 经反复实验发现TF-IDF $> 0.3$ 的词汇可以较好地代表论文的主题。分别计算这些词汇在3类论文的摘要中出现的频次(见表5)。

表5 3类生物医学论文中摘要前20个高频词统计表

排 序	高关注度论文		中关注度论文		低关注度论文	
	词 汇	频 次	词 汇	频 次	词 汇	频 次
1	vaccin	312	vaccin	1 069	women	1 976
2	zikh	270	mutat	997	vaccin	1 597
3	mutat	198	cell	805	group	1 016
4	cell	180	network	669	cell	885
5	memori	169	cancer	578	gene	773
6	infect	153	mortal	475	beta	509
7	network	124	infect	400	cancer	483
8	denv	117	variant	368	mice	426
9	mosquito	108	gene	334	children	409
10	mtb	101	mosquito	298	mortal	401
11	cancer	97	eye	273	mutat	384
12	speci	93	women	205	mirna	350
13	transmiss	80	risk	187	eye	321
14	commun	73	children	176	muscl	318
15	commens	69	malaria	148	health	268
16	dengu	52	muscl	126	variant	219
17	chikv	45	dengu	113	tumor	183
18	replic	42	pregnanc	108	task	156
19	microbiota	36	resist	103	ckd	130
20	dog	30	group	95	intervent	125

疫苗(vaccin)、癌症(cancer)、突变(mutat)、细胞(cell)等主题在3类论文中均得到较高的网络关注, 表明这些主题是生物医学领域普遍受欢迎的研究主题。一些与病毒相关的主题, 如寨卡(zikh)、登革热(denv)和基孔肯雅(chikv)出现在高关注度论文中。zikh病毒可造成大批婴儿脑发育不全, denv病毒感染具有高发病率和高死亡率特征, chikv病毒可引发以发热、皮疹及关节疼痛为主要特征的急性传染病, 相关主题的论文获得了大众的高度关注。基因(gene)、儿童(children)、妇女(women)和眼(eye)这些主题出现在中关注度论文和低关注度论文中。健康(health)仅出现在低关注度论文中。

对比3类论文的研究主题可以发现, 除3类论文有共同的研究主题外, 与人类健康迫切相关的研究主题得到高的网络关注, 较为常规、宽泛的研究主题没有得到高关注度。

#### 3.2 不同关注度论文的来源国家/地区分析

本研究统计了每类论文的来源国家/地区的发文量。

国家/地区的发文量根据论文的通信作者所属国家/地区进行计算。如一篇论文有2名通信作者,分别来自中国和美国,那么中国对这篇论文的贡献记50%,美国记50%。本文统计了每类论文中发文量前20的国家。

美国在3类论文中的发文量均排在首位,且在高关注度论文中占比最大,达36.04%。这表明美国的生物医学研究在网络中受欢迎的程度很高,也意味美国在世界医学研究领域占主导地位。英国、加拿大、巴西、瑞士、苏格兰在高关注度论文中的发文量排名较前,在中关注度和低关注度论文中的排名较后。与此相反,中国、日本、韩国、意大利在高关注度论文中排名较后,在中关注度和低关注度论文中排名较前,表明这些国家发表的生物医学论文在网络中受欢迎的程度不高。

### 3.3 论文关注度与来源期刊的关系分析

7种期刊的扩散趋势如图3所示。可以看出, *PLOS Biology* 期刊的论文平均获得的访问量最高, *PLOS Medicine* 期刊次之, *PLOS Pathogens* 期刊和 *PLOS Genetics* 期刊的论文平均获得的访问量相同。对照表2期刊影响因子可以得出,生物医学论文获得的访问量与所在期刊的影响因子之间存在一定的正相关性,论文在网络中的关注度受所在期刊影响因子的影响,但可以实现局部超越。

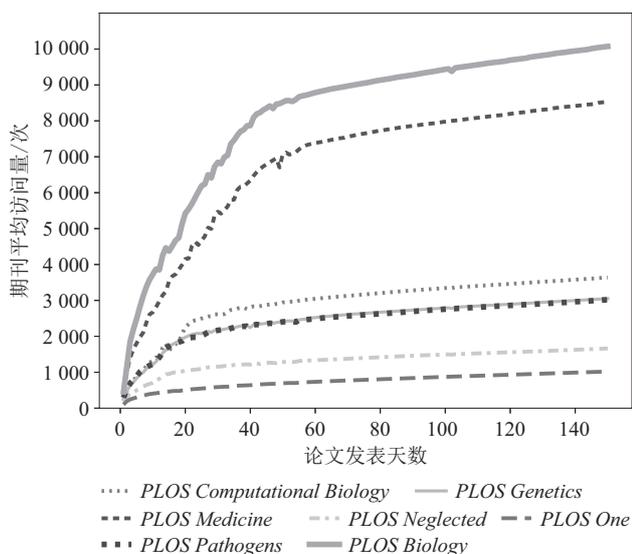


图3 生物医学论文在7个期刊上的扩散趋势

### 3.4 论文关注度与作者影响力的关系分析

本文使用H指数作为作者影响力的测度指标。我们

利用Web of Science下载每篇文章作者发表的论文记录,为减少作者同名对检索结果带来的影响,采用“作者姓名+机构名称”的方式实施检索,并对检索记录进行人工筛选。基于筛选后的数据集,计算每位作者的H指数。对于每篇文章,我们选取第一作者和通信作者的H指数(多个通信作者的选取H指数最大的一个),与第3天Counter值做Spearman相关分析,结果为 $P > 0.05$ ,说明论文第一作者和通信作者的H指数与其Counter值均不存在相关性。这反映出大众在网络中传播学术论文时并不关注论文作者是否具有影响力,论文作者的影响力不是论文网络关注度的影响因素。

## 4 结论

从关注过程看,生物医学论文在Counter、Mendeley、Twitter和PMC 4种媒介上,扩散趋势和覆盖率变化趋势各不相同。因此在扩散论文时,针对不同的媒介,应采取不同的扩散策略。论文发表后第3天是关注的巅峰期,发表后7天内是关注的黄金期。因此要在论文发表初期及时传播论文,使科研成果得到最大限度的扩散。

从关注度差异上看,论文的研究主题、来源国家/地区、来源期刊的影响因子都会影响论文在网络中的关注度,而论文作者的影响力并不会对论文在网络中的关注度产生影响。

(1) 论文的研究主题会影响论文在网络中的关注度。研究主题与人类健康迫切相关的论文获得了高的网络关注,如传染性病毒。研究主题较为常规普通的论文网络关注度较低。

(2) 与美国、英国等国家相比,我国的生物医学论文获得的关注度较低。这个结果可能是由两个因素造成的。首先,尽管近年来我国的生物医学论文质量有很大的提高,但与美国、英国等发达国家相比还有一定的差距;其次,一些主流社交媒体在我国大陆的使用受限,导致我国的科研人员在网上参与度较低。有研究表明<sup>[17]</sup>,论文的扩散首先是从本机构、本国开始的,之后向外扩散。我国的科研人员在网上参与度较低,导致我国的生物医学论文获得较低的网络关注。这提醒我国医学研究人员,要创造条件参与科研网络交流,以促进论文的扩散。

(3) 论文所在期刊的影响因子会影响论文在网络中的关注度。论文所在期刊的影响因子与论文获得的网络访问量存在一定的正相关性,但可以实现局部反超。

(4) 论文作者的影响力不会影响论文在网络中的关注度。高影响力作者的论文与低影响力作者的论文在网络中的关注度不存在显著差异。这说明, 网络社交媒体为论文在网络中的扩散提供了平等机会。

研究展示了生物医学论文网络关注过程, 挖掘了生物医学论文网络关注度的影响因素, 为生物医学论文获得更多的网络关注度提供指导意见。望研究结果能够帮助医学科研学者有效地提高论文在网络中的扩散。

## 参考文献

- [1] TEECED J. Technology transfer by multinational firms: The resource cost of transferring technological know-how [J]. *Economic Journal*, 1977, 87 (346): 242-261.
- [2] 金贞燕, 侯景丽, 孙华丽. Altmetrics数据整合分析工具的现状特点及相关问题研究 [J]. *情报理论与实践*, 2019, 42 (4): 89-95.
- [3] SHU F, HAUSTEIN S. On the citation advantage of tweeted papers at the journal level [J]. *Proceedings of the Association for Information Science and Technology*, 2017, 54 (1): 366-372.
- [4] 魏绪秋, 郭凤娇. 基于动态Altmetrics数据的学术论文持续关注度研究 [EB/OL]. [2019-04-01]. <http://kns.cnki.net/KCMS/detail/11.1762.G3.20190117.1725.004.html>
- [5] WANG X, FANGZ, GUO X. Tracking the digital footprints to scholarly articles from social media [J]. *Scientometrics*, 2016, 109 (2): 1365-1376.
- [6] 王真, 马建华. 单篇论文网络浏览量的跟踪研究 [J]. *中国科技期刊研究*, 2018, 29 (3): 270-277.
- [7] 王真, 马建华. 基于PLOS开放获取数据的单篇论文网络浏览量累积规律的数理统计及分析 [J]. *图书情报工作*, 2018, 62 (12): 72-83.
- [8] BINFLELD P, BEDE G, BAER R, et al. PLoS One: Background, future development and article-level metrics [EB/OL]. [2019-03-31]. <https://core.ac.uk/download/pdf/47275815.pdf>.
- [9] ERDT M, NAGARAJAN A, SIN S J, et al. Altmetrics: an analysis of the state-of-the-art in measuring research impact on social media [J]. *Scientometrics*, 2016, 109 (2): 1117-1166.
- [10] KRAKER P, LEX E, GORRAIZ J, et al. Research data explored II: anatomy and reception of figshare [J]. *Computer Science*, 2015 (3): 77-79.
- [11] Counter指标的介绍 [EB/OL]. [2019-03-31]. <https://www.projectcounter.org/about/>.
- [12] GLÄNZEL W, SCHUBERT A. Characteristic scores and scales in assessing citation impact [J]. *Journal of Information Science*, 1988, 14 (2): 123-127.
- [13] MOHAMMADI E, THELWALL M, HAUSTEIN S, et al. Who reads research articles? An altmetrics analysis of Mendeley user categories [J]. *Journal of the Association for Information Science and Technology*, 2015, 66 (9): 1832-1846.
- [14] HAUSTEIN S, PETERS I, BAR-ILAN J, et al. Coverage and adoption of altmetrics sources in the bibliometric community [J]. *Scientometrics*, 2013, 101 (2): 1145-1163.
- [15] 刘晓娟, 余梦霞. Altmetrics覆盖率影响因素及应用分析 [J]. *图书情报工作*, 2018, 62 (16): 92-101.
- [16] HAUSTEIN S, PETERS I, SUGIMOTO C R, et al. Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature [J]. *Journal of the American Society for Information Science & Technology*, 2014, 65 (4): 656-669.
- [17] WANG X W, FANG Z C, LI Q C, et al. The poor altmetric performance of publications authored by researchers in Mainland China [EB/OL]. [2019-04-01]. <http://xianwenwang.com/>.

## 作者简介

田玥, 女, 1993年生, 硕士研究生, 研究方向: 生物医学知识挖掘。

贺培凤, 女, 1962年生, 硕士, 教授, 研究方向: 医学信息管理。

邵杨芳, 女, 1974年生, 博士, 副教授, 研究方向: 信息组织与语义检索。

吴胜男, 女, 1987年生, 博士, 讲师, 研究方向: 信息计量与知识管理。

张昆, 男, 1996年生, 硕士研究生, 研究方向: 数据挖掘。

于琦, 男, 1982年生, 博士, 副教授, 通信作者, 研究方向: 数据驱动的生物医学知识挖掘, E-mail: yuqi351@gmail.com。

Analysis of the Attention Degree of Scientific Papers on Social Network—Taking Biomedical Papers from PLOS as Example

TIAN Yue HE PeiFeng TAI YangFang WU ShengNan ZHANG Kun YU Qi  
(School of Management, Shanxi Medical University, Taiyuan 030001, China)

Abstract: Using altmetrics data to analyse the attention degree of biomedical papers on social network and explore the factors that affect the network attention of the papers. The 200 consecutive day's PLOS ALMs data for 8 322 papers and bibliographic data are used in this paper. This paper uses Python self-programming program to analyse the attention process and attention degree differences of papers. The attention process includes the diffusion trend

of the paper, the change trend of coverage and the daily visits. The Characteristic Scores and Scales and IF-IDF are used to analyse the attention degree difference and its reasons. In the attention process, biomedical papers have different diffusion trends and increased trends of coverage rate in different media. The 3rd day after the publication of papers was the peak of the attention of the papers, the first week after the publication of papers was the golden period of the attention of papers. For the attention degree differences, the research topic, the country, and the journal impact factor of a paper may affect its attention degree. Papers whose research topics are urgently related to human health obtain high network attention, like infectious diseases. Medical research led by China attracted lesser attention than that by USA and England. The higher the impact factor of the journal in which the paper was published, the paper more likely to obtain high network attention. The influence of author will not affect the attention of papers. This research shows attention process and analyses the possible factors that influence the attention of biomedical papers on the social network, expecting for research scholars' research results provide a foundation and reference for faster diffusion and dissemination.

Keywords: Biomedical Papers; Network Attention; Altmetrics; PLOS ALMs; Counter

(收稿日期: 2019-03-24)

# 2019数字信息资源建设学术研讨会 暨第六届科技文献资源建设研讨会征文通知

国家科技图书文献中心(NSTL)、CALIS管理中心、《数字图书馆论坛》编辑部将联合主办“2019数字信息资源建设学术研讨会暨第六届科技文献资源建设研讨会”，征文有关事项通知如下：

## 一、征文主题：数据融合与开放获取

1. 数字信息资源整合与集成方法
2. 数字信息资源元数据及其标准规范
3. 数字环境下多来源、多类型资源的评估，组织融合及揭示
4. 科技信息资源发现系统与发现服务
5. 开放运动下开放资源建设态势分析
6. 新媒体影响下的学术资源建设方法与实践
7. 数字信息资源版权及其风险防范
8. 大数据时代信息资源建设模式与渠道
9. 大数据时代资源采集人员素质及岗位职责
10. 与信息资源建设相关的其他主题

## 二、征文要求

1. 应征论文须为未公开发表的原创性研究成果或实践总结，格式规范请参考《数字图书馆论坛》投稿要求，征文稿件发送至`ziyuanjianshe@istic.ac.cn`。
2. 所征论文邀请专家进行严格评审，评选出一、二、三等奖。
3. 征文截止日期：2019年6月30日。
4. 具体会议时间和地点另行通知。

## 三、联系方式

联系人：张藤予、丁遒劲

联系电话：010-58882320