

QQ群消息及“代笔”交易的挖掘与学术不端诱因分析

■ 谭春林

收稿日期:2018-12-06

修回日期:2019-05-08

华南师范大学学报编辑部,广东省广州市天河区中山大道西55号 510631

摘要 【目的】分析论文交易的需求特征和学术不端行为的主要诱因。【方法】从QQ群提取2018年9月12日—2018年11月12日的所有聊天记录(4898条),从某电商平台获取“代写”类产品(4473个)截至2019年2月7日20:26的交易记录(174875笔)及部分评价信息等数据。对获取的聊天记录与交易记录消息进行关键词去重、筛选、合并与频次统计,得到信息发布时段分布、需求分类、涉及期刊、“代写”产品的地区分布等统计结果。【结果】QQ群发布论文交易类广告的时段主要集中在工作日的上班时间,内容主要包括论文买卖和期刊级别;某电商“代写”产品累计销售额达769.681万元,其店铺主要分布在北京、广东和上海,其产品主要分为晋升类、工作类和学习类。【结论】学术评价的“唯论文”导向是学术不端行为的直接原因;“代写”产品的“效率高、文笔好、查重率低”是重要诱因;对论文交易双方的惩治较轻、惩治难度较大等导致学术不端行为的实施成本低;期刊编辑随意开具录用通知和社内发票、同意更改作者或署名单位等不规范操作无意间为学术不端打开了“后门”。

关键词 代笔;社交网络;群消息;学术不端;诱导因素

DOI:10.11946/cjstp.201812061076

近年来,国内外学术不端问题越来越严重,其显性、隐性特征及危害引起了学术界与期刊界的广泛关注。刘春明^[1]总结出学术不端有三种常见形式:抄袭与剽窃、捏造与篡改、重复发表;徐婷婷等^[2]指出现有学术不端文献检测系统存在图表数据检测、跨语种检测两大盲区,总结了剽窃、篡改学术论文中的图表(改头换面、局部篡改等)以及跨语种重复发表等“隐性”学术不端现象。除了上述三种学术不端形式,笔者认为代笔(代写)、代理、代投、转让等论文交易行为也是常见的学术不端行为,并且这类行为已经引起了期刊编辑的广泛重视。例如,王雁等^[3]提出了甄别来稿中疑似代写、代投论文的技术手段;王景周^[4]基于采编系统探讨了代投论文的识别方法及其遏制策略。不少研究者^[5-7]对论文代笔现象的原因进行了多层面的分析,但多数研究从学术评价指挥棒、学术道德自律、政策监管、编辑职业操守等角度分析了成因,很少从大数据角度研究“代笔”论文交易乱象及学术不端诱因。

国外关于“代笔(Ghostwriting)”研究的文献较少,在Web of Science数据库的“标题”字段精准检索“Ghostwriting”,共获得83篇文献。1994年Linenthal^[8]发表了《代笔的过去和现在——学术界

的一种疾病》,自此,“Ghostwriting”逐渐引起了学界的注意。这些文献对涉及商业营销行为的“代笔”现象关注较多,笔者将其总结为“医生-药商-学者-出版者”商业营销“代笔”模式。例如,由制药公司策划邀请著名研究学者撰写或“代笔”撰写其药物产品的临床研究报告,通过操纵出版商、同行评审等,利用出版推广药品。大量现代药物的推广都涉及到代笔^[9]。这种模式类似于国内微信营销中的“软文”广告,当然也会涉嫌代理公司与医院或院校作者之间的论文买卖。

目前,国内外从社交群或电商平台交易数据挖掘角度分析学术不端乱象的研究鲜有报道。因此,本研究从QQ群聊天记录及电商“代笔(代写)”产品交易记录大数据中挖掘涉及论文的实际交易信息,从论文交易信息分析论文交易乱象及学术不端的成因,以期防范学术不端行为提供策略和依据。

1 研究对象和研究方法

1.1 样本数据

以社交平台——“万方数据期刊交流”QQ群(群号为133534308)2018年9月12日—2018年11月12日发布的4898条信息文本及其发布者、发

作者简介:谭春林(ORID:0000-0003-3313-6361),博士,编辑,E-mail:tancel@senu.edu.cn。

布时间等信息为研究对象。该群创建于2011年11月16日。截至2018年11月12日,该群有群成员248人,平均每天在线人数约150人,该群基本上无互动性聊天记录,群成员主要发布论文交易类的广告信息,这些样本数据为数据挖掘与数据清洗提供了方便。

以某电商平台上产品标题包含“代写”的4473个产品及其交易记录(单价、交易笔数)、产品所属店铺、店铺所在地等信息为对象,采用分词法分析“代写”产品标题的语义,根据分词语义特征将代写需求分为5个需求层次,并分类统计频次。在确定数据样本的过程中,笔者尝试搜索代写、代发、论文、代投、去重、降重、论文检测、买论文、论文转让等关键词,发现代写涉及到的交易记录最多,覆盖的产品信息最广,因此,本研究以某电商平台上的代写产品交易记录为数据样本。

1.2 研究方法

1.2.1 数据挖掘

在计算机终端QQ软件底部点击“打开消息管理器”,点击“群”,找到“万方数据期刊交流”群,点击右键并在弹出框中点击“导出消息记录”“另存为”文本文件(.txt),完成对该群聊天记录的挖掘。获取的文本文件包含消息文本及其发布时间、网名、QQ号等数据,共计4898条消息记录。

从电商网站搜索“代写”的结果页面,抓取所有“代写”相关产品的基本信息(产品标题、店铺名称、地区等)及交易信息(单价、付款人数),导出到Excel文件中,利用Excel软件对产品标题进行分词处理,在去除重复关键词操作后,提炼出73个关键词,并对产品关键词以及店铺所属地区(31个省市自治区、11个国家)的频次、产品数进行统计,对产品关键词按照代写需求的5个层次进行分类统计。

1.2.2 数据整理与分析

利用EditPlus 3软件和WPS表格对“万方数据期刊交流”群消息记录文本及电商平台代写产品信息文本进行清洗、筛选、分列、分类、关键词提取、去除重复与统计汇总。

以关键词的统计为例:对分列出来的聊天信息记录进行初步浏览,共筛选出29个关键词。根据聊天信息中“《”和“》”之间的内容共提取到56种期刊名刊。利用WPS表格的“COUNTIF()”函数从聊天记录表格列中统计关键词和期刊名称的频次。例如,A列为聊天记录(从A2~A4899共4898条),B

列为关键词,C列为频次。点击C2单元格,在“f_x”公式编辑框中输入统计的函数表达式=COUNTIF(A2:A4899,"*"&B2&"*"),其中“*”为通配符,检查区域为“A2:A4899”。然后双击C2单元格右下角句柄,自动填充其下方的单元格,完成统计。在分列出来的用户时间信息表中,采用“TEXT(C2,"HH")”读取时间(如14:54:48)中的时间段小时数并填入新的表列中,并采用数据透视表功能分别统计不同日期、不同时间段论文交易类广告信息的日均发布频次。其他信息的统计操作步骤与关键词的统计步骤类似。

2 结果与分析

2.1 论文交易类广告的发布频率

为了解“万方数据期刊交流”QQ群中代写、代发、转让等论文交易广告信息的周分布情况,从统计的62天消息记录中,按照星期统计出总频次及平均频次(图1)。从图1可以看出,论文交易类信息的发布主要集中在工作日,特别是在星期三、星期四、星期五,发布总频次均在400条以上。从平均频次来看,工作日和周末的平均频次分别为44条、14条。另外,该群日均发布论文交易广告消息频次的时段分布见图2,总体来看,论文交易类群消息发布时间主要集中在8:00—18:00,发布峰值平均约8次/小时。

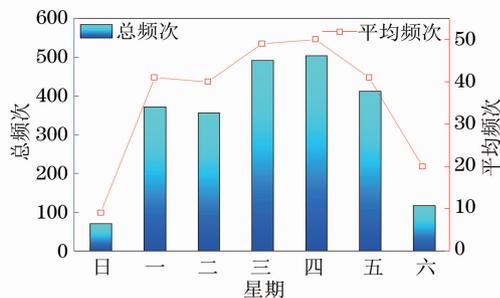


图1 按星期统计论文交易类消息的发布频次

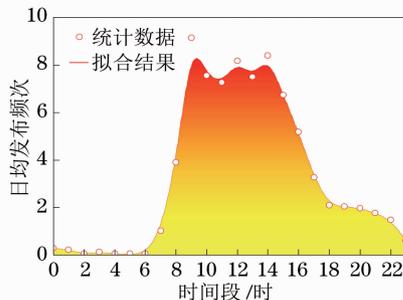


图2 日均发布论文交易广告消息频次的时段分布

2.2 论文交易类广告的关键词分类特征

根据群消息文本,提取出29个关键词,统计

其频次,并将这些关键词按照需求类型分为 5 大类(表 1):级别需求(5862 次)、收录需求(4649 次)、买卖需求(3922 次)、代理需求(3229 次)、时间需求(1121 次)。根据表 1 中具体关键词的统

表 1 群消息中论文交易需求的分类、关键词及其频次

需求类别	关键词(频次/次)	总频次/次
级别需求	国家级(2222)、省级(1350)、普刊(938)、核心(897)、北核(262)、南核(193)	5862
收录需求	知网(1495)、万方(1289)、龙源(1223)、维普(642)	4649
买卖需求	代写(1453)、毕业论文(988)、过稿转让(570)、大量接单(497)、专利(223)、征集*主编(72)、图书(57)、职称(42)、课题(20)	3922
代理需求	社内直收(1342)、修改(817)、全科(372)、包版(259)、论文发表(205)、杂志社发票(158)、定制(76)	3229
时间需求	见刊快(661)、出刊准时(335)、急(125)	1121



图 3 论文交易类消息的关键词标签云

值得注意的是,在买卖需求类别的信息中,代写(1453 次)、毕业论文(988 次)、过稿转让(570 次)、大量接单(497 次)等关键词反映出的论文买卖现象也不容忽视。另外,除了大量与论文写作和发表相关的信息以外,还有少量涉及专利与软件著作权包办、课题申报代写、主编挂名等涉及学术交易的信息,这些与代写论文一样都属于学术不端的范畴。

在本研究获取的群聊天记录中,收集到部分“过稿转让”论文的题目(表 2)。从隐匿了关键词的题目中依然可以发现,医学和经济领域“过稿转让”问题较严重。

表 2 过稿转让涉及的论文篇名举例

序号	篇名
1	不同手术方式治疗假性胰腺囊肿的效果对比分析
2	专业市场空间***研究——基于***的视角
3	不同**渠道对消费者***的影响机制研究
4	***环境下基于**的整合营销方式探索
5	消费者**、服务**与心理契约
6	椎体成形术、***成形术治疗老年骨质疏松***椎体压缩性骨折临床疗效对照分析
7	盐酸**、神经节苷脂钠同时用于***帕金森病患者治疗认知、运动功能改善效果研究
8	重症颅脑损伤***精细化护理临床效果观察
9	盈余***激励以及***理的相关性
10	审计质***性的相关性分析

“过稿转让”是指代笔公司(或个人)以作者身份向编辑部提出要求并更改、增加作者和机构单位,

计频次绘制出具体需求信息的“标签云”图(图 3),直观地将需求热度显示出来,其中消息发布总频次排名前 5 位的分别是国家级、知网、代写、省级、社内直收。

通过这种方式将“过稿”(录用稿)转让给“论文消费者”。少数编辑部可能遇到过拟刊发稿件的作者要求新增第一作者或通信作者的情况,此操作可能存在“过稿转让”的风险。

2.3 论文交易类信息涉及的刊名

从挖掘的群内消息记录中,提取涉及的相关期刊名称,统计出刊名在记录中出现的次数(频次),结果如表 3 所示。在统计过程中发现,群内发布的广告内容中有少数内容采用“*”隐匿了刊名,也有少数内容虚构了刊名,这些刊名不属于本研究的统计范畴。

由表 3 可知,涉及期刊共计 56 种,其中单刊频次在 300 次以上的期刊有 9 种,分别是《特别健康》《健康必读》《健康大视野》《文体用品与科技》《汽车实用技术》《医药界》《海外英语》《英语画刊》《英语教师》,单刊频次占总频次的比例的平均值(平均热度)约 5.0%,这 9 种期刊的出现频次之和占总频次的比例(热度)约 45.1%;单刊频次在 200~299 次的期刊有 18 种,其平均热度约 2.3%,热度约 41.1%;单刊频次在 199 次以下的期刊共有 29 种,其平均热度约 0.5%,热度约 13.8%。上述调查结果表明,健康、医药、体育及其交叉领域以及英语学科等专业期刊在广告消息中出现的频次以及期刊的种数尤为突出;另外,对于单刊频次相对较少的情况,不排除为临时发布的征稿信息,可能不属于代理广告信息。

2.4 电商“代写”类产品的交易记录统计分析

前述 QQ 群广告信息不涉及具体的交易信息(店家来源、产品标题、单价及交易量等),为了进一步了解“代写”交易的实际情况,笔者从某电商平台获取并统计分析了“代写”类产品的交易情况。在该电商平台商品搜索栏输入“代写”,检索到 4404 个产品,截至 2019 年 2 月 7 日 20:26,累计交易量达 174875 笔,总销售额达 769.681 万元。

表3 群内消息涉及的期刊名称及其出现频次

涉及的期刊名称	频次	涉及的期刊名称	频次	涉及的期刊名称	频次
特别健康	819	医师在线	262	轻工标准与质量	56
健康必读	815	中国医学人文	262	炎黄地理	56
健康大视野	792	石油化工	256	中国科学探险	49
文体用品与科技	726	农家参谋	240	中国蒙医药	44
汽车实用技术	513	体育风尚	234	维吾尔医药	34
医药界	454	中国误诊学杂志	224	商业故事	25
海外英语	311	大众医学	208	幸福生活指南	25
英语画刊	311	药物与人	208	电脑与电信	23
英语教师	311	疯狂英语(理论版)	192	丝路艺术	20
武术研究	294	东方	153	西部论丛	18
教育	276	中国战略新兴产业	145	商品与质量	26
当代体育科技	269	家庭医药	144	区域治理	8
体育时空	269	东方藏品	131	数码设计	5
体育世界	269	中国保健营养	73	中华骨与关节外科杂志	4
田径	269	中国经贸	60	城市周刊	4
运动精品	269	当代体育科技文献通报	58	饮食保健	4
中华武术(研究)	269	电工技术	58	中学生报	2
航空军医	262	校园英语	58	领导科学	2
健康世界	262	英语广场	58		

我国各地区“代写”类产品数量的分布如表4所示,“代写”类产品数量在200个以上的省份(含直辖市)有北京(887)、广东(563)、上海(434)、浙江(328)、江苏(324)、山东(322)、湖北(203);产品数量在100~199个的有四川、河南、湖南、福建4省;其余省市的产品数量均在100个以下。另外,在其他国家中,“代写”类产品数量排名前列的国家为

英国(23)、美国(18)。

根据电商“代写”产品描述信息提取关键词并统计其频次(表5),关键词涉及生活类、学习类、工作类、晋升类代写需求的方方面面,其中后3类关键词的总频次均在3500次以上,关键词“文章”的出现频次(1458)最高,该结果说明文章代写需求最旺盛。

表4 某电商“代写”类产品的数量与地区分布

地区	数量/个	地区	数量/个	地区	数量/个	其他国家	数量/个
北京	887	陕西	84	香港	20	英国	23
广东	563	辽宁	83	云南	17	美国	18
上海	434	安徽	72	贵州	14	澳大利亚	7
浙江	328	河北	71	甘肃	13	日本	4
江苏	324	重庆	67	内蒙古	10	新加坡	4
山东	322	天津	55	新疆	10	加拿大	3
湖北	203	江西	42	海南	6	法国	2
四川	191	黑龙江	28	宁夏	2	意大利	1
河南	157	广西	27	西藏	1	韩国	1
湖南	125	山西	26			德国	1
福建	106	吉林	22			其他	9

表5 电商交易平台“代写”产品关键词及其频次

交易分类	关键词(频次/次)	总频次/次	热度/%
生活类	情书(352)、贺卡(241)、检讨(72)、对联(53)、春联(35)、保证书(35)、诉状(30)、转让(30)、申诉(19)、律师函(15)、贺词(12)、请帖(9)、举报信(6)、喜联(4)、辩护(4)、上访信(2)	919	6.6
学习类	演讲稿(682)、读后感(424)、代做(412)、程序(314)、剧本(253)、推荐信(199)、笔记(139)、体会(132)、作文(129)、散文(114)、诗词(114)、帮写(113)、脚本(108)、代码(108)、解说词(106)、代抄写(77)、通讯(51)、美术(26)、辩论稿(12)	3513	25.4
工作类	报告(668)、总结(657)、软文(570)、代笔(478)、策划(364)、PPT(272)、翻译(215)、计划书(195)、新闻稿(176)、发言稿(175)、教案(120)、公文(106)、述职报告(102)、调研(93)、合同(89)、汇报(79)、主持稿(49)、简报(35)、投标书(33)	4476	32.3
晋升类	文章(1458)、原创(883)、征文(400)、英文(387)、留学(384)、润色(345)、简历(260)、写手(229)、创意(132)、毕业(110)、申报(64)、检测(55)、统计(49)、隆重(47)、电大(45)、规划(40)、MBA(30)、职称(26)、评优(3)	4947	35.7

注:热度为关键词频次占总频次的比例;统计时间为2019年2月7日20:26。

3 讨论

3.1 学术评价的“唯论文”导向是直接原因

首先,大学生各类论文需求催生了论文代写市场。大学生的研究生保送申请、各类奖项评比、创新项目申报与结题,以及毕业答辩等均对发表论文级别和数量有要求。由表 5 可知,大学生对代写的需求主要涉及文章、留学、简历、毕业、润色、降重等方面。中国高校传媒联盟针对论文代写现象展开了调查,随机抽取了 318 名大学生,其中,31.1% 的大学生表示考虑过找人代写,73.9% 的大学生身边出现过论文代写现象。由此可见,大学生已成为“代写”市场的主要消费群体和学术不端的重灾区^[10]。

其次,在各类行政职务、岗位的晋升、竞选过程中,唯论文、唯职称、唯学历、唯奖项的“四唯现象”不仅扩大了论文需求量,而且最易诱发论文交易行为。虽然某些行政岗位的竞选没有明确要求有高层次论文,但实际上,在竞选者都具备各种硬实力(项目、论文或获奖)的情况下,大多数竞选以论文成果竞争力强者胜出。迫于“四唯”的无形压力,不少人心存侥幸,求助代理机构代写代发论文。

有关研究表明:缺乏科学公正的学术评价和考核制度,以及监督制度的不完善是造成学术不端的主要环境因素^[11],因此,打破“唯论文”的学术评价机制,可望从根本需求层面大幅降低论文交易乱象程度。2018 年 10 月,国家科技部、教育部、人力资源和社会保障部、中国科学院、中国工程院五部门联合通知并开展了清理“四唯”专项行动,从某种程度上说,这一举措正是针对“四唯”导致的论文交易乱象这一社会问题提出的专项行动。这一专项行动的开展,将会在很大程度上整治学术不端问题。

3.2 代写的高效率、好文笔、低查重率具有吸引力

在电商销售记录的评价消息中,针对顾客购买“代写”产品的动机,统计部分关键词的频次发现:13% 的顾客以急、忙、懒为理由选择购买“代写”产品;27% 的顾客因为“写不出来、没头绪”选择“硕博团队、专业写手”代写;58% 的顾客评价写手“文笔好、质量高、速度快”,能“解决”问题,并“顺利通过”;也有 2% 的顾客只是“尝试一下”寻求代理,但最终“还没有自己写得好”而感到“失望”。

总体来看,效率高、文笔好、查重率低,能为顾客解决大问题,这是“代写”产品的重要卖点,更是引发学术不端行为的重要诱因。

3.3 相关部门的监管乏力导致学术不端犯罪成本低

大多数情况下,具有剽窃行为或通过代写交易发表的论文,在发表后的 1~2 年内,才被数据库出版商(如中国知网等)检测为高相似度的论文(甄别滞后)。而这对于这些论文作者来说,其晋升、评职或毕业目的早已达到,不会明显影响后续工作(影响小)。

对于“专业写手”或大学生“兼职写手”而言,他们利用自己的写作或专业优势,将所谓的知识或剽窃篡改技巧转化为财富,没有直接风险,也乐此不疲。据《中国青年报》报道^[10],不少大学生、研究生投身到代写行业,从中赚取生活费,并且有大学生做了代理。代写电商拥有一条组织严密的代写产业链,大多数商铺声称“查不出来的,放心下单”。这种“自信”从某种意义上说明了目前对代写产业链的打击存在短板。

长期以来,相关部门对于“代写”买卖双方的学术不端行为监管与防治乏力,导致学术不端行为的实施成本低。

而对于受害者之一的编辑部而言,一方面只能发布撤销申明,并在各大收录数据库撤销、删除相应论文^[12];另一方面,只能通过电话或者书面文件等向涉事作者所在单位人事管理部门申明其学术不端行为,以及建议该单位作相应处罚(如撤销职称等)。有研究表明,论文对科学界的影响不因撤销而轻易消退,科学界自我净化的时间成本很高^[13]。另外,在实际过程中,这种不良影响和论文痕迹难以被彻底消除,例如,很难从互联网的庞大社交网络、各类文库(如百度文库、爱学术、豆丁网、道客巴巴等)中删除这些学术不端论文。由于相关部门的监管乏力,涉事作者应该承受的处罚往往因各种原因遭遇“执行难”。这种大环境进一步降低了论文抄袭、代写、买卖等学术不端行为的实施成本与风险,难以对学术不端行为人的发展造成负面影响,消费者对论文交易等学术不端行为无所顾忌,共同导致学术不端行为屡禁不止。因此,对学术不端行为的监管与防治乏力,导致学术不端行为的成本低廉,甚至在一定程度上放任了论文交易市场的“恶性”发展。

鉴于此,相关部门或者期刊界有必要建立一个权威的学术诚信监控预警平台,借鉴国际上治理学术不端行为的政策与策略^[14]。一方面,对学术不端行为的涉事作者进行实名记录,为毕业答辩、职称、项目、招聘、晋升等事项组织单位提供审查通道;另

一方面,严查、严打代写、代发、转让等代理公司及个人,严查社交网络媒体的交易广告信息,提供举报通道,全民监督。

3.4 办刊的不规范无意间为学术不端敞开“后门”

在期刊的编审流程中,一方面,目前大多数采编系统存在技术上的漏洞(例如无验证码登录等),使代写代理公司或个人可以通过他们的软件快速注入式自动投稿;另一方面,编辑热心为作者行方便(如同意变更第一作者和通信作者、同意变更文章发表刊期、随意开具录用函和“社内发票”等),这为代写代发公司或个人提供了可乘之机。深谙各期刊编辑部“习性”的代理公司或个人,利用这些由期刊编辑无意间留下的“后门”,研究出了一系列成熟的、可助力成功发表的“渠道”。从前文挖掘的论文交易广告内容看,他们声称能“代写、代发、包修改,成功后再付款”,甚至能将通过终审的录用稿以更换第一作者或新增通信作者的形式转让出去(即“过稿转让”)。

升级采编系统、加强办刊的规范性、重视论文署名的严肃性、建立作者贡献声明与监督机制、加强初审和审稿环节的把关^[15]、提高甄别学术不端论文的能力等,有利于从源头阻断论文交易乱象。

4 结论

通过对“万方数据期刊交流”QQ群涉及论文交易类广告消息(4898条)、电商“代写”类产品(4473个)交易记录(174875笔)及其评价信息等数据进行提取、整理与统计,分析了论文交易乱象的主要诱因。

经过对信息数据的获取、整理与分析,主要得到以下结论。

(1) 论文交易类广告的发布时间主要集中在工作日,且集中在每天8:00—18:00,发布峰值达8次/小时,这不仅打扰了群内成员的学习和工作,而且在不断强化论文交易的欲望。

(2) 在论文交易类广告中,论文交易的需求主要分为:级别需求、收录需求、买卖需求、代理需求、时间需求,其广告信息的高频关键词主要集中在前三种需求上。“代写”“毕业论文”“过稿转让”等关键词的出现频次相当高,其中“过稿转让”现象需引起编辑部高度重视。涉及的期刊主要集中在健康、医药、体育及其交叉领域以及英语学科等专业期刊。

(3) 在某电商“代写”类产品交易记录中,截至2019年2月7日20:26,累计交易量达174875笔,

总销售额达769.681万元。“代写”产品商铺主要分布在北京、广东、上海,其次分布在浙江、江苏、山东、湖北等省。“代写”产品关键词可分为晋升类、工作类、学习类、生活类,其中晋升类(35.7%)、工作类(32.3%)和学习类(25.4%)的“代写”产品占比比较大。

对数据结果进行了讨论分析,总结出学术不端行为的诱因如下:(1)学术评价的“唯论文”导向是直接原因,但2018年国家出台的清理“四唯”专项行动有望整治学术不端问题;(2)“代写”产品具有的“效率高、文笔好、查重率低”等特点,是学术不端行为的重要诱因;(3)不端论文的甄别滞后、对学术不端行为人的负面影响小、代写产业链的组织严密、对学术不端行为的监管与惩治乏力等导致学术不端行为的实施成本低;(4)建议相关部门或期刊界建立一个权威的学术诚信监控预警平台,为各类学术评价提供审查通道、为公众提供举报通道,以有效治理学术不端乱象;(5)部分期刊的不规范办刊无意间为学术不端敞开了“后门”,特别是随意开具录用通知和社内发票、同意更改作者与单位等不规范操作,为各类代写代理机构提供了可乘之机。

本研究利用大数据挖掘与分析方法,对学术不端现象进行了研究。后续研究将利用Python编程软件对其他相关研究进行更大数据量的挖掘与分析。本研究方法可为期刊编辑从事学术研究提供参考。

致谢 感谢广东省科学技术期刊编辑学会的支持。

参考文献

- [1] 刘春明. 防范学术不端的实践与思考[J]. 中国科技期刊研究, 2019, 30(1): 6-8.
- [2] 徐婷婷, 曹雅坤, 曾礼娜, 等. 关于防范科技论文中“隐性”学术不端行为的建议[J]. 编辑学报, 2018, 30(1): 58-60.
- [3] 王雁, 苟莉, 刘丹, 等. 代写代投来稿的甄别及防范措施[J]. 编辑学报, 2018, 30(2): 171-173.
- [4] 王景周. 代投论文的甄别与防范[J]. 中国科技期刊研究, 2018, 29(6): 557-562.
- [5] 孔凡柱. 论文代写代发成因的多角度分析及对策研究[J]. 淮阴工学院学报, 2014, 23(4): 65-68.
- [6] 刘玉邦, 赵枫岳, 黄君录. 论文代写代发乱象产生根源及其对策[J]. 成都电子机械高等专科学校学报, 2012, 15(3): 89-91.
- [7] 胡泽文, 武夷山. 论文代写发现象之研究[J]. 中国软科学, 2012(7): 78-90.
- [8] Linenthal A J. Past and present. Ghostwriting: a "disease of academia" [J]. *The Pharos of Alpha Omega Alpha - Honor*

Medical Society, 1994, 57(4):40.

- [9] Chen C L. Assessing potential legal responses to medical ghostwriting: Effectiveness and constitutionality [J]. *Journal of Law and the Biosciences*, 2018, 5(1):84-102.
- [10] 朱星宇. 每到三月倍儿忙, 市场交易很红火——大学生论文代写乱象揭秘[N]. 中国青年报, 2015-03-16(12).
- [11] 王育花, 童成立. 科技期刊编辑和审稿专家对学术不端的认知及其防范对策[J]. 中国科技期刊研究, 2018, 29(11): 1127-1135.
- [12] 杨珠. 学术不端论文未被撤销现象及其治理——基于董鹏学术不端事件的个案分析[J]. 中国科技期刊研究, 2018, 29(7):691-697.
- [13] 姚长青, 田瑞强, 杨冬雨, 等. 撤销论文及其学术影响研究[J]. 中国科技期刊研究, 2014, 25(5):595-604.
- [14] 田瑞强, 姚长青, 刘泮颖, 等. 学术不端治理政策及案例计量研究[J]. 中国科技期刊研究, 2018, 29(4):355-361.
- [15] 陈志贤, 黄仲一. 防范学术不端论文出版的立体策略[J]. 中国科技期刊研究, 2017, 28(7):605-609.

Text mining of QQ group messages and ghostwriting trading records and the inducements for academic misconduct

TAN Chunlin

Editorial Office of Journal of South China Normal University, 55 West Zhongshan Avenue, Tianhe District, Guangzhou 510631, China

Abstract: [Purposes] This paper aims to analyze the demand characteristics of ghostwriting trading and the main inducement for academic misconduct. [Methods] All chat records (4898 items) were extracted from QQ group from September 12, 2018 to November 12, 2018. The data such as transaction records (174875) and evaluation information of ghostwriting products (4473 products, as of 20:26 February 7, 2019) were obtained from an e-commerce company. Through keyword de-duplication, filtering, merging, and frequency statistics of chat messages and transaction records data, the statistical results of information release period distribution, demand classification, involving journals, and regional distribution of ghostwriting products were obtained. [Findings] The release time of ghostwriting advertisements mainly focuses on the working day, and the advertisement content mainly focuses on the ghostwriting trading and journal level. The cumulative sales of ghostwriting products in e-commerce have reached 7.69681 million yuan, the stores are mainly located in Beijing, Guangdong, and Shanghai, and the products are mainly divided into promotion, work, and study categories. [Conclusions] The orientation of academic evaluation of "paper first" is the direct inducement for academic misconduct; the high efficiency, good writing, and low rate of checking duplication of ghostwriting products are the important inducement; the low cost of academic misconduct is caused by the lighter punishment and the greater difficulty of punishment for both sides of the ghostwriting transaction; the irregular operation of issuing acceptance letters and invoicing and changing authors and institution, unintentionally opens a "back door" for academic misconduct.

Keywords: Ghostwriting; Social network; Chat message; Academic misconduct; Inducing factor

(本文责编:梁永霞)