

## COVID-19 科普微信公众号与文章影响力及影响因素分析\*

中南大学湘雅公共卫生学院(410008) 郑 铭 李 婕 成佩霞 宁佩珊 胡国清<sup>△</sup>

**【摘要】目的** 探讨 COVID-19 科普微信公众号与文章影响力及影响因素。**方法** 使用基于 Python 语言的 Scrap-Redis 爬虫程序抓取科普类微信公众号于 2019 年 12 月 31 日-2021 年 7 月 31 日发布的文章。经关键词筛选和人工筛选两步确定纳入的 COVID-19 科普文章。通过正则匹配文章的网页源代码,获取文章的标题与正文的编辑情况。采用逐步法构建多元线性回归模型,分别以微信传播指数(Wechat communication index, WCI)和阅读量为因变量,以标准化偏回归系数评价不同影响因素的作用大小。**结果** 共纳入 201 个科普类微信公众号及 22173 篇 COVID-19 科普文章。科普类微信公众号和 COVID-19 科普文章的影响力均呈右偏态分布。公众号 WCI 和文章阅读量的中位数分别为 239.4 和 101,四分位数间距分别为 215.5 和 408。微信公众号的 WCI 与原创文章数( $b'=0.55$ )和开通视频号( $b'=0.23$ )呈显著的正相关( $P<0.01$ )。科普文章的阅读量与文章来源公众号的 WCI( $b'=0.78$ )、消极情绪的标题( $b'=0.02$ )、标题使用感叹号( $b'=0.01$ )、文字加粗( $b'=0.04$ )、段落居中( $b'=0.06$ )、使用分隔线( $b'=0.05$ )和插入音乐( $b'=0.03$ )等因素呈正相关( $P<0.05$ ),但与文章使用表格( $b'=-0.10$ )呈负相关( $P<0.01$ )。**结论** 高影响力公众号与发布较多原创文章相关,与 COVID-19 科普文章较高阅读量有关的因素包括消极情绪的标题、标题使用感叹号、文字加粗、段落居中、插入视频和音乐。

**【关键词】** COVID-19 微信公众号 科普文章 Python

**【中图分类号】** R193 **【文献标识码】** A **DOI** 10.11783/j.issn.1002-3674.2024.01.010

基于新媒体的健康传播是提高我国居民传染病健康素养的有效途径<sup>[1]</sup>。约 82% 的国内网民曾主动查找微信公众号获取健康信息<sup>[2]</sup>,基于微信公众号的健康传播具有重要的公共卫生应用价值。在新冠肺炎流行期间,公众对疫情相关信息高度关注<sup>[3]</sup>,尽管微信公众号能够提供便利的科普渠道,但是由于发布的文章数量庞大,质量良莠不齐,权威、高水平的公众号与科普文章的传播效果可能受到限制。

目前已有许多研究指出微信公众号影响力的影响因素包括文章推送频率<sup>[4]</sup>、推送时间<sup>[5]</sup>、月发文数<sup>[6]</sup>等。既往研究报道,微信公众号文章影响力的影响因素包括文章的原创性<sup>[7]</sup>、标题特点<sup>[8]</sup>、图文数量<sup>[9]</sup>、和文章排版<sup>[10]</sup>等。然而,现有研究存在以下不足:一是主要通过人工筛选文章,所研究文章的数量不超过 1000 篇,无法反映所有相关科普文章的情况;二是既往研究均是针对其他健康问题的科普文章,目前尚没有针对 COVID-19 科普文章影响力的影响因素研究。本研究以科普类微信公众号和 COVID-19 科普文章为研究对象,使用基于 Python 语言的 Scrap-Redis 爬虫框架抓取文章,并采用正则匹配网页源代码的方式获取文章的编辑情况,探索科普类微信公众号及 COVID-19 科普文章影响力的相关因素,为微信平台中的科普创作者提供具体可行的文章编辑和排版建议。

## 对象与方法

### 1. 影响力评价指标的选择

(1) 科普类微信公众号:目前微信公众号的影响

力评价一般采用清博数据平台提供<sup>[11]</sup>的微信传播指数(Wechat communication index, WCI),WCI 从公众号所发布文章的阅读数、在看数和点赞数三个指标综合评价公众号的影响力,取值范围一般在 0~2000 之间。WCI 数值越高,公众号影响力越大。

(2) COVID-19 科普文章:阅读量。

2. 获取科普类微信公众号

(1) 在清博数据平台,以“科普”为关键词对公众号的名称进行检索(检索时间:2021 年 6 月 1 日)。

(2) 剔除部分不符合要求的公众号。排除标准:①过去 6 个月内未发布过文章;②已注销。

(3) 在筛选后的公众号中,随机抽取其中 50% 的微信公众号纳入研究。

3. 获取 COVID-19 科普文章

Python 网络爬虫技术可通过模拟用户网络浏览行为从网页中自动、大量提取信息,适用于互联网中的健康相关信息的分析<sup>[12]</sup>。

(1) 选择基于 Python 开发的 Scrapy 开源框架,引入 Redis 开源框架实现多机分布式爬虫<sup>[13]</sup>,抓取纳入研究的科普类微信公众号的所有历史文章。参照武汉市卫生健康委的首次通告时间<sup>[14]</sup>,抓取时间范围选择为 2019 年 12 月 31 日-2021 年 7 月 31 日。

(2) 筛选 COVID-19 科普文章。第一步,参照宁佩珊等<sup>[15]</sup>提供的新冠流行期间谣言相关的高频词汇,选择“肺炎”、“病毒”、“新冠”、“新型”、“感染”、“传播”、“隔离”作为检索词,筛选出在文章正文中包含以上任意检索词之一的文章;第二步,在关键词筛选的结果基础上,由 3 位研究人员独立根据文章内容进行人工筛选,将人工筛选判断为 COVID-19 科普文章者纳入研究。

\* 基金项目:国家社会科学基金重大项目(20&ZD120)

<sup>△</sup>通信作者:胡国清, E-mail: huguqing@csu.edu.cn

### 4. 影响因素

#### (1) 微信公众号

根据清博指数提供的公众号信息, 纳入因素包括: 是否为原创号、是否开通视频号、是否定期发文、月发文数、原创文章数、发文频率。

#### (2) 文章

①根据张兰等<sup>[9]</sup>的研究结果, 纳入因素包括: 来源公众号的 WCI、标题情感倾向、标题的标点符号使用(感叹号、问号); ②根据微信公众号图文消息编辑栏的内容, 纳入因素: 文字调整(加粗、斜体、下划线、删除线、文字颜色、文字背景颜色)、段落调整(居中、右对齐、首行缩进、两端缩进、调整行间距、调整字间距)及特殊元素使用(序号列表、表格、分隔线、图片、视频、音乐)。采用 SnowNLP 情感分析模型<sup>[16]</sup>分析文章标题的情感倾向。采用 Python 中 re 语句正则匹配获得标题的标点符号、文字调整、段落调整及特殊元素的使用情况。

### 5. 统计学分析

使用中位数和四分位数间距描述纳入研究的公众号与文章的基本特征, 采用逐步法构建多元线性回归模型, 分别以 WCI 和阅读量为因变量, 以标准化偏回归系数评价公众号和文章影响力的影响因素作用大小,  $\alpha_{入} = 0.05$ ,  $\alpha_{出} = 0.10$ 。

## 结 果

### 1. 检索结果与基本情况

通过检索, 共纳入 201 个科普类微信公众号及 22173 篇 COVID-19 科普文章(图 1)。

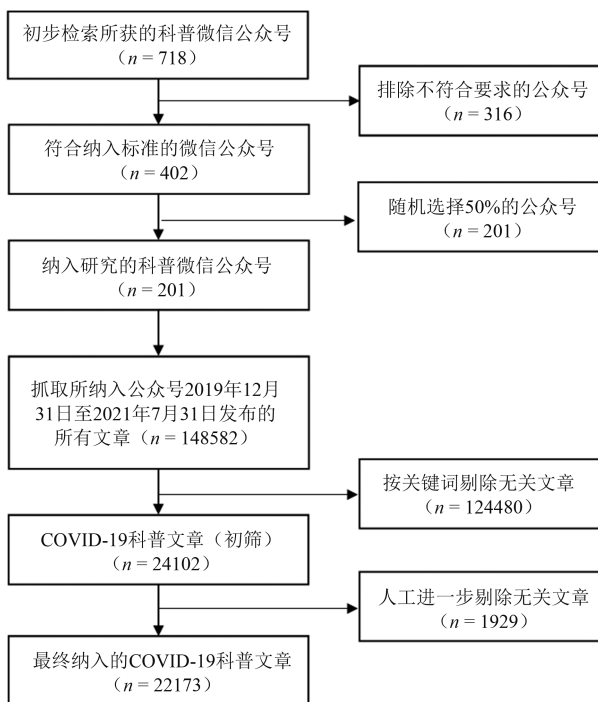


图 1 科普类微信公众号及 COVID-19 科普文章筛选过程

微信公众号 WCI 与科普文章阅读量均呈右偏态分布。纳入研究的科普类微信公众号 WCI 中位数和四分位数间距分别为 239.4 和 215.5, 整体的 WCI 偏低。文章阅读量中位数为 101 次, 四分位数间距为 408。阅读量最大值为 119377。点赞量的中位数为 1 次, 四分位数间距为 3, 文章最高获得了 4187 次点赞(表 1)。

表 1 纳入研究的公众号与文章影响力的基本情况

指标	样本量	中位数	$P_{25}$	$P_{75}$	四分位数 间距	最小值	最大值
WCI	198	239.4	138.4	353.9	215.5	4.2	1667.0
阅读量	21251	101	27	434	408	0	119377
点赞量	21251	1	0	3	3	0	4187

\*: 3 个公众号的 WCI 值缺失, 920 篇文章的阅读量和点赞量值缺失。

### 2. 微信公众号 WCI 与科普文章阅读量的多因素分析结果

根据 WCI 影响因素的分析结果(表 2), 较高 WCI 的科普类微信公众号与“是否开通视频号”和“原创文章数”有关( $P < 0.01$ ), 其中“原创文章数”对影响力的影响最明显( $b' = 0.55$ )。

表 2 微信公众号 WCI 影响因素多元线性回归分析结果

影响因素	$b$	$s_b$	$b'$	$t$	$P$	$b$ 的 95% 置信区间	
						下限	上限
是否开通视频号 (参照 = 未开通)	161.08	39.84	0.23	4.04	<0.01	82.51	239.66
原创文章数	0.37	0.04	0.55	9.73	<0.01	0.30	0.45

\*: 调整  $R^2 = 0.403$ 。逐步法中被排除的因素包括: 是否为原创号、是否定期发文、发文频率。

表 3 显示, 影响科普文章阅读量最主要的因素是文章来源公众号的 WCI( $b' = 0.78$ ), 其他与阅读量呈正相关的因素包括: 采用消极情绪的标题、标题使用感叹号、文字加粗、段落居中、使用分隔线和音乐。其中作用最大的是段落居中( $b' = 0.06$ ), 作用最小的是标题使用感叹号( $b' = 0.01$ )。

与科普文章阅读量有关的因素包括: 标题使用问号、使用下划线、右对齐、首行缩进、调整行间距与字间距、插入表格。其中作用最大的因素是插入表格( $b' = -0.10$ ), 作用最小的因素是调整行间距( $b' = -0.01$ )。

## 讨 论

本研究采用基于 Python 的网络爬虫技术获取 COVID-19 科普文章, 较以往的人工采集信息更为高效, 并且本研究首次使用了审查网页源代码的方法获取了文章被编辑的详细信息, 探讨了与 COVID-19 科普文章、科普类微信公众号影响力有关的影响因素。

在微信公众号的影响力方面, 本研究发现原创文章的数量对科普类微信公众号影响力的作用最大, 原创文章数量越多, 公众号的影响力越大; 开通视频号

表 3 科普文章阅读量影响因素的分析结果

影响因素	<i>b</i>	<i>s<sub>b</sub></i>	<i>b'</i>	<i>t</i>	<i>P</i>	<i>b</i> 的 95% 置信区间	
						下限	上限
来源公众号的 WCI 标题	37.52	0.21	0.78	177.36	<0.01	37.10	37.93
情感倾向得分	-627.36	180.74	-0.02	-3.47	<0.01	-981.62	-273.10
感叹号	283.95	128.29	0.01	2.21	<0.05	32.48	535.41
问号	-578.00	141.79	-0.02	-4.08	<0.01	-855.87	-321.74
文字调整							
使用加粗	1256.72	138.58	0.04	9.07	<0.01	985.09	1528.34
使用下划线	-1037.07	299.90	-0.02	-3.46	<0.01	-1624.90	-449.24
段落调整							
居中	1604.07	127.14	0.06	12.62	<0.01	1354.86	1853.29
右对齐	-588.81	136.25	-0.02	-4.32	<0.01	-855.87	-321.74
首行缩进	-1092.16	124.11	-0.04	-8.80	<0.01	-1335.42	-848.89
调整行间距	-559.85	198.67	-0.01	-2.82	<0.01	-949.26	-170.44
调整字间距	-2193.68	257.48	-0.04	-8.52	<0.01	-2698.37	-1688.99
特殊元素使用							
插入表格	-7933.83	353.30	-0.10	-22.46	<0.01	-8626.32	-7241.33
插入分隔线	1301.46	130.99	0.05	9.94	<0.01	1044.70	1558.21
插入音乐	5625.46	911.71	0.03	6.17	<0.01	3838.43	7412.48

\* : 调整  $R^2 = 0.653$ 。逐步法中被排除的因素包括: 文字调整(斜体、删除线、文字颜色、文字背景颜色)、段落调整(两端缩进)及特殊元素使用(序号列表、图片、视频)。

功能的科普类公众号影响力相对较高。微信平台以低影响力的科普类微信公众号为主,这可能与大部分科普类微信公众号的原创性较差有关。有学者指出,原创性是微信公众号影响力的重要影响因素<sup>[17]</sup>,在纳入研究的公众号中,原创文章数量在 200 篇以下占 86.07%,且仅有 30 个公众号通过了原创号认证。积累原创、优质的科普文章,是科普类微信公众号提高影响力的最重要途径。视频号的直播功能在微信公众号中的嵌入可以极大程度方便用户对信息的获取<sup>[18]</sup>。科普类微信公众号应当抓住当前视频号的发展红利期<sup>[19]</sup>,通过多种途径扩大影响力。本研究也发现公众号影响力与是否定期发文、发文频率的关系无统计学意义。有研究指出,过于频繁的推送文章并不能增加文章的平均阅读量,公众号过于频繁推送文章反而可能会降低用户的关注度<sup>[4]</sup>。

在文章的影响力方面,本研究发现 COVID-19 科普文章的影响力主要受到文章来源公众号自身影响力的影响。公众号的影响力越高,订阅人数越多,相应地越容易出现高影响力的文章。与科普类微信公众号相似,低影响力的 COVID-19 科普文章占到了绝大多数,阅读量在 10000 以下的文章占 99.55%。尽管微信平台中 COVID-19 科普文章数量很多,但其内容大多较为冗余,大部分科普文章并未获得良好的传播效果。与高阅读量有关的影响因素包括:采用消极情绪的标题、标题使用感叹号、文字加粗、段落居中、使用分隔线和音乐。从文章标题的情感倾向来看,传递消极情绪的标题更能吸引人们的阅读。有研究指出在 Twitter 平台中,用户感知到信息的消极情绪与转发行

为有关<sup>[20]</sup>。本研究发现文章标题使用感叹号与较高的阅读量相关,而使用问号则与较低的阅读量相关,这一结果和张兰等<sup>[9]</sup>的研究结果类似。本研究纳入的大部分因素对文章阅读量的影响具有统计学意义,其中对提高文章阅读量作用最大的是段落居中。大部分有统计学意义的影响因素对文章阅读量呈负相关,其中使用表格的负相关程度最大。这可能提示文章的发布者可能无需用到大部分的编辑功能,而是应当更加注意提高科普文章内容本身的原创性和质量。

本研究结果有以下应用价值:①科普类微信公众号运营者应当更注意内容本身的原创性和质量,可以通过开通视频号功能,撰写高质量原创文章来提高公众号的影响力;②文章作者和编辑可通过使用文字加粗、段落居中、插入视频和音乐等手段在一定程度上提高科普文章的影响力。

本研究仍存在一些不足之处。首先,文章的阅读量是一个动态变化的数据,发布较晚文章的阅读量可能会有所增加,尽管文章阅读量通常会在发布 48 小时后进入增长衰亡期<sup>[4]</sup>。总体来看,本研究纳入的绝大多数文章的发布时间都已超过 48 小时,故其对分析结果的影响不大。考虑到语言的复杂性,本研究未对具体文本内容展开深入分析。未来可通过自然语言处理手段,从文章内容的角度展开进一步研究。

参 考 文 献

[1] 晋菲斐,田向阳. 不同健康传播方式对中国居民传染病健康素养影响. 中国公共卫生, 2018, 34(8): 1125-1128.  
 [2] 丁香园. 2019 国民健康洞察报告. (2019-01-10) [2022-03-01]. [https://resource.chemlinked.com/cn/old/cdn/food/file/health\\_](https://resource.chemlinked.com/cn/old/cdn/food/file/health_)

china\_2019.pdf

- [ 3 ] Mheidly N, Fares J. Leveraging media and health communication strategies to overcome the COVID-19 infodemic. *J Public Health Policy*, 2020, 41(4): 410-420.
- [ 4 ] 张玲. 新媒体平台上的科学传播效果分析:以科学类微信公众号为例. *科技传播*, 2018, 10(6): 55-56.
- [ 5 ] 方婧, 陆伟. 微信公众号信息传播热度的影响因素实证研究. *情报杂志*, 2016, 35(2): 157-162.
- [ 6 ] 王雅倩, 曹高辉. 基于双路径视角的图书馆公众号文章传播效果影响因素研究. *图书馆学研究*, 2020(9): 18-26.
- [ 7 ] 钟若曦, 马晓燕, 马文军, 等. 公共卫生类微信公众号文章发布情况与传播效果研究. *中国健康教育*, 2018, 34(9): 800-803.
- [ 8 ] 石婧, 段春波, 周白瑜, 等. 科技期刊应用微博微信平台影响力评价初探. *中国科技期刊研究*, 2014, 25(5): 655-660.
- [ 9 ] 张兰, 陈信凌. 社科类学术期刊微信公众号传播效果影响因素实证研究:以 CSSCI 来源期刊(2019-2020)为例. *中国科技期刊研究*, 2019, 30(9): 1014-1021.
- [ 10 ] 李明德, 高如. 媒体微信公众号传播力评价研究——基于 20 个陕西媒体微信公众号的考察. *情报杂志*, 2015, 34(7): 141-147.
- [ 11 ] 清博指数. 微信传播指数 (v14.2). (2021-06-01) [2022-03-01]. <https://www.gsdata.cn/site/usage>.
- [ 12 ] 周江杰, 王胜锋, 李立明. Python 爬虫技术在信息流行病学中的应用. *中华流行病学杂志*, 2020, 41(6): 952-956.
- [ 13 ] 白鹤, 汤迪斌, 王劲林. 分布式多主题网络爬虫系统的研究与实现. *计算机工程*, 2009, 35(19): 13-16+19.
- [ 14 ] 武汉市卫健委. 武汉市卫健委关于当前我市肺炎疫情的情况通报. (2019-12-31) [2022-03-01]. [http://wjw.wuhan.gov.cn/gsgg/202004/t20200430\\_1199576.shtml](http://wjw.wuhan.gov.cn/gsgg/202004/t20200430_1199576.shtml).
- [ 15 ] Ning P, Cheng P, Li J, et al. COVID-19-Related Rumor Content, Transmission, and Clarification Strategies in China: Descriptive Study. *J Med Internet Res*, 2021, 23(12): e27339.
- [ 16 ] Github. SnowNLP: Simplified Chinese Text Processing. (2020-01-19) [2020-10-25]. <https://github.com/isnowfy/snownlp>
- [ 17 ] 匡文波, 武晓立. 基于微信公众号的健康传播效果评价指标体系研究. *国际新闻界*, 2019, 41(1): 153-176.
- [ 18 ] 薛欢雪, 郭山. 基于 SIR 模型的公共图书馆“三微一端”营销策略研究. *图书馆学研究*, 2021(18): 57-65.
- [ 19 ] 谭春林. 公众号、视频号与微信群协同推动学术期刊的“主动传播”. *编辑学报*, 2021, 33(5): 549-552.
- [ 20 ] Meng J, Peng W, Tan PN, et al. Diffusion size and structural virality: The effects of message and network features on spreading health information on twitter. *Comput Human Behav*, 2018, 89: 111-120.

(责任编辑:郭海强)

(上接第 48 页)

- [ 4 ] 刘倩, 李春燕. 增强子的鉴定及其在肿瘤研究中的应用. *遗传*, 2020, 42(9): 15.
- [ 5 ] Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*, 2015, 12(5): 453-457.
- [ 6 ] Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics*, 2013, 14: 7.
- [ 7 ] Roh W, Chen PL, Reuben A, et al. Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci Transl Med*, 2017, 9(379): eaah3560.
- [ 8 ] Tian W, Chen K, Yan G, et al. A Novel Prognostic Tool for Glioma Based on Enhancer RNA-Regulated Immune Genes. *Front Cell Dev Biol*, 2021, 9: 798445.
- [ 9 ] 严光灿, 田伟, 刘美娜. 基于特异性增强子的低级别胶质瘤分型研究. *中国医院统计*, 2021, 28(6): 502-507.
- [ 10 ] Stintzing S, Modest DP, Rossius L, et al. FOLFIRI plus cetuximab versus FOLFIRI plus bevacizumab for metastatic colorectal cancer (FIRE-3): a post-hoc analysis of tumour dynamics in the final RAS wild-type subgroup of this randomised open-label phase 3 trial. *Lancet Oncol*, 2016, 17(10): 1426-1434.
- [ 11 ] 卢志斌, 廖景升, 刘琦. MTHFR C677T 基因型与结直肠癌患者化疗消化道不良反应的关系分析. *结直肠肛门外科*. 2019.25(1): 62-64.
- [ 12 ] Fauriat C, Long EO, Ljunggren HG, et al. Regulation of human NK-cell cytokine and chemokine production by target cell recognition. *Blood*, 2010, 115(11): 2167-2176.
- [ 13 ] 王鹿鸣, 谢尚. 自然杀伤细胞在头颈部鳞癌发展及治疗中的研究进展. *医学新知*, 2021, 31(1): 33-41.
- [ 14 ] Duan S, Guo W, Xu Z, et al. Natural killer group 2D receptor and its ligands in cancer immune escape. *Mol Cancer*, 2019, 18(1): 29.

(责任编辑:张悦)