

基于数据分析透视网络科普舆情特点

钟 琦 王艳丽*

(中国科普研究所, 北京 100081)

[摘 要] 在海量信息时代, 开展科普工作需要具有针对性, 了解网络科普舆情特点可以让科普工作更加有的放矢。本文从中国科协、中国科普研究所和新华网共同开展的“科普中国实时探针”舆情监测系统数据入手, 分别从不同媒介载体舆情热度对比、同类媒介载体舆情热度对比、科普专题舆情热度对比、舆情热度趋势图规律分析(日、周、月)四个维度进行分析, 对网络科普舆情规律和特点进行了提炼和总结, 力求为科普工作提供参考借鉴。

[关键词] 网络科普舆情 科普工作 数据分析

[中图分类号] G206

[文献标识码] A

[DOI] 10.19293/j.cnki.1673-8357.2016.06.005

根据中国历史文献记载,“舆”字的基本含义是“众”。《左传》“僖公二八年”中说,“舆者,众也”。按照《辞源》的解释,舆情的字面含义,是指“民众的意愿”^[1]。所谓舆情,通俗来说,一般指公众对人、事、物或某种现象的情绪、态度和意见,舆情是民意的反应,对政府决策部门具有一定的推动作用。科普舆情是指公众对科普领域各类信息、人物、事件与现象的情绪、态度和意见,它不仅表现在公众的态度如何,还包括公众重点关注哪类信息或者议题,以及其关注背后的原因是什么。网络科普舆情是指借助互联网,通过连接到网络里的各种设备获取到的相关科普舆情,网络是一种数据获取手段。

网络科普舆情是社会舆情系统的重要组成部分。随着中国经济社会的发展,多元化文化背景逐步确立,公众的个性化需求越来越丰富,经济、文化越发达的地区个性化差异就表现得越强烈,人们的具体要求就越多,表现出

微观和琐屑,对于科普工作者而言,这导致了科普对象的不断复杂,科普需求日趋宽泛和精细,科普工作的开展只有关注到对象需求的多样性和个性特点,在具体工作中做好对象的细分和做到以人为本,才能使科普服务的效果最大化^[2]。大数据时代,网络科普舆情不仅可以快速产生,而且可以快速发酵,进而对科普领域主管部门产生一定的影响力,这种影响力可能是正面推动,也可能是负面施压。因此,及时了解网络科普舆情,对于科普领域主管单位来说意义重大:一方面可以对正面科普舆情进行相关强化与扩散,让其促进科普领域工作;另一方面对于一些带有负面情绪的科普舆情则要找到其背后的深层原因,有针对性地进行疏解,为公众释疑解惑,从而让科普工作做到良性发展。

本文依托全媒体数据支持,通过舆情监测系统对科普领域相关报道及网友评论进行整合分析,形成舆情趋势图。分析结果除了

收稿日期: 2016-09-21

* 通讯作者: E-mail: 491078584@qq.com。

对科普领域相关单位来说有重要价值之外,对于媒体的科学传播工作来说同样意义重大。媒体通过该研究可以及时了解网络科普舆情,从而了解公众关注的信息、话题以及其观点和态度,这对于媒体确定选题、传播科学知识、澄清科学领域谣言以及进行舆论引导都有很大的帮助。此外,网络科普舆情研究还向科普信息化建设的项目承担者、各级科协、学会的工作者提供科普内容的选题以及切入方向的参考,并为科协系统及相关部门提供和科普相关的热点焦点、突发重要新闻预警,支持科协系统工作。

1 网络科普舆情数据分析方法

1.1 网络科普舆情数据平台建设

科普舆情分析需要大量的数据。为了获取数据,2015年中国科协与中国科普研究所和新华网合作,共同建设了科普舆情数据监测系统——科普中国实时探针,通过该平台可以实现对数据的实时获取,从而为后续数据分析服务。

根据网络科普舆情关注重点及科普舆情特点,科普舆情系统监测数据平台共分为八个版块,分别是舆情总览、分析、科普热点、科普关键词、微博监测、负面预警、搜索、报告。“舆情总览”里可以看到科普热度走势图、载体分布图和科普舆情地域分布等内容;“分析”版块里可以通过选择日期看到当时(日、周、月)的舆情曲线图,也可以看到媒体趋势图和科普专题分布等内容,由此可以看到媒体平台信息量分布及热点科普专题等情况;“科普热点”版块则可以按照平台类型及日期查看阅读量大的热门科普新闻;“科普关键词”版块,可以通过选择不同的关键词看到该类型专题的发文数;“微博监测”版块里可以看到微博账号发文总数排行榜及博主发博排行榜;“负面预警”版块里则有一些需要预警的负面信息;“搜索”版块里可以通过日期、标题、作者等信息搜索站内或者全网信息;“报告”版块则可以生成日报或者简报。

目前监测系统主要涵盖了七个媒介平台

形态,分别是论坛、博客、网络新闻、微博、纸媒、微信、APP新闻,基本涵盖了数据可以搜索获取的主要媒介样态。监测系统对监测内容进行不同科普主题的划分,给每个主题有针对性地进行设置科普领域监测关键词,并及时进行研判更新。后台通过这些监测关键词可以在各大监测平台实时获取数据,并进行数据存储与分析,后期可以采用人工参与的方式对系统获取到的数据进行深度解读。例如,根据阅读量与回复量等内容统计出重点及热点科普信息;对网友评论和媒体态度进行提炼概括;思考科普启示等。最后根据这些内容形成文字版周报、月报和年报。

1.2 网络科普舆情数据分析方法

网络科普舆情系统的建立为科普舆情数据的实时抓取和收集提供了重要平台,通过对这些在线数据进行整合梳理分析,找出其中的趋势和规律,可以了解网络科普舆情的状态。网络科普舆情研究周报、月报等以舆情系统收集的数据作为起点,结合人工方式对不同时间段的网络科普舆情情况进行解读,产生了具体直观的阶段性成果,对其研究数据进行整合分析可以从更加整体的视角诠释网络科普舆情规律。而通过系统在线数据曲线图的直接生成,则可以看到不同时间周期的舆情趋势变化,对于了解网络科普舆情来说效果更加直接生动。

出于以上考虑,本文数据分析主要从两个方面来开展:第一个方面是将数据监测系统2015年10月至2015年12月共3个月的12份周报和3份月报进行数据提取,结合统计学的系统抽样方法,对被抽取出来的周报和月报样本数据进行分析,得出数据规律及相关结论;第二个方面是对舆情监测系统里面的相关数据进行在线二次分析,寻找舆情趋势的规律。比如选定特定日期,对数据监测系统里面的趋势图进行生成和分析,发现和总结其中的规律。

根据系统抽样原则,本文对12份周报样本进行了间隔取样,共获得6份样本;对3份月报样本进行了全样本分析。在现有12份周报样本中,2015年12月21日—2016年1月

3 日这期周报因假期及项目进度原因发布的是双周报，其余的 11 份都是周报，现将这 12 份周报进行数字编码，设置为 1~12 号，在这 12 份样本中进行间隔抽取后确定 6 份样本，即序

表 1 不同时间段的周报样本统计

样本 1	样本 2	样本 3	样本 4
20151005—20151011	20151012—20151018	20151019—20151025	20151026—20151101
样本 5	样本 6	样本 7	样本 8
20151102—20151108	20151109—20151115	20151116—20151122	20151123—20151129
样本 9	样本 10	样本 11	样本 12
20151130—20151206	20151207—20151213	20151214—20151220	20151221—20160103

除了对周报进行抽样分析，还将对 3 份月度样本进行全样本分析。

2 从数据看网络科普舆情特征与规律

正式分析前，将研究中会涉及的词汇——“热度”释义如下：舆情分析中所提及的“热度”和通俗意义上表达温度的“热度”含义不同，舆情分析中的“热度”通常指新闻或者其他信息的热门程度，这种热门程度通常体现为文章发布量、用户阅读量或者网民评论回复数等，一般通过数字或者分析百分比等指标来体现。从一定程度上来说，通过热度指标可以看出研究对象（这里指科普相关信息）被用户关注或者热议的程度。

本文研究基于两个研究入口开展：第一种是对舆情数据监测系统的研究周报和月报进行分析，针对以下三个方面的内容得出相关结论，即：不同载体舆情热度对比、同类载体舆情热度对比、科普专题舆情热度对比；第二种是对舆情数据监测系统在线数据分析，得出科普舆情趋势曲线图规律，即：舆情热度趋势图分析（日、周、月）。

（1）不同载体舆情热度对比显示：网络新闻、论坛、微信三个平台新闻量大。

载体这里指媒介形态（下同），具体指网络新闻、博客、论坛、APP 新闻、微信、纸媒、微博这七类数据源。舆情热度指不同媒介平台抓取的新闻量，新闻量越大，说明该媒介形态的科普信息越多，作为科普平台的功能性越强，网友也越关注。通过对不同载体信息量的分析，可以知道当前哪类载体的

号为 1、3、5、7、9、11 的这 6 份周报。通过对这六期周报的数据进行录入分析，总结出数据规律（为便于文字表现，接下来仅以样本序号代替样本的时间周期）。

科普信息量较多。

研究对 6 份样本中不同载体的新闻量进行统计，以不同形态载体在七类载体中新闻量占比情况作为对比参数，对不同载体信息量情况进行统计分析，从而了解科普舆情信息的主要阵地。

表 2 不同载体舆情热度对比：新闻量占比（%）

载体	样本序号					
	1	3	5	7	9	11
网络新闻	38	41	43	41	42	37
博客	7	6	7	9	10	19
论坛	17	17	17	16	17	19
APP 新闻	3	3	3	4	4	4
微信	17	15	15	15	10	10
纸媒	4	5	5	5	5	4
微博	14	13	10	9	10	7

从这组数据中可以看出，不同类别平台的新闻获取量在总量占比中呈现比较平稳的态势。综合来看，网络新闻、论坛、微信这三个平台分别排名前三位；APP 新闻平台和纸媒平台列最后两位；博客和微博的新闻获取量则保持在一个大致相当的浮动范围，但微博比博客的信息量略占优势。如表 2 所示。

（2）同类载体舆情热度对比显示：搜狐新闻、新浪微博、百度贴吧等在全类载体中排名领先。

为了了解同类载体中表现比较突出（排名前三位）的佼佼者，从而确定舆情聚集的重点媒介，因此做了这一对比排名。排名主要分析同类载体舆情热度排名前三位的媒介，这一排名可以为科普传播工作提供目标聚焦点，在信息投放时可以确定舆情关注的优先选项，做到有的放矢，不至于在众多同类型媒

介中眼花缭乱。

通过综合数据可以看出，网络科普舆情数据来源最多的新闻类网站平台是搜狐新闻，博

客平台是新浪博客，论坛平台是百度贴吧，新闻 APP 是 Zaker，纸媒是南方日报，微博是新浪微博。表 3 是详细数据表。

表 3 同类载体舆情热度对比：排名前三位

载体	样本序号					
	1	3	5	7	9	11
网络新闻	搜狐新闻	搜狐新闻	搜狐新闻	搜狐新闻	搜狐新闻	搜狐新闻
	网易新闻	潍坊新闻网	光明网	光明网	中国网	中国网
	光明网	光明网	中国网	中国网	光明网	光明网
博客	新浪博客	新浪博客	新浪博客	新浪博客	新浪博客	新浪博客
	天涯博客	天涯博客	天涯博客	天涯博客	天涯博客	天涯博客
	和讯博客	和讯博客	和讯博客	和讯博客	和讯博客	和讯博客
论坛	百度贴吧	百度贴吧	百度贴吧	百度贴吧	百度贴吧	百度贴吧
	网易论坛	万户论坛	万户论坛	网易论坛	温商社区	青岛社区
	天涯	网易论坛	网易论坛	东方财富网	青岛社区	网易论坛
APP 新闻	Zaker 新闻	Zaker 新闻	Zaker 新闻	Zaker 新闻	百度新闻	百度新闻
	搜狐新闻	人民新闻	中国经济网	百度新闻	Zaker 新闻	Zaker 新闻
	人民新闻	搜狐新闻	搜狐新闻	人民日报	搜狐新闻	搜狐新闻
微信	—	—	—	—	—	—
纸媒	南方日报	南方日报	南方日报	南方日报	山西商报	山西商报
	金华晚报	中国环境报	中国环境报	山西商报	南方都市报	南方日报
	山西农民报	人民日报	深圳特区报	深圳特区报	南方日报	南方都市报
微博	新浪微博	新浪微博	新浪微博	新浪微博	新浪微博	新浪微博
	腾讯微博	腾讯微博	新浪长微博	腾讯微博	和讯博客	和讯博客
	新浪长微博	新浪长微博	腾讯微博	和讯博客	腾讯微博	腾讯微博

论坛：从该数据中可以看到，百度贴吧在 6 期抽样中稳定排在第一位，可以看出百度贴吧在用户中受到极大欢迎；网易论坛则在这 6 期抽样中，5 次排名前三位，综合排名仅次于百度贴吧，位列第二位。其他如万户论坛和青岛社区也在前三名的序列中时有出现，可以看出这两个平台在论坛用户中也有一定的吸引力。

博客：在该载体类别的 6 期抽样中，新浪博客、天涯博客、和讯博客这三个平台以稳定的排名分别位列第一、二、三位。

网络新闻：在网络新闻平台中，搜狐新闻以绝对优势在 6 期样本中连续排名第一位，可以看出它是最主要的信息获取平台；光明网则在 6 期抽样中连续 6 期排名前三位，可以看出它是排名第二位的网络新闻来源平台；中国网则在 6 期抽样中有 4 期排名前三位，在本平台中排名第三位。

纸媒：在纸媒类载体中，南方日报在 6 期样本的前三位中均有出现，并位列 4 期样本的第一位，可以看出在纸媒中具有很强的辐射

力，在该类载体中排名第一位；山西商报在 6 期样本中有 3 次排名在前三位，综合排名第二，有些出人意料；其他如中国环境报、南方都市报、深圳特区报分别有 2 次出现在本次前三名的排名中，可以看出这几个媒体也是纸媒信息比较重要的信息获取平台。

新闻 APP：在 6 期抽样中，Zaker 新闻每期都排在前三位，其中四次排名第一，可以看出在新闻 APP 中，Zaker 新闻综合排名第一位；搜狐新闻则在 5 期抽样的前三名中都有出现，综合排名第二；百度新闻则以 3 期入选前三名的成绩位列新闻 APP 的第三名。

微博：新浪微博以绝对优势占据 6 期抽样的第一位，是信息获取排名第一位的微博平台；腾讯微博在 6 期抽样中均排名前三，综合排名位列微博平台的第二位；和讯博客和新浪长微博在 6 期抽样中分别在前三位中出现 3 次，可以看出这两个平台仅次于新浪微博和腾讯微博。

(3) 科普专题舆情热度对比显示：健康医疗类信息始终位列第一位。

为了更有针对性地对科普领域信息进行监测，本文对科普领域信息进行了专题划分，具体分为应急避险、食品安全、生态环境、前沿科技、健康医疗、能源利用、事故、信息通信技术、航空航天、伪科学等多项类别。每类专题里面包括众多关键词，关键词采用迭代更新机制，根据热点焦点科普内容，定期进行新词增加和补充。舆情监测系统通过专题名称及关键词可以实现对科普信息的即时抓取，抓取的信息会自动归类在各自的专题中形成信息总量。

在众多科普专题中，会有一些专题的信息量相对来说较多，另外一些专题的信息量则相对较少。通过科普专题舆情热度进行对比，可以更多地了解信息量较多的专题领域，通常这些领域和公众相关度更大，可以为科普工作提供借鉴视角。

研究分别把 6 份样本中排名前三位的科普专题进行提取排列，形成如表 4 所示的内容。可以看到排名具有一定的规律性，并可以看

表 4 科普专题舆情热度对比：排名前三位

排名	样本					
	1	3	5	7	9	11
1	健康医疗	健康医疗	健康医疗	健康医疗	健康医疗	健康医疗
2	事故	生态环境	生态环境	生态环境	生态环境	生态环境
3	生态环境	事故	事故	信息通信技术	事故	信息通信技术

出信息量较大的科普专题。

通过 6 份样本排名前三位的科普专题排名可以看到，健康医疗类信息始终位列第一位，生态环境则以极大优势排名第二，事故类科普专题排名第三，信息通信技术专题在不同周的报告中也偶有出现。综合来看，排名前三位的专题都是和公众生活密切相关的领域。

(4) 日、月、周舆情热度趋势图呈现出显著规律。

本次舆情热度趋势图主要从舆情监测系统数据平台出发，对每日、每周、每月的舆情趋势图进行分析，得出三个不同时间周期的舆情特点如下。

① 从每日舆情热度走势看，上午 9 点和下午 3 点是高峰时段。

每日舆情热度趋势图选取 6 份样本的每周三数据进行分析。从信息载体 24 小时的舆情热度分布来看，从凌晨 4 点到上午 9 点呈现信息逐渐上升的态势，9 点到达峰顶之后到中午 12 点之间呈现逐渐下降趋势，12 点到达波谷，后又逐渐呈现上升趋势，直到下午 3 点出现新一轮小的波峰，下午 3 点后直到下午 6 点呈现逐渐回落趋势，下午 6 点至晚间呈现明显低线走势。由此可见，人们发布和接触信息的两个高峰时段分别是上午 9 点左右和下午 3 点左右。如图 1 所示。

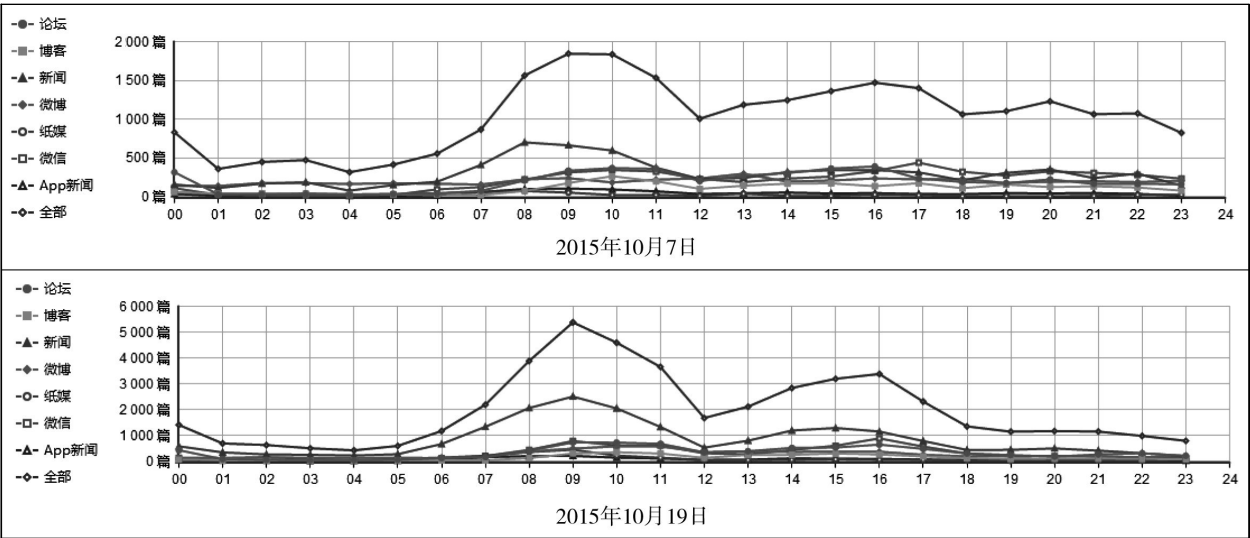
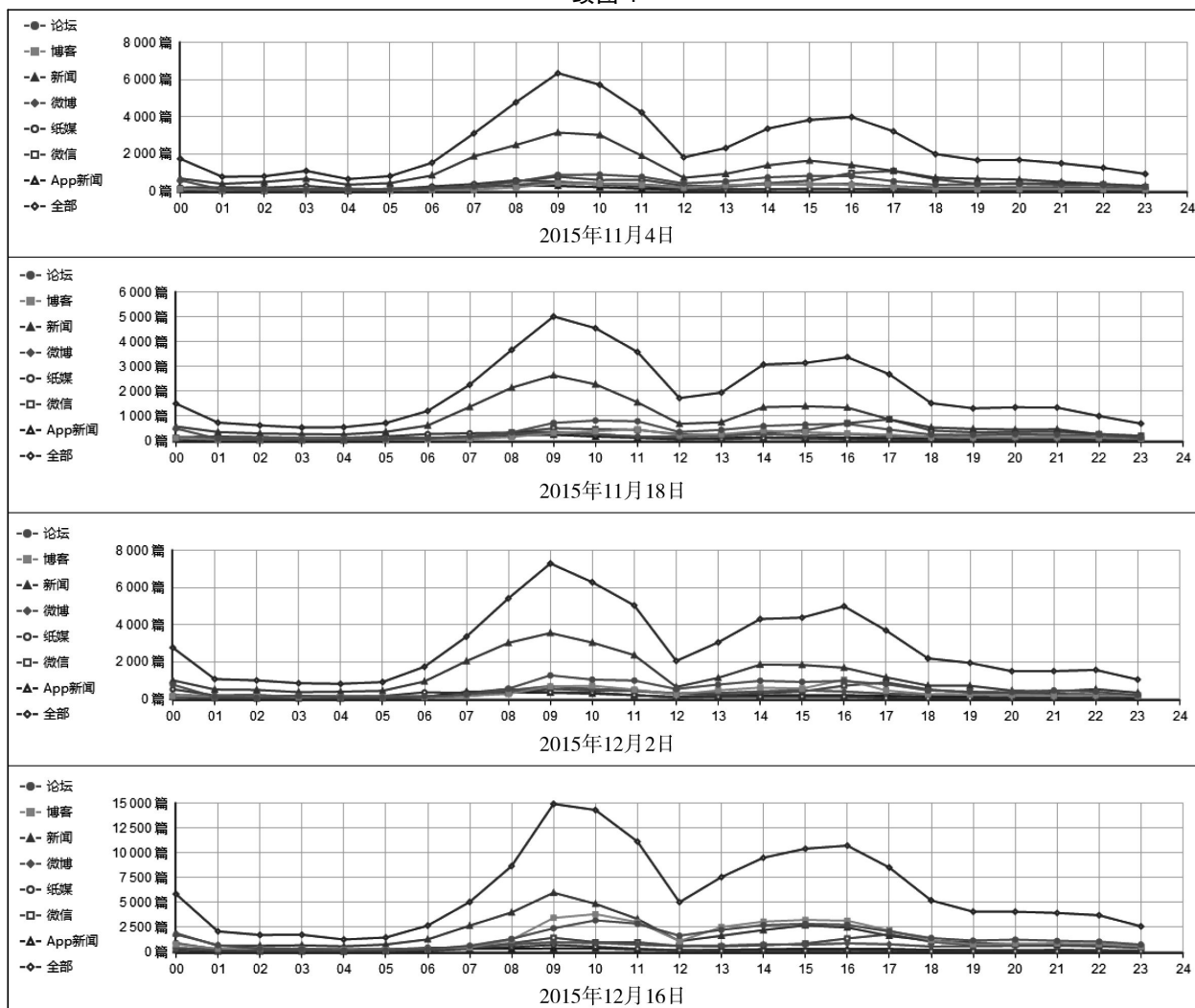


图 1 所选 6 份样本的每周三舆情热度趋势图

续图 1



② 从每周舆情热度走势看,周三和周四通常是高峰,周末处于较为明显的低谷。

图 2 是本次选取的 6 份样本的每周舆情热度

度趋势图。样本 1 的周一至周三都在国庆节假期,信息量相对较少,10 月 8 日上班后信息呈现大幅上升,在之后连续上班的三天保持明显高

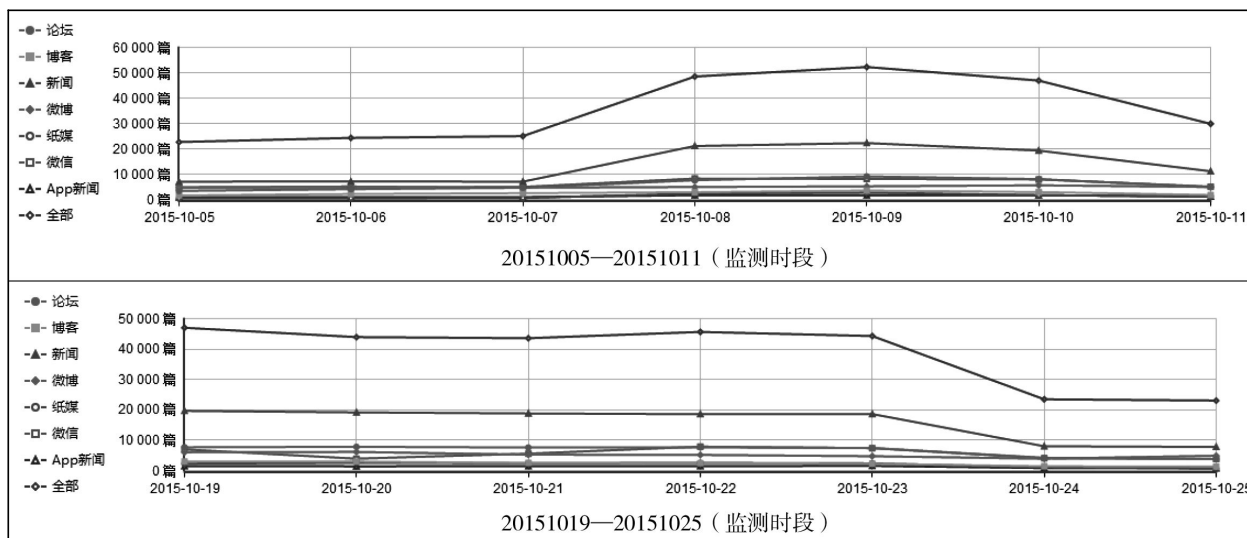
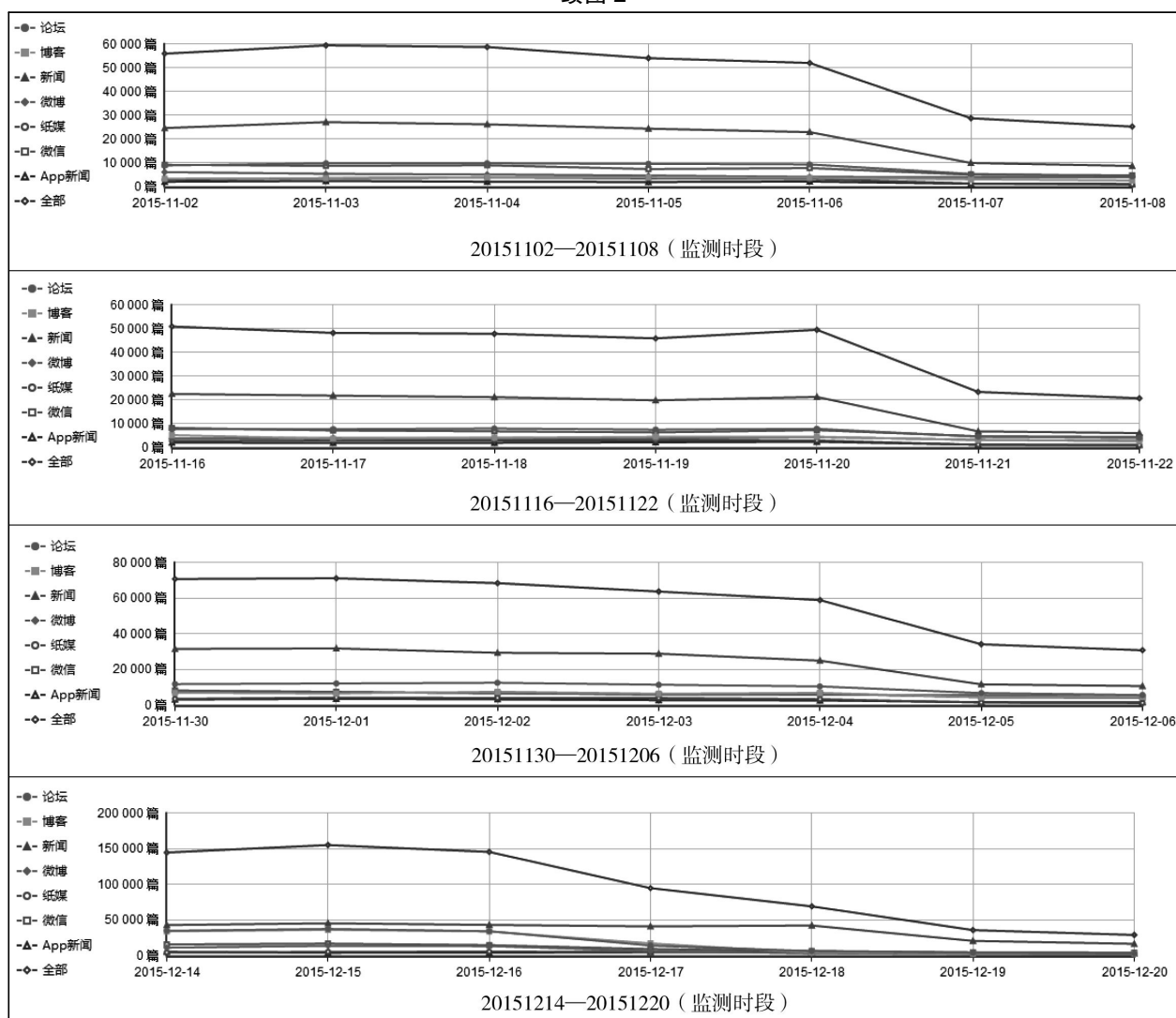


图 2 所选 6 份样本的每周舆情热度趋势图

续图 2



位态势。其他 5 份样本则大体保持了类似的热度趋势：从周一开始，信息量呈现逐渐上升趋势，到周四通常会形成信息总量的一个小高峰，然后呈现逐渐下降趋势，但是周五通常仍然会保持略低于周四的信息量，周六和周日的信息量下降比较明显，周日通常会降到一周的最低点，到了周一，信息量又会呈现比较明显的上升趋势。

③ 从每月舆情热度走势看，四个星期之

间形成较为固定及明显的波峰和波谷。

从图 3 可以看到，每个月四个星期之间通常会形成较为明显的波峰和波谷线，这和每周工作日舆情信息量偏多而周末的信息量偏少有关。从每个月来说，用户在工作日通常生活比较规律，使用互联网比较便捷，对科普信息的阅读、分享、评论会较为活跃；周末时，用户通常选择休息、娱乐或者是看一些“轻新

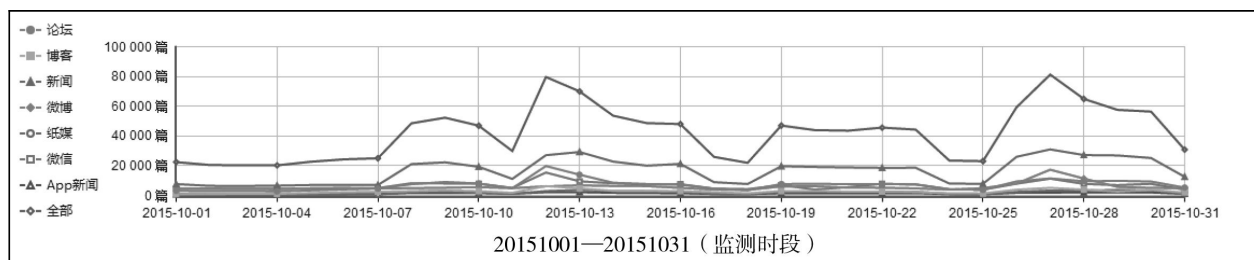
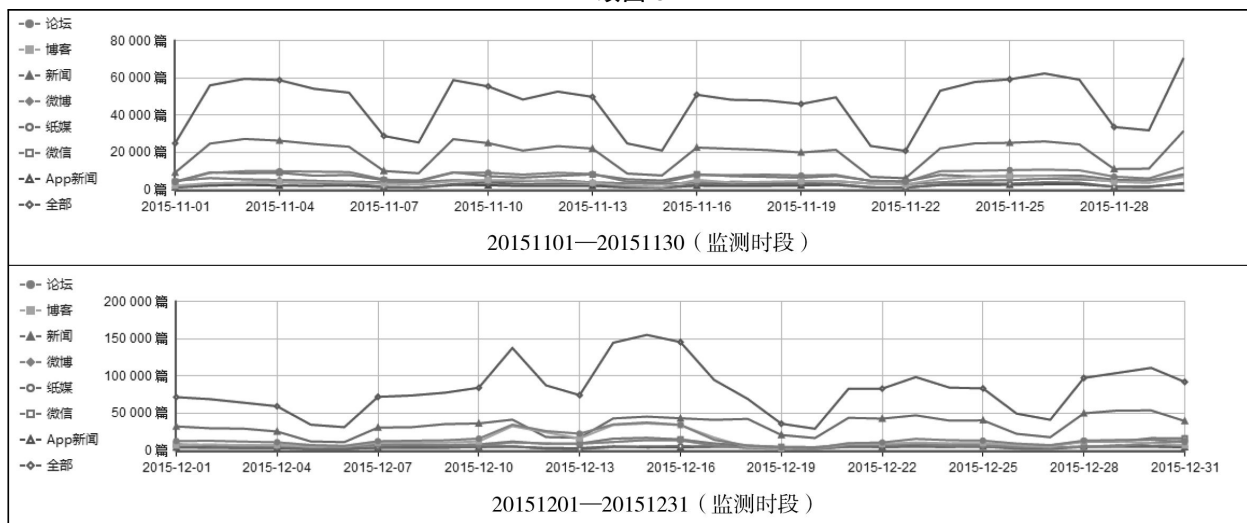


图 3 全部传播载体月度舆情热度走势图

续图 3



闻”，对科普信息的关注度会有所下降。从科普传播工作来说，建议在有重要的活动或是重要新闻需要发布时，选择合适的传播时机，这样才能达到更好的传播效果。

3 研究结论

在本文语境下，网络科普舆情研究主要关注以下几个层次：一是在不同类媒介载体的对比中，公众更喜欢从哪类载体获取科普信息；二是在同类媒介载体中，哪些媒介平台更受公众青睐；三是在不同类别的科普专题中，哪类题材的科普内容更为受众所关注。除此之外，当分别将日、月、周设定为科普信息的传播周期时，公众对科普信息的关注度是否有变化及是否有规律可循。通过数据分析，我们可以将研究结论综合概括如下。

从不同载体舆情热度对比数据中可以看出，网络新闻、论坛、微信这三个媒介平台分别排名前三，APP 新闻平台和纸媒平台列最后两位。也就是说，相对于其他平台来说，前三个渠道的科普信息存量或者受众关注评论数都比较多，APP 新闻平台的热度排名相对靠后与 APP 新闻平台数量偏少或许有关，而纸媒的热度排名靠后则侧面体现了其在现实媒介生态中呈现出的式微态势。同类载体舆情热度对比数据显示，科普舆情数据来源最多的博客平台是新浪博客，论坛平台是百度贴吧，新闻 APP 是 Zaker，纸媒是南方日

报，微博是新浪微博。从一定程度上来说，这个数据结果同公众对于这些载体平台的品牌认知有相当的契合。在科普专题舆情热度对比中，可以看到健康医疗类信息始终位列第一位，生态环境类专题排名第二，事故类科普专题排名第三位。可以看出，排名前三位的科普专题都是与公众生活密切相关的领域。

通过分析舆情热度趋势图可以看出，每日、每周和每月的舆情趋势图都呈现出一定的规律和特点：每日舆情数据在每天上午 9 点左右和下午 3 点左右是两个较为明显的高点；每周舆情数据在工作日和周末之间有较为明显的分界点，工作日尤其是每周三和周四通常是信息量的高峰，周六和周日两天则处于较为明显的低谷状态；每月舆情数据从趋势图上来看，最为明显的特征是每周之间形成比较规律的波峰波谷线。

以上这些网络科普舆情特征及规律可以为科普系统工作者及媒体工作者提供参考，从而挑选合适的时机来进行信息发布及了解舆情。

参考文献

- [1] 王来华. 舆情研究概论·理论、方法和现实热点[M]. 天津: 天津社会科学院出版社, 2003.
- [2] 谢小军. 我国科普工作面临的形势和挑战[J]. 中国科技信息, 2012(22): 52.

(编辑 袁博)