

计算视角下的虚假信息传播:内容、主体与模式

吴晔^a,李永宁^a,张伦^b

(北京师范大学 a.新闻传播学院;b 艺术与传媒学院,北京 100875)

摘要: 虚假信息的传播在政治、经济等多个领域影响着社会正常运作,深入理解虚假信息的传播机制并遏制其传播是传播学、政治学、计算机科学等多个学科领域的共同研究目标。基于社会化媒体对于虚假信息内容及传播踪迹的记录,以及信息传播相关计算算法的广泛采用,传播学在当前阶段得以对虚假信息传播进行经验性深入探讨。基于此,将从虚假信息内容特征、传播者用户特征和虚假信息传播模式 3 个角度系统综述计算方法对于虚假信息传播研究的推进,探讨未来可行的跨学科合作模式。

关键词: 虚假信息;计算方法;内容特征;用户画像;传播模式

中图分类号: G206.3

文献标志码: A

虚假信息的传播研究作为传播学的重要研究领域,近几年引起了包括经济学、心理学、政治学、计算机科学等多个学科的关注^[1]。2018 年 3 月, *Science* 刊登的 *The Spread of True and False News Online*^[2] 收集了 Twitter 中约 12.6 万条消息,囊括了超过 300 万人约 450 万次的信息传播数据,并对比真假信息传播内容、主题和模式的差异。该论文一经发表便引起了包括信息科学、计算机科学以及传播学者在内的广泛关注与讨论。

文献^[1]指出虚假信息使得信息传播的成本提高,从而引致相关领域需要付出巨大的社会代价。1) 虚假信息影响受众获取信息的准确度,从而影响其社会决策行为;2) 在选举活动中,虚假信息影响受众决策的正确性进而削弱整个社会系统民主选择的能力;3) 虚假信息可能导致受众对合法的新闻信息源产生怀疑,影响媒体公信力;4) 虚假信息刺激信息生产方通过生产虚假信息的方式获取关注。

社会化媒体助长了虚假信息的传播。皮尤研究中心发布的 *News Use Across Social Media Platforms 2017*^[3] 指出,67% 的人通过社交媒体获取信息,新媒体技术加速了信息的分享和大规模传播,加之社交媒体信息生产成本大幅降低,使得社交媒体助长和加速了虚假信息的传播^[1]。而另一方面,社会化媒体记录了信息传播的“数字足迹(Digital Trace)”,使得数据的大规模获取和分析都具有了可操作性^[4]。基于社会化媒体,虚假信息的传播得以被经验性研究。面对虚假信息这类海量且非结构化的数据处理,传统的传播学定量研究方法需要消耗大量的人力和时间,而基于文本挖掘和模型建构的计算机算法的加入,能够从技术层面助力传播学深入理解虚假信息特质以及传播过程,最终拓展虚假信息传播的理论视野。

综上所述,在计算时代,虚假信息成为影响社会经济活动的重要议题;多个学科致力于如何控制和遏制虚假信息的传播研究——即探究虚假信息内容特征、传播者特征以及传播模式,从而减少信息在社会层面的传播成本。更重要的是,不同于传播学传统研究范式,计算方法得以使该议题能够被系统经验性研究。基于此,本文将综述如何利用计算方法对于虚假信息传播进行研究。具体而言,本文将从虚假信息内容分析、传播者用户画像以及传播过程 3 个方面来综述当前多个学科对于虚假信息这一议题的研究。

收稿日期: 2018-12-17.

基金项目: 国家自然科学基金重点项目(71731002);中央高校基本科研业务费专项资金资助项目;教育部人文社会科学研究一般项目(18YJC860050).

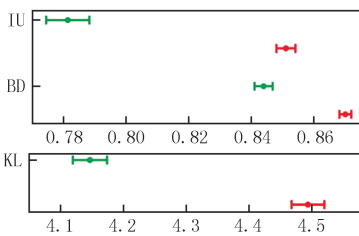
作者简介(通信作者): 吴晔(1982-),男,福建莆田人,北京师范大学教授,博士生导师,研究方向为计算传播学, E-mail: wuyue@bnu.edu.cn.

1 对虚假信息内容特征的判定

对虚假信息内容特征的研究,是及时遏制虚假信息传播、降低虚假信息控制成本的关键.对虚假信息内容特征的判定,是理解虚假信息传播研究的重要基础.判断虚假信息的最理想方式是对比第三方真实数据(Ground Truth).例如,Vosoughi 等人的研究所使用的真假信息数据集,来自于6个独立的事实检测组织共同判定^[2].但是依靠第三方平台检测信息真实性,需要较高的人工成本.

对虚假信息内容进行判定,目前广泛采用的是文本挖掘方法.一般而言,文本挖掘是指为了发现知识,从文本数据中抽取隐含的、未知的、潜在有用的模式的过程,包括文本结构分析、文本摘要、文本分类、文本聚类等.借助这些文本信息的挖掘和分析,能够从海量数据,特别是非结构化数据中发现文本背后潜在的规律和价值.

在虚假信息传播研究中,常使用文本分类和文本聚类技术,分析虚假信息的内容特征以及借助内容特征检测信息真实性.1)文本分类可以理解为将信息数据按照已有的类目或给定的分类标准进行筛选分类,常采用机器学习算法,提高分类的准确性.例如 Mitra 等人通过对 Twitter 消息的可信语料库的研究,确定15个语言维度,然后采用文本分类,将文本用词划分到不同级别的可信度当中,通过某些用词及其相关短语可以预测虚假信息事件^[5].Conroy 等人认为,从语言线索入手检测虚假信息应该是多层次的;即从词汇层次、语法层次,最终到达语篇层次,以达到更好的检测效果^[6].2)文本聚类则是按照文本的相似性对文本自动集聚成群,通过聚类能够得到集群各自的特点,并以此为根据识别潜在的同类信息.例如 Vosoughi 等人利用 LDA 技术(即文档主题生成模型)对信息进行聚类后,以“信息独特性(Information Uniqueness, IU)”,“巴氏距离(Bhattacharyya Distance, BD)”和“K-L 散度(K-L Divergence, KL)”为指标,测量信息的“新奇性(Novelty)”,3个指标反映的都是信息差异性.指标越大,则信息异质性越强,即“新奇性”越强.如图1所示,绿色和红色分别代表真实信息和虚假信息.研究发现虚假信息比真实信息更具有“新奇性”^[2];通过将包含“is this true”“really”等表达怀疑态度的推文围绕特征词进行文本聚类,基于文本特征及时识别出有争议话题,能够实现谣言之早期识别^[7].



| | mean | | variance | | ks-test |
|----|------|-------|----------|---------|-----------------------|
| | true | false | true | false | |
| IU | 0.85 | 0.78 | 0.005 2 | 0.007 2 | $D=0.457, p \sim 0.0$ |
| KL | 4.49 | 4.15 | 0.161 8 | 0.094 8 | $D=0.433, p \sim 0.0$ |
| BD | 0.87 | 0.84 | 0.000 8 | 0.000 8 | $D=0.415, p \sim 0.0$ |

图1 真实信息和虚假信息“新奇性”对比^[2]

Fig.1 Differences of novelties between true (green) and false (red) rumor tweets^[2]

在对虚假信息内容特征的判定方面,计算机科学的数据处理能力带给人文社科传统的内容分析新的概念测量的突破.尽管在一些研究中,计算机算法不能完全取代人工编码过程,但是计算机算法能够简化人工编码前的信息检索和收集过程.例如 Kim 等人结合用户主动标注可疑信息的行为,设计了一种 C_{URB} 算法,利用算法决定何时、将哪一条被标注的可疑信息发送到信息核查平台.基于算法的判定,再介入人工编码,实现高效及时的检测^[8].从虚假信息识别和检测的角度而言,相比于传统的后判方式,利用计算方法减少了识别虚假信息的时间成本,在虚假信息形成大规模扩散之前实现较为准确的识别,遏制其传播.尽管如此,文本挖掘作为一个多学科交叉的领域,涵盖了数据挖掘和语言处理等多方面的技术.目前在对虚假信息内容特征分析方面,不仅识别和检测算法设计有待提高,更需要如语言学等领域专业知识的注入,多学科共同努力,提高分析的精确度.

2 虚假信息传播者特征构建

对参与虚假信息传播的用户进行特征刻画,是研究虚假信息传播过程的重要问题.社交媒体平台的开放

性为研究用户属性和行为特征提供了重要数据^[9]。当前对于传播者特征进行研究的计算方法,主要集中于“用户画像”研究。传播学理论研究的特征阐释流派认为,你在特定的时段如何进行传播信息取决于你的个性和你所处的具体条件^[10]。这一理论为用户画像提供了理论依据,即用户发布、评论或转发虚假信息等行为和用户个人的特征紧密相关。

用户画像即根据用户的人口统计学信息、历史行为信息、社会关系信息等直接和间接数据,对用户特征进行量化和归纳。具体而言,用户画像可以从属性特征和行为特征对虚假信息传播者进行特征建构。

1)从属性特征角度来看,用户画像可以量化传播真假信息用户的属性差异。一方面可以研究用户社交网络结构特征(例如关注数、粉丝数)对虚假信息传播的影响,比如 Zubiaga 等人研究发现,具有高关注率(关注率=关注数/粉丝数)的用户倾向于对流言表达肯定态度,而且倾向于引用外部证据来支持自己的观点^[11]。Vosoughi 等人通过对传播真、假信息的用户进行比对发现传播虚假信息的账号,在“关注数”和“粉丝数”等指标中,都高于传播真实消息的人^[2]。Jang 等人研究发现,大多数虚假信息都是由普通用户发布,而不是记者、政客等特别身份的人^[12]。另一方面,社交机器人(Social Bots)已被广泛用于渗透政治话语、操纵股市、窃取个人信息、传播错误信息^[13]。Twitter 公布其平台上有 9%到 15%的活跃账号是社交机器人;Facebook 推算其平台上有近 6 000 万的社交机器人^[14]。利用属性特征对机器人账号进行用户画像可以用于对社交机器人的检测,比如建立在机器学习算法之上的 BotOrNot 系统^[15],就是利用了账号包括多种属性特征(如用户元数据、好友统计、词性和情绪分析等)在内的超过 1000 个特性,评估 Twitter 账户与社交机器人的已知特性之间的相似性,为 Twitter 账号计算“机器人分数”。Shao 等人利用这一系统发现,在该研究的样本中只有 8%的账号被判定为社交机器人,但是这 8%的账号传播的未被证明真假的信息数量占到总体的 36%^[16]。

2)从行为特征角度来看,用户画像方法旨在发现某类特定行为,对于虚假信息传播者的指征意义,例如转发行为、点赞行为、提及行为、内容重复发布行为等。从转发、点赞等行为特征入手,能够发现虚假信息传播者在群集维度上的特点。例如,Bessi 等人研究发现,关注“阴谋论”类型信息的极化用户(即对某类信息点赞数量占到总点赞量的 95%的用户)更倾向于评论“阴谋论”类型的帖子^[17],如图 2 所示,“科学类”信息(Scientific news)极化用户的评论行为中,9.71%是评论给“阴谋论”信息(Alternative news),而“阴谋论”信息极化用户的所有评论行为中,只有 0.92%是评论给“科学类”信息。根据社会判断理论,当信息与个人的立场较为接近的时候,它就会被吸收;如果信息与个人的立场大相径庭,反而会加强个人原有的立场,这就是所谓的“回力飞镖效应”^[18]。围绕信息类型的不同,可以将用户划分成不同的社群群体,甚至是划分出极化群体。这一类研究从量化的角度对经典理论进行了新的解读和验证。从提及行为、内容重复发布等行为特征入手,可以作为检测社交机器人的指标。社交机器人通常在虚假信息发布的初期进行大量转发,而且试图通过评论和提及的方式影响真实用户^[16]。在 BotOrNot 系统用于计算机器人分数的特性中,也包含了如活动的时态模式等行为属性。

通过用户画像的方式,能够描述出参与虚假信息传播者的账号属性和行为特点,归纳虚假信息传播者和易感人群的群体特征。计算时代的用户画像,能够利用大样本、甚至全样本数据,提供更具普适性的研究结果。这种“非介入性(Unobtrusive)”方法适合于对“少数”和“敏感”人群进行区分和鉴别。这是传统的基于调查法进行的用户特征研究所无法实现的研究目的。其根本原因在于,调查研究无法询问被调查者是否及如何传播虚假新闻等负面行为。更重要的是,计算方法在研究信息传授者特征时可以超出传统的人口统计学特征范围,加入社交属性、用户历史记录等,更准确地实现对用户行为数据和社交关系的画像。此外,对于广泛存在的“社交机器人”,只有通过算法才能进行有效鉴别,研究社交机器人对信息传播的效果评估,并实现对社交机器人的及时管控。

3 虚假信息的传播模式

通过“数字足迹”研究信息传播模式,能够帮助理解虚假信息的扩散过程。社交网络中的传播行为包括点击、评论、喜欢、转发等多种形式。利用这些行为记录可以描述信息传播模式。具体而言,通过描述一条信息完整的传播过程,得到按照时间次序的每一个转发者的信息,以及转发者相互之间的信息转发情况,从而构建

出单条信息扩散结构^[19].比如通过分析每一个转发的信息来源,可以得到一个完整的信息转发网络.

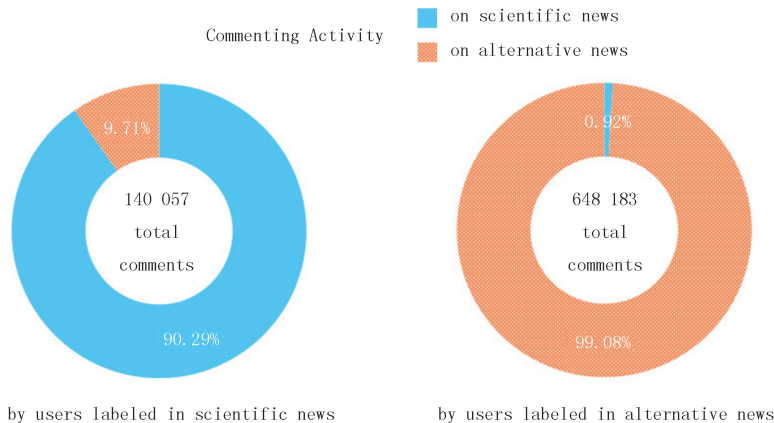


图2 极化用户的评论行为^[17]

Fig.2 The commenting activity in terms of polarized users^[17]

从计算科学的角度来看,信息传播模式可以从多个维度进行测量,比如信息传播的范围、广度、深度,以及信息扩散的时间分布特征和空间分布特征等.Goel等人提出了信息传播过程中“结构性扩散度”的概念:1)在深度固定的情况下,结构扩散度随着分支数的增加而增加;2)在分支数固定的情况下,结构扩散度随着深度的增加而增加;3)和病毒式传播不同,广播式传播覆盖的节点数不影响其结构扩散度^[20].结构扩散度能够测量信息传播的结构特征;当该值取1时,即指示信息传播呈广播式传播(例如传统媒体一个信源向其多个受众直接传播信息)、当该值远大于1时,即指示信息呈病毒式传播(即多代分支结构传播模式,例如个体之间一对一、一对多相互传播信息).当结构扩散度从2增加到50时,传播的扩散结构如图3所示.图3每个结构示意图下方,展示了当前结构扩散度下,该结构的级联随着时间变化的曲线.

研究在线社交网络中虚假信息如何传播,一方面可以通过实证数据分析虚假信息和真实信息在传播模式方面的差异;另一方面可以对虚假信息的传播模式进行建模预测.首先,从实际数据入手量化真假信息传播模式差异,可以帮助理解虚假信息的传播效果.例如图4所示,绿色和红色分别代表了真实信息和虚假信息的传播模式,Vosoughi等人从传播效果入手,发现虚假信息在传播深度(Cascade Depth,即信息被转发层数)、传播范围(Cascade Size,即信息覆盖的总用户数)、传播最大宽度(Cascade Max-Breadth,即一次转发后信息到达的最大用户数)、以及结构性扩散度(Structural virality,即信息平均传播路径)四个维度皆领先于真实信息.虚假信息比真实信息得到了更多的转发,而且在传播模式上不仅仅是广播式的传播,还通过广泛的人际传播形成了病毒式的传播结构^[2].Vicario等人研究了“科学类”信息和“阴谋论”信息传播中的级联(Cascade)变化,发现2种信息的传播模式存在差异,“科学类”信息在级联达到200左右时信息的生命周期(从第一次信息发出至最后一次信息转发的时间间隔)最长,而“阴谋论”信息的生命周期与级联数呈现出正相关关系^[21].Jang等人通过研究真实和虚假信息的演变过程发现,2种信息的内容演变过程存在差异,在一次次的转发中,虚假信息文本发生的变化更多^[12].

其次,对虚假信息传播模式进行建模,既可以预判虚假信息传播效果,又可以预判虚假信息阻碍策略的有效性.对于虚假信息传播效果的研究,可以引入如传播时间(即虚假信息要到达其他节点所需的最小时间)、传播因子(即某一给定节点与其他节点交换虚假信息的平均最大可能性)等变量^[22],或者加入虚假信息传播时空数据(即同一话题2次转发之间的时间间隔和空间距离)^[23],构建虚假信息传播模式,预测信息扩散的速度、范围等.虚假信息传播阻碍策略研究的目的在于借助传播模式,推断如何能够以最小成本实现最有效的抑制虚假信息传播.He等人研究表明实时优化(real-time optimization, RTO,即在一定时间内,以最小成本,结合多种方式抑制虚假信息传播)和“间歇性传播纠正”(pulse spreading truth and continuous blocking rumor, PSCB)都是有效的虚假信息抑制策略^[24].通过建立信息在社会网络中的传播模式模型,预测虚假信息传播态势,及时采取抑制策略,降低虚假信息传播的负面影响.

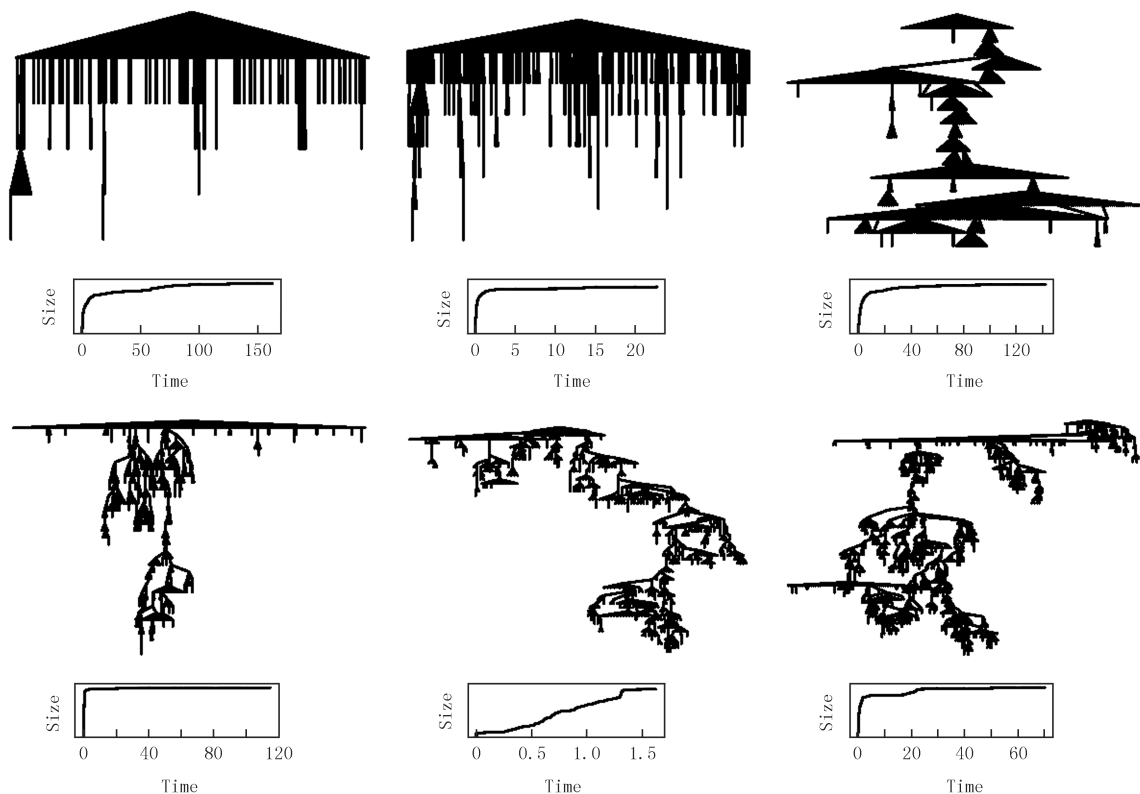


图 3 结构扩散度从 2 增加到 50 示意图^[20]

Fig.3 A Random Sample of Cascades Stratified and Ordered by Increasing Structural Virality, Ranging from 2 to 50^[20]

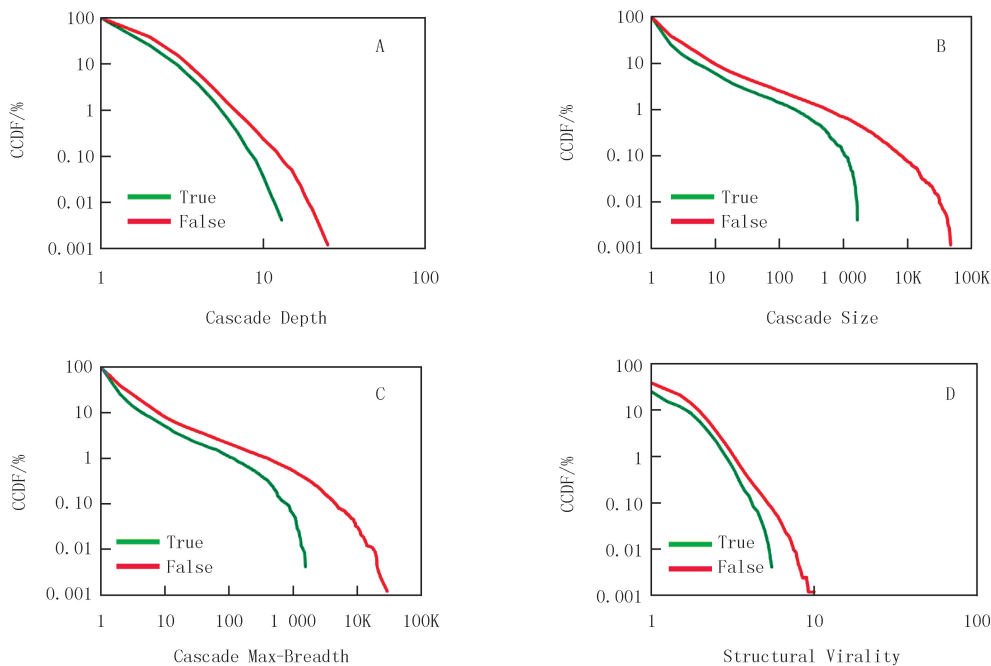


图 4 真实信息和虚假信息传播模式图^[2]

Fig.4 Complementary cumulative distribution functions (CCDFs) of true and false rumor cascades^[2]

信息传播模型是传播学的经典研究问题之一,同时也是复杂网络的重要研究问题.目前,相关学科已提

出多种信息传播模型.但这些模型主要探讨的是信息的一般性传播结构模式,且绝大部分停留在理论推演阶段.而目前对于虚假信息的传播模式,特别是导致其传播模式特征的微观动力的研究阙如.在目前数据获取便捷的环境下,未来研究可以利用实证数据逐一检验研究经典的信息传播模式对描述、解释和预测虚假信息传播的有效性.

4 结 语

通过上述文献梳理可见,计算方法的技术支持能够实现对虚假信息传播较为精准的现象描述,即,虚假信息是什么;虚假信息的特征是什么;其和真实信息有何差异;虚假信息传播的参与者具备何种特征;初步描述虚假信息传播模式特征,具备何种网络结构特征等.相比而言,传统传播学领域对于虚假信息传播忽略了2个重要层面:1)真假信息的传播有何差异;2)哪些因素能解释这些差异^[2].这2个问题的解决,一方面依赖于相关理论探究动因,另一方面更需要数据和技术提供帮助.计算方法的加入所带来的技术支持和研究成果,使得传播学学者对这一领域的把握更加清晰准确,从而能够在经验证据的基础上进一步理解虚假信息传播.

此外,对于虚假信息传播的研究,绝非仅凭方法的改进就能够得到根本的进益.深入理解虚假信息的传播,还需要基于对信息传播个体行为的深入理解.信息传播作为传播学研究的核心内容之一,在传播学领域具有深厚的理论积淀,比如传统的信息生产理论、信息接收和处理理论等.但是目前学科之间的割裂,致使对虚假信息传播的理论解释流于表面.例如,传播学详尽可能性模型(ELM)等多个理论探讨了个体信息接收和传播的微观作用机制.但是,在目前的虚假信息研究中,除了提出信息“新奇性”是引起更多转发行为的影响因素之外,计算科学相关领域几乎没有对信息传播的动力进行微观层面的深入解释.因此,探究虚假信息的微观传播机制,从传播学理论入手,借助计算方法实现,连通经典理论和实证数据,是未来对于虚假信息传播研究进行跨学科合作的研究进路.

参 考 文 献

- [1] ALLCOTT H, GENTZKOW M. Social Media and Fake News in the 2016 Election [J]. *Journal of Economic Perspectives*, 2017, 31(2): 211-235.
- [2] VOSOUGHI S, ROY D, ARAL S. The spread of true and false news online [J]. *Science*, 2018, 359(6380): 1146-1151.
- [3] Pew Research Center's Journalism & Media [EB/OL]. [2017-09-07]. <http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/>.
- [4] WATTS D J. A twenty-first century science [J]. *Nature*, 2007, 445(7127): 489.
- [5] MITRA T, WRIGHT G P, GILBERT E. A Parsimonious Language Model of Social Media Credibility Across Disparate Events [C]// *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. New York: ACM, 2017: 126-145.
- [6] CONROY N J, RUBIN V L, CHEN Y. Automatic Deception Detection: Methods for Finding Fake News [C]// *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*. Silver Springs: American Society for Information Science, 2015: 82.
- [7] ZHAO Z, RESNICK P, MEI Q. Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts [C]// *Proceeding of 24th International Conference on World Wide Web*. Geneva: International World Wide Web Conferences Steering Committee, 2015: 1395-1405.
- [8] KIM J, TABIBIAN B, OH A, et al. Leveraging the Crowd to Detect and Reduce the Spread of Fake News and Misinformation [C]// *the 11th ACM International Conference on Web Search and Data Mining*. New York: ACM, 2018: 324-332.
- [9] ZUBIAGA A, AKER A, BONTCHEVA K, et al. Detection and Resolution of Rumours in Social Media: A Survey [J]. *ACM Computing Surveys*, 2018, 51(2): 32.
- [10] Miller L C, Cody M J, Mclaughlin M L. Situations and goals as fundamental constructs in interpersonal communication research [M]// *Handbook of interpersonal communication*. 1994: 162-197.
- [11] ZUBIAGA A, LIAKATA M, PROCTER R, et al. Analysing How People Orient to and Spread Rumours in Social Media by Looking at Conversational Threads [J]. *PLoS ONE*, 2016, 11(3): e0150989.
- [12] JANG S M, GENG T, LI J Y Q, et al. A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis [J]. *Computers in Human Behavior*, 2018, 84: 103-113.
- [13] FERRARA E, VAROL O, DAVIS C, et al. The Rise of Social Bots [J]. *COMMUNICATIONS OF THE ACM*, 2016, 59(7): 96-104.

- [14] LAZER D M J, BAUM M A, BENKLER Y, et al. The science of fake news[J]. *Science*, 2018, 359(6380): 1094-1096.
- [15] DAVIS C A, VAROL O, FERRARA E, et al. BotOrNot: A System to Evaluate Social Bots[C]// Proceedings of the 25th International Conference Companion on World Wide Web. Geneva: International World Wide Web Conferences Steering Committee, 2016: 273-274.
- [16] SHAO C, CIAMPAGLIA G L, VAROL O, et al. The spread of low-credibility content by social bots[EB/OL]. [2018-12-02]. <https://arxiv.org/abs/1707.07592>.
- [17] BESSI A, COLETTI M, DAVIDESCU G A, et al. Science vs Conspiracy: Collective Narratives in the Age of Misinformation[J]. *PLoS ONE*, 2015, 10(2): e0118093.
- [18] 斯蒂芬·李特约翰. 人类传播理论 [M]. 史安斌, 译. 7 版. 北京: 清华大学出版社, 2004: 154-156.
- [19] 许小可, 胡海波, 张伦, 等. 社交网络上的计算传播学[M]. 北京: 高等教育出版社, 2015: 8-9.
- [20] GOEL S, ANDERSON A, HOFMAN J, et al. The Structural Virality of Online Diffusion[J]. *Management Science*, 2016, 62(1): 180-196.
- [21] VICARIO M D, BESSI A, ZOLLO F, et al. The spreading of misinformation online[J]. *PNAS*, 2016, 113(3): 554-559.
- [22] LIND P G, SILVA L R D, ANDRADE J S, et al. Spreading gossip in social networks[J]. *PHYSICAL REVIEW E*, 2007, 76(3): 036117.
- [23] DOMENICO M D, LIMA A, MOUGEL P, et al. The Anatomy of a Scientific Rumor[J]. *SCIENTIFIC REPORTS*, 2013, 3: 2890.
- [24] HE Z, CAI Z, YU J, et al. Cost-Efficient Strategies for Restraining Rumor Spreading in Mobile Social Networks[J]. *IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY*, 2017, 66(3): 2789-2800.

False information dissemination from the perspective of computing: content, subject and pattern

Wu Ye^a, Li Yongning^a, Zhang Lun^b

(a. School of Journalism and Communication; b. School of Arts & Communication, Beijing Normal University, Beijing 100875, China)

Abstract: The spread of false information affects the normal operation of society in many fields such as politics and economy. It is a common research goal of communication, political, computer science and other disciplines to deeply understand the communication mechanism of false information and curb its spread. Based on the record of false information content and digital trace by social media, as well as the wide adoption of computing algorithms related to information spreading, communication is able to conduct an empirical and in-depth discussion of false information spreading at the current stage. Based on this, this study will systematically review the advancement of computing methods in the study of false information spreading from the perspectives of false information content characteristics, communicator characteristics and dissemination patterns of false information, and explore the feasible interdisciplinary cooperation mode in the future.

Keywords: false information; calculation method; content features; user portrait; dissemination-pattern

[责任编辑 杨浦]