



跨媒介舆情网络环境下信息传播机制研究与进展

饶元*, 吴连伟, 张君毅

西安交通大学软件学院社会智能与复杂数据处理实验室, 西安 710049

* 通信作者. E-mail: yuanrao@163.com

收稿日期: 2017-03-06; 接受日期: 2017-06-07; 网络出版日期: 2017-11-30

国家国际科技合作专项 (批准号: 2015DFA81780)、陕西省科技厅协同创新项目 (批准号: 2015XT-21)、教育部“云数融合”基金项目 (批准号: 2017B00030) 和中央高校基本科研业务费 (批准号: zdyf2017006) 资助项目

摘要 随着移动网络、云计算以及大数据技术的应用发展, 网络舆情的传播呈现出了传播跨媒介、内容跨媒体以及覆盖区域跨语言等新的特征, 并成为了新一代舆情事件分析与研究的前沿. 本文从跨媒介舆情网络环境下的信息传播机制出发, 通过对网络信息传播过程中的信息传播者 P 、信息接收者 R 、信息内容 C 、传播媒介 M 以及传播的结果评价 A 5 个核心要素进行建模, 形成了 PRCMA 多元相关的信息传播模型. 在此基础上, 结合上述 5 个要素对舆情信息传播过程中的影响作用和存在的关键性问题, 即: 传播者影响力识别与度量以及对信息传播影响作用问题、不可信信息的识别及可信度的实时度量问题、跨媒介信息传播动力学机制、信息接收者个体兴趣偏好与群体行为对信息接收影响程度的度量问题, 以及跨媒介信息传播效果的评价等问题与技术挑战进行了理论探讨、实际应用以及发展趋势分析, 为解决跨媒介舆情信息的传播机制提供了新的研究线索与方向.

关键词 跨媒介, 舆情网络, 信息传播, 大数据, 评估机制

1 概述

舆情的生产与传播方式对国家安全、社会和谐稳定以及经济发展均可能产生影响与作用. 随着移动网络、云计算以及大数据技术的快速发展, 舆情事件通过网络传播过程中所呈现出来的新特征与新机制特别引人关注和思考, 例如, 在 2016 年美国总统大选中, 有人利用 Facebook 和 Twitter 等社交媒体以及电视与新闻媒介散布的一些虚假、偏激或者恶意的信息来左右选民的政治倾向与态度, 并通过跨媒介信息的快速交叉传播影响甚至改变了选情结果^[1,2]. 李彦宏指出: “在社交媒体上, 那些被传播极其广泛的, 很多是阴谋论、假新闻和各种各样极端感情的抒发”¹⁾. 与此同时, 在我国发生的舆情事件也具有相似的鲜明特征, 例如, 发生在 2016 年 2 月 6 日 19 时 28 分的“上海女逃离江西农村”事

1) 李彦宏. 第三届世界互联网大会演讲. 2016.11.16. <http://www.52dwx.com/article/97285.html>.

引用格式: 饶元, 吴连伟, 张君毅. 跨媒介舆情网络环境下信息传播机制研究与进展. 中国科学: 信息科学, 2017, 47: 1623–1645, doi: 10.1360/N112017-00053
Rao Y, Wu L W, Zhang J Y. A survey of information propaganda mechanism under the cross-medium (in Chinese). Sci Sin Inform, 2017, 47: 1623–1645, doi: 10.1360/N112017-00053

件²⁾。该事件引爆了关于城乡差异、地域歧视、婚嫁观念、贫富差距等热门的民生话题, 激发了网民的集体参与感, 随后被 @ 华西都市报、@ 新浪江西、腾讯、凤凰等微博以及新闻媒体报道, 还上了北美最流行的网站文学城新闻。其中, 以 @ 成都商报发起的“见到第一顿饭后想分手”为核心话题展开的网络讨论, 阅读量已超过 1.1 亿人次, 话题讨论量近 10 万条。随后, 各大媒体纷纷跟进, 如《人民日报》发表评论文章《农村, 说声爱你太沉重》, 光明网发表评论《私域的爱情, 公域的乡愁》, 但通过后续的调查发现该帖子是一个虚假新闻。上述网络热点事件一方面在信息内容上往往存在着较大的争议或者是不确定性^[3]; 另一方面通过不同类型的媒介相互影响并交互传播舆情信息, 体现出当前网络舆情事件所具有的传播跨媒介、内容跨媒体以及跨语言的新特征。如何在这些新的特征条件下进一步探索舆情信息在跨媒介范围内的交叉传播规律、相互作用和动力学机制, 并针对舆情事件内容可信度判断的延迟性与响应滞后性提供快速的分析判断能力, 为舆情信息传播的动态演化过程和趋势预测提供必要的支撑手段, 已成为了当前学术界、工业界以及政府部门共同关注的研究焦点。本文主要针对跨媒介条件下舆情信息传播过程中相关的技术问题挑战以及研究进展进行分析与综述。

2 舆情信息传播核心要素模型与挑战

2.1 舆情信息传播核心要素模型

信息传播是人类社会的一种基本功能, 它主要针对信息的传播者将相应的信息内容通过传播媒介到达信息接收者的一个动态过程, 在此基础上通过传播效果的评价来实现信息传播质量的控制以及传播效果的增强。因此, 可针对传播过程中存在的核心要素进行建模, 并形成了一个多元相关的信息传播 RPCMA 模型, 该模型可以形式化地表示为一个五元组

$$IP = \{P, R, C, M, A\}, \quad (1)$$

其中, IP (information propaganda) 表示信息的传播能力; P 表示信息传播者构成的集合, $P = p_1, \dots, p_i, \dots, p_n$; R 表示信息接收者构成的集合, $R = r_1, \dots, r_j, \dots, r_m$; C 表示信息内容构成的集合, $C = c_1, \dots, c_k, \dots, c_v$; M 表示传播媒介构成的集合, $M = m_1, \dots, m_h, \dots, m_p$, 并且信息在不同的传播媒介之间也存在着相互的传播与影响; A 表示信息传播效果的评价, 它由传播效果评估指标集合 $a_1, \dots, a_i, \dots, a_n$ 和相应权值 $w_1, \dots, w_i, \dots, w_n$ 通过运算而得。通过传播效果评价, 一方面可以对传播过程的质量进行评估分析; 另一方面可以对传播者提供信息反馈, 从而实现传播的控制与优化。上述 5 个关键特征及其关系如图 1 所示。

本文进一步从 PRCMA 模型中存在的多个要素特征的视角出发, 通过对舆情信息在传播过程中存在的 key 问题以及核心技术挑战进行分析, 从而对当前舆情传播领域中存在的技术问题与研究现状来展开分析和讨论, 为跨媒介条件下舆情信息传播动力学机制的研究提供有效依据和基础。

2.2 舆情信息传播过程的核心问题与挑战

根据 PRCMA 模型, 舆情信息在传播过程中, 传播者的影响力大小、传播内容的争议性与可信度大小、多媒介在信息传播过程中的交叉影响与相互作用、信息接收者的兴趣偏好和群体相互作用对接

2) 凤凰新闻. 上海女孩跟男友回农村过年, 见第一顿饭后想分手. 2016.02.07. <http://i.ifeng.com/news/sharenews.f?aid=106051113>.

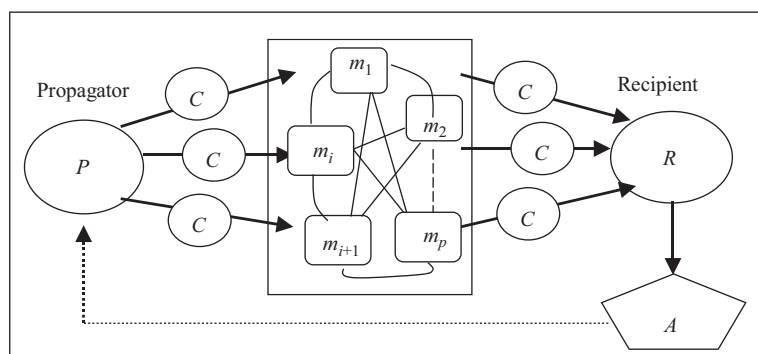


图 1 舆情信息传播 PRCA 多要素模型

Figure 1 Sentiment propaganda model with PRCA multi-factors

收者影响等多种问题和因素的交织影响,使得舆情信息在传播过程中的效果评价具有较高的复杂度,其中,最为核心的问题与挑战如下。

挑战 1. 跨媒介条件下,传播者影响力识别与影响力最大化问题。在信息传播过程中,那些具有不同公信力或影响力的信息传播者(包括新闻媒体与网络社交媒体)在改变信息的传播方向、传播速度、传播周期以及实际传播效果上反映出明显的差异。如何度量这种跨媒介条件下的多层次、不同粒度条件下网络节点传播效果的差异以及影响力特征,并深入研究引起这种传播差异的内在动力学机制与作用机理,为探索信息传播的影响力最大化或最小化的条件和约束机制提供理论与技术支撑,成为目前一个重要的技术挑战。

挑战 2. 不可信信息的识别以及内容可信度的实时度量问题。在传统的舆情传播过程中,信息的真实性与完整性鉴别存在着明显的滞后与时延,往往已经产生并形成了舆论热点或者次生影响后,才逐步被人们发现和证伪,这种基于事后的内容真实性判别结果直接影响到了舆论场的形成以及舆情传播的方向与趋势。为了消除不同类型的不可信信息在传播中对社会公众产生的负面影响,实时或者准实时地识别不同类型信息内容的可信度已成为目前舆情分析与预防过程中迫切需要展开的研究领域。

挑战 3. 跨媒介信息传播的动力学机制问题。由于不同媒介的受众群体之间存在着拓扑结构特征的差异性与用户群体的交叉性,因此,多源异构的信息通过跨媒介的交叉传播与共振,使得传统的基于单一数据源以及单一群体的仓室模型(compartment model)已无法有效地模拟和仿真实际网络环境下舆情信息的传播机制,迫切需要借助于大数据以及复杂网络等技术来探索跨媒介条件下新的信息传播模式与动力学机制。使得跨媒介、多层次耦合的关系网络环境下的传播动力学机制的研究受到了越来越多的重视。

挑战 4. 信息接收者个体兴趣偏好和群体行为对信息接收的影响和度量问题。一方面,接收者个体对信息内容的兴趣偏好直接影响到了信息的接受或者转发等行为的发生概率;另一方面,群体行为也会通过从众的方式影响到个体的价值判断,并最终导致了信息的接受。因此,将个体偏好特征与群体行为特征相结合,形成信息传播过程中接受程度的度量判据,将直接影响到信息传播的质量与效果。

挑战 5. 跨媒介舆情信息传播的效果评价问题。在跨媒介信息传播的过程中,不仅存在着多个要素特征和对象,并且这些特征和对象之间也存在着相互的影响与作用,如何建立一个合理、有效的信息传播效果的客观评价体系,这对于舆情信息的传播、控制以及预警均具有重要研究意义和实际价值。

通过对上述问题和挑战的深入研究,进一步揭示出舆情信息在跨媒介网络传播过程中存在的关系拓扑结构、网络群体行为与网络信息内容之间复杂的交互关系和互动规律^[4],并为研究跨媒介条件下

的舆情信息传播动力学机制奠定基础. 本文将围绕上述 5 个技术挑战, 分析并论述目前最新的技术成果与研究进展.

3 跨媒介信息传播的特征

3.1 传播媒介的研究历程和方向

传播介质是指信息进行传播的载体, 由于载体本身所具有的特征和性质差异, 对舆情信息传播的速度、覆盖范围以及影响力等将产生不同的影响和作用. 根据信息传播媒介的特征可以将相关的工作分为以下 3 个阶段.

第一个阶段, 针对报纸、电视、广播以及门户网站等传统传播媒介的特征进行研究, 它们均以“一对多”的方式来实现信息的“单向”传播, 由媒体与读者之间的接触而形成的用户网络结构呈现出一种多层次的星型辐射特征, 因此, Lazarsfield 将这种传播定义为“二级”传播模型^[5], 即大众传播只有通过“意见领袖”的中介作用才能发挥影响, “意见领袖”的数目以及影响力大小都会直接影响到信息传播的效果与质量; 但是“意见领袖”往往受到区域、大众群体以及媒体的所有权与控制权等问题的局限, Rogers 则进一步提出了“多级”传播模型, 即将“传播流”分解为了“信息流”和“影响流”两个部分, 其中“信息流”的传播可以是一级的, 但是“影响流”的传播则必须是多级的, 从而反映出了媒介影响大众的模式具有显著的不确定性和易变性特征.

第二个阶段, 则针对 Web 2.0 与移动互联时代下以各种社交媒体为介质的网络平台的传播特征研究, 这些网络社交平台使得信息通过 UGC (user generated content) 和“多对多”的方式“双向”或者“多向”地展开互动式、高实时性的快速传播. Schifferes^[6]指出新闻传播技术的变革并没有消除传统媒体, 例如网络社交媒体中的主要内容的创造者仍然来自于那些传统的媒体机构. 但是, 最大的变化在于“意见领袖”在整个信息传播网络中的位置被极大的“前置化”^[7], 并介入到早期的信息发现和挖掘阶段, 从而通过整合社会的“信息流”和“影响流”来左右舆情信息的传播方向和模式.

第三个阶段, 针对网络舆情信息主要通过新闻、微博、微信、论坛等不同类型的网络媒体来扩散传播, 这些异构媒体之间的交叉传播与相互作用突破了信息在单一媒介中简单传播的模式局限, 特别是在不同媒介间信息的交叉传播和相互作用下, 导致了舆情事件产生的环境、主体、客体和媒介等要素都发生了变化, 使得媒体交叉影响与互动传播过程中舆情的形成和爆发期大大缩短. 而这些异质、多源的媒介也导致了信息的交叉传播与整合, 并体现为多个平台、不同媒体和用户群体间的合作、共生、互动与协调^[8], 对于媒体中的热点事件或者话题, 新闻网站中往往会对事件的发生与后续影响和发展进行了更为深入的报道, 而社交媒体则更多体现了民众对事件的关注度以及主观反响, 从而形成了多视角、多来源、多主题下的舆情事件分析结果. 因此, 如何在多源异构跨媒介的条件下, 通过信息的感知、融合和关联分析, 以及不同媒介的用户群体网络结构特征和节点影响力特征, 来深入研究舆情信息在交叉传播过程中的相互影响与传播的动力学机制已成为了目前最具挑战的研究任务之一.

3.2 传播介质的核心特征

社会化的网络传播平台突破了传统时空的局限性, 不仅拥有了规模巨大且类型多样的网络用户, 并呈现出了拓扑结构的复杂性、连接的多样性以及网络生成与演化的动态性等特征, 从而具有了复杂网络中小世界、无标度网络以及涌现等新特性^[9,10]. 由于小世界网络下具有较高的网络聚类系数和相对稳定的平均路径长度, 从而一些在局部高度聚集的节点与一些有助于产生较短路径的长距离节点之

间随机连接, 并产生较快的信息传播途径. 而对于无标度网络中少量具有高度数的中心节点, 不仅反映了个体的交际能力, 同时也反映了占用网络资源的一种能力, 其本质上反映出了“意见领袖”的部分特征. 特别是具有小世界特征的网络与具有节点动态生长和偏好依附特征的 B-A 网络结构相比, 呈现出了更加显著的社区结构^[11]. 而现实的传播网络在结构上往往是不均匀的, 常由多个子网来构成, 在相同的网络舆情演化规则下, 不同传播媒介的特征模型往往会产生具有一定差异甚至是完全相反的结果^[12]. 因此, 本文从以下两个方面对媒介的特征进行分析.

(1) 传播网络节点的结构与分布特征. 像新浪、搜狐和网易等在线新闻门户网站, 网页被点击的次数满足幂指数在 0.60 ~ 1.03 之间的幂律分布, 而网站访问量则满足幂指数接近 1 的幂律分布^[13]. Kou 等^[14] 证明了针对 BBS 简单回复关系所建立的用户网络具有小世界和无标度特性; Li 等^[15] 则进一步验证了 BBS 在话题与浏览数、话题数与回复数等特征上分别服从指数为 1.395 和 1.1177 的长尾分布; Lin^[16] 也指出微博中个人信息转发数的分布呈现出显著的长尾现象. Weng 等^[17] 针对 Twitter 研究发现, 当用户关注数小于 105 时, 社交网络中节点的度分布服从指数为 2.267 的幂率分布. 从上面的研究结果中可以获得一个需要进一步证实的假设, 如下所述.

假设 1. 不同传播媒介的交互性越强, 幂指数就会越高, 分布所呈现出的重尾现象则会更加显著, 即信息在交互性越强的媒介中传播则对“意见领袖”的依赖作用更加显著.

另外, 传播的网络结构与分布模式对信息传播能力的影响存在着显著差异, 例如: 在小世界网络中会加速流行病的传播, 而在无标度网络环境下存在着零传播的阈值或者传播临界值的著名现象^[18]. Zhang 等^[19] 也指出背景网络结构和参数设置对传播具有影响, 特别是无标度网络中存在着不稳定性较高的 Hub 节点, 这导致了无标度网络结构下传播的可靠性要比小世界网络结构低, 但传播效率却更高. Wu 等^[20] 针对度分布和聚类系数可调的社区网络模型中的传播规律研究发现: 流行病传播的效率与网络的度分布和聚类系数相关, 并随着聚类系数的增大而降低, 即随着网络度分布的非均匀性的增强而增大, 因此, 网络社区结构的存在将显著地降低传染病流行的危险性. 而网络中的社区结构会降低流行病传播的阈值, 但是一旦爆发, 则相同传染概率下的感染人群数目却会出现下降^[21]. 因此, 不同传播媒介下的用户群体网络结构特征与分布模式对信息传播速度、覆盖度以及深度等传播动力学机制存在着显著差异, 本文进一步提出如下假设.

假设 2. 不同媒介的用户所形成的网络拓扑结构中, 聚类系数越高, 则网络结构社区化的特征就越明显, 即信息在该媒介上的传播速度越低.

如何根据媒介网络结构特征的差异建立与信息传播能力之间的联系, 并通过建模和定量化分析对上述假设进行验证则是目前一个重要的问题.

(2) 舆情网络节点属性特征. 传统的独立级联模型和线性阈值模型均将信息传播视为离散的、等概率的传播过程, 对传播网络的结构特征、信息发布者的影响力、权威度、信息内容的话题属性、可信度、信息受众的兴趣偏好等影响因素考虑不足. Yang 等^[22] 研究了信息传播的速度、范围和深度问题, 发现用户的个体属性和信息的话题分类在决定信息传播的速度、范围和深度方面发挥着不同作用; Lagnier 等^[23] 则在一个改进的线性阈值模型 DRUC 的基础上进一步研究了信息内容与用户轮廓模型对信息传播概率的影响, 并指出用户对信息内容的兴趣度、用户传播信息的意愿以及已感染信息的邻居节点的数目直接会影响到信息传播概率的大小. 为了对微博信息的传播过程进行动态预测, 周东浩等^[4] 发现传播网络节点属性特征 (如兴趣偏好、行为特征)、内容的负面情感特征以及信息发布者的权威度等因素将导致信息传播过程中出现异步性, 这种异步性从侧面反映出了信息在网络传播过程中的动力学特征. 可见, 网络节点属性特征和舆情信息的内容特征对舆情信息的传播概率以及信息传播过程中的预测能力将产生重要的影响作用.

3.3 跨媒介传播过程中的信息延迟特征

跨媒介传播是指信息在不同媒介之间的流动, 它至少包含两层含义: 其一是指同一条信息在不同媒介之间的交叉传播与整合; 其二是媒介之间的合作、共生、互动与协调. 本文主要研究跨媒介传播的第一层含义. 对此, 许多学者做出大量的研究工作, 例如: Fang 等^[24] 率先将政府网站、门户网站新闻、传统媒体以及网络社区 4 类不同结构特征的媒介组织成一个跨媒介的传播网络, 通过建立 4 个流位变量、8 个流率变量等共 33 个变量形成的一个跨媒介传播的舆情热度流模型, 尽管该模型也考虑到了时间对舆情事件热度产生的影响, 但是没有深入考虑到信息在不同媒介中传播的差异性特征. Di 等^[25] 则分析了事件作用、网民作用、媒体作用和政府作用 4 个方面对网络舆情热度的影响并对舆情热度流图进行了进一步的仿真验证, 发现基于网络用户的线上与线下异步行为对舆情传播也产生影响. Spiro 等^[26] 在研究 Twitter 信息传播延迟中也发现: 用户活跃度、权威度和可信度的高低, 以及消息内容是否包含标签和 URL 链接等特征对信息的转发延迟将产生直接的影响作用. 特别是, Twitter 中发布的新闻信息有时要比传统新闻机构发布的更快且更细致, 甚至还包含了传统媒体中没有涉及到的一些细微的事件信息^[27].

斯坦福大学的 Leskovec 等^[28] 利用所提取的 9000 多万条博客以及主流媒体的新闻信息进行挖掘时发现: 绝大多数情况下, 主题事件均由主流媒体向博客进行传播, 传播峰值的延时约为 2.5 小时, 而且仅有 3.5% 的事件是由博客向主流媒体传播的. 另外, Heimbach 等^[29] 在一个德文新闻网站中抽取的 4278 篇新闻样本来验证了它与 Twitter, Facebook 和 Google+ 等不同在线社交媒体之间通过内容分享所反映出的跨媒介信息传播的差异问题, 其中, 通过对媒体内容的情感、作者、位置、内容主题分类以及时间分布等特征属性, 发现内容特征与媒介之间存在着强烈的相关性, 并且不同的内容特征会直接影响到信息内容从网站向上述 3 个不同的社交媒体之间分享与传播的能力与效果. 另外, 体育类新闻以及内容具有强烈情感特征, 特别是有趣或者刺激让人产生忿怒情绪的文章往往在这 3 个社交媒体中都易于转发和传播. 那些具有负能量的新闻则更易于在 Twitter 和 Google+ 中传播, 而具有正能量的信息却更易于在 Facebook 中传播, 但是内容可用性高的文章却很难在 Facebook 中传播; 另外, Twitter 和 Google+ 的用户更倾向于政治、商业、技术与科学主题的内容, Google+ 的用户同时倾向于传播生活方式与职业生涯相关的主题文章, 但是这 3 个社交媒体都倾向于采用和传播原创的内容, 而新闻机构所提供的文章很少会引发传播. 该研究结果反映出信息传播与信息内容特征以及媒介用户群体的偏好特征之间存在着显著的依赖关系. 上述研究工作再次表明: 不同媒介在舆情信息传播过程中的延迟以及对内容偏好的差异反映或者影响了媒介传播能力的差异, 这为研究跨媒介环境下的信息交叉传播与相互作用以及传播动力学内在机制奠定了基础.

综上, 跨媒介的舆情信息交叉传播与相互作用颠覆和改变了传统单媒介条件下的舆情形成、传播与演化机制, 并使得舆情传播与复杂网络的动力学机制相融合. 另外, 由于不同媒介的特征差异性, 对于跨网络环境下不同用户群体的接受行为的影响也存在着明显差异, 通过一些媒介的信息传播, 甚至产生用户线下与线上行为的共振^[30], 推动或者加剧了舆情在不对称条件下的传播、扩散与演化, 从而成为了目前人们研究和关注的热点^[31, 32].

4 信息内容特征对传播的影响

随着用户创造内容成为 Web 2.0 媒体的主流模式, 在享受数据信息快速获得与共享的同时, 人们不仅深刻地感受到在网络中传播的海量错误、谣言, 未证实或被操纵, 甚至是虚假信息所产生的负面影

响,同时发现识别出真实可信的内容信息所需要的技术复杂度也越来越高. Howell^[33] 将海量数字化的虚假内容信息列为了影响现代社会发展的重大威胁之一,不仅微博被大量的垃圾和谣言信息所严重污染^[34],而且在线博客中也包含着许多偏激或者是虚假的内容,甚至在线新闻媒体中也充斥着大量不可靠且没有被证实的新闻,这直接影响到了主流媒体^[35]. Gupta 等^[36] 的研究结果显示:在 Twitter 中有将近 52% 的内容是确定可信的 (definitely credible)、35% 的内容是大致可信的 (seems credible)、13% 的内容是确定不可信的 (definitely incredible). 而 Zhou 等^[4] 在对微博传播能力研究中发现,负面消息往往比正面消息更容易吸引注意力,并具有更强的传播能力. 当针对虚假信息或者谣言内容不加甄别地通过社交媒体平台进行快速传播的过程中,极大地渲染了消极和负面的社会情绪,甚至损害了公众对社会与政府的公信力,以及法律和道德的底线. 美国计算机协会 ACM Transactions on Information Systems Journal 在 2015 年 5 月和 2016 年 8 月均针对社交媒体中信息的真实与可信性出版了一份专刊 (*Special Issue on Trust and Veracity of Information in Social Media*),而在 SemEval-2017 大会³⁾中针对谣言的检测任务也提出了明确的测试任务要求. 因此,在舆情传播过程中,针对具有高度社会性与敏感性的信息内容的真实性、可信度以及事实的检测与评价方法已成为目前重要的研究热点之一^[37]. 下面针对不可信信息在网络媒体中的传播特征以及相关的问题与研究进展进行分析综述.

4.1 不可信信息的分类与特征识别

根据文献 [38~42], 本文将网络传播过程中的不可信信息总结并分为: 网络谣言、垃圾信息、误报信息、虚假信息和偏激信息 5 种类型^[38]. 所谓谣言, 是指在社会中出现并流传的未经官方公开证实, 或者已经被官方所证伪的信息. 谣言往往具有争议性, 且与事实的真实性存在偏差甚至相反的一种陈述, 通过传播谣言使信息受众做出一系列举动以达到特定的目的. 而垃圾信息通常包括了无用且重复的信息以及黄色反动信息等. 误报信息则是由于人为错误而导致发布的信息存在错误或偏差, 一般地, 误报信息也是信息的一种特殊类型. 虚假信息则是在与他人交流的过程中, 通过扭曲事实或脱离客观事实, 传播不真实的信息以满足人们猎奇和表现欲的心理. 而偏激信息往往是人们对现实社会认知和个体情绪的一种主观反映, 通过夸大事实信息、断章取义或者以偏概全等言论在一定程度上影响了公众对社会舆情信息的立场和价值判断.

根据上述定义以及不同类型信息特征的分析, 如何通过主题检测、作者特征抽取以及数据来源检测等内容特征来检测和识别信息所属的类型. 其中, 针对谣言的判别, Takahashi 等^[43] 通过对信息内容进行命名实体的抽取, 从而获取日期、地点、人物和组织机构等实体特征, 从而形成对事件的判断并实现了对这些信息特征的真实性检测. 而 Hamidian^[39] 利用了时间、标签、URL、转发等 Twitter 和特定网络特征, 以及内容的一元和二元模型等语法特征信息, 并首次采用了 100 维 Twitter 潜在向量 (TLV) 语义特征进行谣言检测. 针对垃圾信息的识别, Wang^[44] 利用了基于用户粉丝和关注的有向图特征以及 Twitter 信息内容本身的特征 (如: 重复 Tweet、评论与 @ 某人、URL 链接和话题等) 进行融合, 并实现了对 Twitter 信息内容的自动检测. 而 Ratkiewicz^[40] 在 Twitter 数据集上, 利用垃圾信息内容中的 Meme 类型、用户行为、网络结构等 18 种特征信息对垃圾信息进行了分类. 而针对误报信息的识别, Qazvinian 等^[41] 使用了基于内容的特征、基于网络的特征和 Twitter 特定的 Meme 特征相结合, 并实现了误报信息的检测. Kumar 等^[45] 发现网络社区中聚集了大量具有共同兴趣和偏好的人, 社区内的成员更倾向于发表看法或态度相近的观点, 这会导致误报信息在网络上很长的时间内不易消失; 同时, 也导致了人们将错误的世界观和意识形态在公共、政治以及宗教问题上持有某种偏

3) SemEval-2017. <http://alt.qcri.org/semeval2017/>.

见. 为了避免单纯的内容特征对信息判别上产生的误差, Karlova 等^[46] 则采用了众包的方式利用网络群体的智慧来实现误报信息的识别与控制. 另外, 针对蓄意虚假信息的识别, Karlova 和 Lee^[42] 从真实性、准确性、完整性、时效性和欺骗性 5 个维度来识别和区分误报信息、虚假信息和政治宣传信息, 其中, 蓄意欺骗的信息均属于虚假信息.

4.2 信息可信度定义与判别

由于网络信息来源的不确定性, 特别是一些具有人为操作的偏激观点、虚假信息, 甚至是谣言的传播将会直接影响了信息受众对热点新闻事件的态度与观点倾向, 并改变了舆情传播的方向和结果^[47, 48]. 为了更有效地识别这些虚假的内容信息, 人们定义了信息的可信度评价指标. 根据相关研究成果, 本文将信息内容的可信度度量分为两类: 一是将信息内容的属性特征用来作为信息可信度的评价研究对象; 而另一类则是指一个人相信信息的真实性并能够接受的程度, 它基于内容, 但也会受到不同人员偏好和环境共同的影响.

针对第 1 类度量方法, 早期的研究主要针对在线新闻门户中信息的可信度分析, Abdulla 等^[49] 在对新闻媒介可信度分析的基础上, 利用可靠度、流行度和倾向度 3 个维度来对在线新闻内容的可信度进行评估. 随着社交媒体在舆情信息传播过程中的重要性越来越高, 特别是社交媒体中所具有的短文本、高实时性、内容与用户特征相互交织的一些新特征, 对信息内容以及信息质量评价提出了更高的要求. Li 等^[50] 针对社交网络中的相关特征, 指出社交网络的结构以及人们的行为倾向对信息传播过程会产生非常大的不确定性, 并且传播内容的可信度与网络的结构特征、个体行为和信息传播的初始状态之间存在着密切的关系. 例如: Schwarz 等^[51] 提出了在线文本的有用性与价值性作为可信度的判断标准; Odonovan 等^[52] 针对 Twitter 内容可信度提出了关注数、URL、转发数和内容的长度这 4 个最佳的评估指标; Schwarz 等^[51] 和 Metzger 等^[53] 则认为: 信息的可视化模式比信息内容以及来源对可信度评估的结果影响更大. 而 Gill 等^[54] 在构建公众信任时, 将参与度、完整性以及目的性作为分析的关键, 他利用自己所开发的 TweetLevel 工具对 Twitter 中的信任信息进行了度量, 结果表明网络中的个体更愿意信任那些和自己建立联系的用户所发表的, 并具有一定转发数与引用数的信息内容. Xu 等^[55] 则针对 Web 信息可信度的时效性、权威性、影响力和关注度 4 个特征进行验证, 并提出了一个多维度加权的可信度计算方法. Gupta 等^[36] 总结了评估 Twitter 中内容可信度的 10 个重要特征, 即: 推文 (tweet) 中包含的链接、特征数目、独特的特征数、字数、用户简介中的地域信息、转发数、推文的寿命、用户粉丝列表与粉丝最新信息 (statuses/follower)、用户好友与粉丝数的比值 (ratio of friends/follower) 以及所含的 URL 等. Hardalov 等^[35] 也提出了一个基于语言学、可信性 (大小写、发音、拼写与情感) 以及语义 (embedding and DBPedia data) 三者融合的富特征条件下, 自动化的英文信息可信度识别方法, 实验结果表明在特定的测试集下, 内容可信度的识别率居然达到了 99.36%.

针对第 2 类度量方法, Fang 等^[56] 则针对网络中信息内容安全提出了一个基于 ICCON 的控制与评价机制, 并将信息内容识别、人员识别以及行为动机识别作为信息内容安全控制的基础框架, 率先将信息内容的特征识别与动机识别检测纳入到安全信息传播与控制过程之中. Castillo 等^[57] 将信息来源、主题、信息传播特征与作者的声誉和写作风格特征相结合, 并且通过时间特征与 Twitter 中的信息可信度紧密相关. Zubiaga 等^[58, 59] 认为在社交媒体中声誉高的用户所传播的内容往往具有更高的可信度, 同时, 那些具有高信任关系的用户, 他们之间的相似度远大于那些不具有信任关系的用户之间所具有的相似度, 这反映出人们在接受信息时的某种心理倾向. 综上所述, 本文将信息内容可信度的相关指标或线索从信息来源、信息内容、主题、传播者以及传播特征 5 个角度进行总结如表 1 所示.

表 1 网络社交媒体中内容可信度评价线索与指标

Table 1 The evolution clue and index for content reliability in social media network

Class	Index and clue
Information source	Basic features: visit num/prize num/rank/official ownership/domain types/topic/is ICP/is well-known/is authenticity. Structure features: simplicity navigation/website structure/editable reply advise/readable strength/is public aesthetic/source advertisement num/advertisement quantity/credibility other data credibility in website. Interactive ability: user search/information validation/timeliness of user response.
Information content	Basic features: the length of content/is spelling error/is syntax error/the reference of content and data/contain URLs/contain label/timeless of content. Participant degree: focus num/resport num/popularity. Syntax features: is true/text case/punctuation/personal pronoun/the length of content/URLs num. Semantic features: the professional of content/understandability/rationality/valuability/usefulness/reliability/orientation.
Topic	The content of topic features: topic tags/topic content segmentation extraction/topic semantic features/positive scoring/negative scoring/topic sentiments. Topic total features: content num in topic/the average length of content/the field of topic/expression num in content in topic. Topic user features: average age of user/user num/validation user num/the num of topic is referred by users/the sex of users in topic.
Information disseminator	Basic attribute features: sex/location/liveness/social platform/portrait/identity validation/level/types of user/frequently of post/focus num/follow num/browser num of user home page. Preference features: the motivation writing and spreads/reputation/similarity degree. User behaviour features: report comment syntax/semantic/comment sentiment/supported.
Propaganda	Propaganda quantity features: propaganda degree num/max num/audience num. The network of content propagation: propagation subtree structure/average depth of propagation/max depth/propagation path/content dispersed structure/content delivery structure/strongly connected component/network density/average clustering sparse/average path length/the distribution of degree/power law distribution/match pattern. A single user attribute in a network of content propagation: content dissemination form a network penetration/degree/clustering coefficient/betweenness/closeness/interest rate/the degree and the degree of correlation/weak star nodes and star effect/homogeneity.

4.3 传播信息内容的实时检测

由于网络中传播的内容信息具有可传播性、真假未知性、时效性和爆发性等特点,特别是针对互联网上快速判断谣言传播真假的迫切需求,除了 DEFFUANT 和 SZNAJD 等经典的模型外,基于 Markov 链的 D-K 模型、物理相变理论的 Potts 模型、元胞自动机模型以及基于复杂网络条件下的谣言传播模型成为了人们研究与关注的焦点^[19]. 随着信息内容的规模以及处理实时性的要求越来越高,这对内容可信性的评估方法与计算能力也都提出了更高的要求. 在上述模型与特征的基础上,人们利用机器学习算法对原有的模型进行优化与工具化,例如 Ratkiewicz 等^[60] 针对微博信息传播过程中的内容可信度评价开发了一个 Truthy 工具来实现内容的快速评估; Castillo 等^[61] 则采用决策树算法对内容可信度进行计算和优化; Xia 等^[62] 则利用基于监督学习的 Bayes 网络进行了算法的改进; Zhao

等^[63]则提出了一个概率图模型用来推断信息的真实性和来源的质量, 上述研究工作面对海量实时数据处理过程中的实时响应能力与评估的准确性仍有待进一步提高. Gupta 等^[36]在总结 Twitter 内容可信度特征的基础上, 利用 SVM-Rank 方法开发了一个 TweetCred 工具来实现对 Tweet 内容可信度的实时评价, 其中在训练和测试数据集下的处理时间分别为 9 ~ 10 秒和 1 秒内, 而归一化累积获得折扣值 (normalized discounted cumulative gain, NDCG) 最高达到了 0.7219, 并且在 6 秒的时间内对 540 万的 tweet 内容中 80% 数据进行了信用评分计算与结果展示. 但是面对海量且不断动态增加的数据集合, 除了准确检测出热点的舆情议题以及突发事件外, 如何结合更有效的大数据处理手段来协助解决上述问题也成为目前研究的前沿任务之一^[64].

5 传播者动机与影响力特征对信息传播的影响

在跨媒介信息交叉传播过程所形成的网络拓扑结构中, 传播媒介则成为了异质网络中的 HUB 节点, 特别是在传统媒介与新媒介之间的互作用与交叉影响下, 信息传播也具有了“优先连接”的特征, 即传播能力与信息传播节点 (包含传播媒介) 的重要性和影响力大小具有密切的相关性. 如何采用定量分析的方法来度量跨媒介信息交叉传播的复杂环境下, 传播节点的重要性程度、影响力特征以及对信息传播能力的影响作用, 则是基于复杂网络条件下, 舆情信息传播过程中亟待解决的重要问题之一^[65]. 下面从传播者动机以及影响力特征对信息传播的影响作用来进行分析.

5.1 传播者动机对信息传播的影响

节点影响力大小往往受到传播行为和动机以及内容等多种要素的作用和影响, Granovetter^[66]曾根据人们之间的互动频率、感情力量、亲密性程度以及互惠交换能力 4 个要素来度量人际关系在社会网络中的传播能力, 特别是将互惠交换能力定义为人际网络中能够进行交互传播的一个重要动机来源, 即分享与传播信息的目的也是为了更好的获得对等的信息资源. Bessi 等^[3]在针对 Facebook 的研究中也证明了类似的结果, 即用户选择转发或传播信息的类型 (如科学类或谣言类) 往往和与自己观点相同的好友数目有关. Zhou 等^[4]则将微博看作是一个融合了社交图谱和兴趣图谱的关系网络, 每一个网络节点均融合了关系特征和兴趣偏好特征, 并且网络舆情信息传播的概率与网络节点之间的结构相似度以及节点用户对信息内容的兴趣偏好具有强相关性, 同时, 用户兴趣偏好在一定条件下也容易转化为信息传播的动机和行为. 为了进一步梳理用户兴趣和情感与传播动机的关系, Qian^[67]针对国内虚拟品牌社区中的信息传播过程抽象出了积极、消极和中立 3 大类、13 项传播动机的核心要素. 其中, 积极动机包括分享快乐、提升影响力、奖励动机、帮助企业和帮助他人动机; 消极动机包括负面情感发泄、心理安慰、警示他人、寻求补偿和报复; 中立动机包括参与互动、支持社区发展和习惯性动机等. 这些动机的划分为使得人们可以在对用户兴趣偏好建模的基础上, 为用户传播信息的动机和意图识别建立了一定的映射关系, 从而为进一步的深入研究奠定了基础.

5.2 传播者影响力的评价与度量方法

在信息传播网络中, 传播节点的影响力不仅可以通过基于节点自身的物理属性特征和网络拓扑结构特征来进行度量, 同时可以利用邻居节点之间的相互链接特征来弥补上述度量方法的不足, 并成为目前的主要的度量方式. 下面对这 3 种主要的影响力评价与度量方式来进行介绍.

(1) 基于节点物理属性特征的影响力度量方法. 该方法主要采用网络中节点自身的物理特征属性作为度量节点影响力大小的依据. 例如, 传统新闻媒体的发行量、电视节目的收视率、社交网络中节点的粉丝数、好友数、用户被提及数、原创内容被转发数和回复数, 以及用户所具有的朋友圈数或者是参与的社区数等等. 在这些特征中, 面对社交网络最常采用的是“粉丝数 + 被提及数 + 被转发数”和“粉丝数 + 回复数 + 转发数 + 评价数”等组合来进行度量, 我们在针对 BBS 社区网络节点影响力的研究过程中, 也曾将内容的浏览次数、用户的文章帖子数、内容的回复数和粉丝数这 4 个基本特征在 sigmoid 函数下进行加权处理来实现用户影响力的计算^[68]. 这些静态特征虽然能够在一定程度上反映出个体在网络中重要性与影响力, 但存在两个明显的不足: 一是无法充分体现出社区中的动态特征与演化行为以及社区网络环境对节点重要性的影响; 二是在大量虚假信息中, 使得原有特征的评价结果失去了真实性与有效性. 根据 Yahoo 的报道^[69]: 许多 Twitter 中的明星用户拥有巨大数量的虚假粉丝用户群, 例如: 奥巴马粉丝中的虚假用户数达到了 42%, 而比尔·盖茨的粉丝中虚假用户数居然高达 50%, 这些虚假用户信息直接影响到了用户的影响力、个性化推荐计算结果的准确性, 同时也带来了虚假信息和虚假广告传播的空间. Shen^[34] 进一步指出: Twitter 上 62% 的虚假用户属于发贴数与粉丝数均小于 10 的垃圾用户, 并且 57% 的虚假用户属于不活跃的“僵尸”用户, 为了区分出这些虚假用户, 他提出了基于“贴子特征 + 用户特征 + 演化特征”相结合的用户识别方法. 正是由于存在着大量的虚假粉丝用户, 使得这种基于单纯物理特征的分析方法失去了客观性和真实性, 因此, 该方法常常被用作数据的预处理.

(2) 基于网络拓扑结构的影响力度量方法. 该方法借助于社会网络分析中的中心性指标 (centrality) 来度量节点在网络中的作用、权力以及中心地位, 从而反映出该节点在网络中的重要性程度和影响力. 一般地, 中心性包括了局部中心性和全局中心性两类. 局部中心性反映了网络中一个节点与其他节点之间存在的直接联系, 如果联系较多, 则节点就居于网络的中心地位, 并拥有较大的影响力. 这种方式与节点物理特征中的粉丝数度量方法存在一定的相似性, 为了进一步改进, 研究人员一方面从网络结构的不同角度提出了不同的中心性度量方法, 如接近中心性 (closeness centrality)、介数中心性 (betweenness centrality) 和特征向量中心度 (eigenvector centrality) 等; 另一方面对于中心性度量方法进行了改进. 例如, Chen 等^[70] 基于节点之间的拓扑关系, 通过计算相应节点的所有邻居节点以及次近邻居节点数之和来扩展了一个多级邻居中心度指标 (multi-level local centrality). Centola^[71] 发现节点在信息传播过程中的重要性与该节点在网络中的聚集性有关, 并且在高聚集网络中的信息传播效率更高. Ugander 等^[72] 研究发现: Facebook 中朋友关系网络下, 邻居节点的绝对数目不是影响节点重要性的决定性因素, 起决定作用的是邻居节点之间连通子图的数目. Kitsak 等^[73] 则在连通子图的基础上提出了一个基于 K -核分解的网络核心度 (coreness) 概念, 由于该指标考虑到节点周围邻居的聚集性程度和影响力大小, 因此, 具有高核心度 (K -core) 的节点在网络中往往也具有较高的影响力. 同时, 通过在社会关系网络、邮件网络、病人接触网络中心的实证结果表明, 在单信息来源的网络条件下, HUB 节点或者具有高介数中心度的节点不一定是最有影响力的节点, 而通过 K -核分解确定的核心节点才具有最大的影响力; 但在多信息来源的条件下, 具有高度数的 HUB 节点却比 K -核节点具有更高的传播效率. Zeng 等^[74] 进一步考虑到 K -核计算过程中被移除的度数为 $K-1$ 层的节点对 K 层节点的影响作用, 提出了将移除节点的度与 K -核节点的度进行加权处理的优化混合度分解法 (MDD).

而全局中心性则反映了节点对网络资源控制的能力和程度, 它从全局化网络的角度来审视了网络中节点的重要性与影响力大小^[75]. 除了采用经典的特征向量中心度、节点可达性 (accessibility) 指标、接近中心度、紧密度等全局网络中心性指标外^[76], 一些学者则从网络抗毁能力的角度来分析节点的

影响力,即具有重要性的网络节点一旦从网络中被删除,则将会对网络的连通性直接造成影响,且节点的重要性越高,则对网络的破坏性就越大^[77].例如,Li等^[78]在度量网络节点(集)的重要性指标时提出:当删除网络中某一个节点后,网络中所形成的不连通节点之间的最短路径的倒数之和反映了该节点的影响力大小;而Yu等^[79]则利用移除节点后的网络直径与网络连通度变化的梯度大小来评估节点影响力的大小.此外,Huang等^[80]针对美国1996~2006年间公司董事共现网络结构的实证研究发现,公司董事的影响力取决于他能够获取掌握公司信息量的大小以及与其他董事之间最短的联系路径,即网络中信息内容的流量大小以及网络拓扑结构对节点的影响力来说都至关重要.

(3) 基于网络链接的节点影响力度量方法.该方法是一种基于随机游走的节点重要性排序方法,本质上反映出了节点之间的链接关系,并通过随机游走策略来识别相链接的邻居节点的重要性,从而判断出该节点的重要性排序.其中,典型的方法均是在PageRank算法的基础上,将节点自身的连接度与周围邻居节点的影响力大小相结合,当邻居节点的影响力越大,该节点影响力大的可能性就越高,从而实现了对节点重要性与影响力的计算与评价.为了解决传统PageRank算法中存在的主题漂移、忽略用户兴趣及偏向旧网页等问题,研究人员提出了大量的优化算法^[81~85].在此基础上,Weng等^[17]考虑到节点特征以及链接结构特征,提出了Twitter中用户影响力的评估指标TwitterRank,实验表明用户粉丝的质量与微博内容的质量和用户的TwitterRank值大小具有相关性;Lü等^[86,87]考虑到网络结构特征的变化会直接影响到PageRank算法中的链接“跳转概率”,同时为了保证网络的连通性,通过增加一个“Ground Node”节点来实现与所有节点之间的连接,在此基础上提出了LeaderRank算法以及通过加权优化来计算网络意见领袖的影响力大小.Xu等^[88]进一步考虑到用户之间存在的观点差异以及恶意注册用户对算法的影响,对LeaderRank算法进行了改进并提高了算法的准确性与抗干扰能力;Guo等^[89]在对所提出的UserRank模型进行优化的基础上,将网络用户的影响力分析应用到一个云存储系统中,实现了基于用户影响力大小的个人隐私数据保护和优化策略.

上述3种方法分别从网络节点物理属性、网络拓扑属性以及网络链接3个不同特征的角度对传播或者关系网络中节点的重要性与影响力进行了识别判断,越来越多的研究工作不仅采用了用户节点的相关特征,同时也将信息内容的特征、用户情感与行为特征以及不同媒介的可信度特征融入到相关的影响力评估算法之中,并成为了目前节点重要性与影响力评估与优化的重要方向.

5.3 节点影响力对信息传播过程的影响

由于社交网络连边上存在的关系强度的异质性导致了只有部分节点之间的连边对信息传播存在实质性影响,即网络中,部分节点之间存在的关系或者说部分节点所具有的影响力对信息的传播产生直接的影响作用.Huang等^[90]将对信息传播产生直接作用的节点和连边的拓扑结构抽象成一个网络影响力骨架,并从微观和宏观两个层面研究了节点和连边对传播的作用以及影响力骨架的连通性和扩散率,发现具有结构洞地位的局部中心节点导致了信息传播过程中存在一定程度的脆弱性,且其传播效率低于对社交网络本身的预期.

由于这些局部中心节点有时会表现为社交网络中的意见领袖,为了进一步度量意见领袖的影响力对传播过程的影响,Wang等^[91]采用了动态有向图来描述消息在微博网络中的传播过程,并将这些过程分解为由各个意见领袖所驱动的子过程,从而对节点的初始影响力、影响力衰减指数以及影响力持续时间等指标进行了度量,发现尽管意见领袖的影响力衰减指数大小与影响力持续时间的长短以及信息传播过程中粉丝的数量几乎无关,但是,信息传播的范围与意见领袖的参与数量存在着弱相关.而在实际情况下,用户影响力与其身份地位以及发布的内容质量高度相关,用户在该领域的影响力和权威度越高、发布信息的质量往往越高、内容的吸引力与新闻性越强,对该消息感兴趣并愿意传

播的人数就会越多. 此外, 关系网络中的一个用户也往往会受到多个信息来源的影响, Bakshy 等^[92]将在 Twitter 网络中信息传播过程中用户之间的影响力分为: 最先影响力 (first-influence)、最后影响力 (last-influence) 和均分影响力 (split-influence) 3 种类型, 它们分别表示: 在所有的影响来源中, 最先 (或最后) 发布该条信息的节点对其产生的影响力最大, 或者是所有影响源具有等价的影响力, 并且可以归一化平均所有的影响力, 根据这 3 种影响力分类策略可以建立不同的信息级联传播模型.

此外, 除了网络节点的拓扑空间对节点影响力产生影响外, Goyal 等^[81]以 Jaccard 系数和 Bernoulli 分布为基础, 在离散和连续时间两种情况下发现网络中节点间的影响力大小随时间而衰减. 上述研究工作表明, 节点的影响力是一个动态变化的指标, 它随着网络空间结构、时间以及网络信息内容和流量的不断变化而发生动态变化, 即在动态的关系网络演化过程中, 用户的影响力将随着主题事件以及时间的变化而呈现出不同的大小和作用. 因此, 节点影响力对信息传播过程的影响不仅需要考虑到网络的宏观与微观的静态特征, 同时也需要考虑到动态的内容主题、用户行为特征以及网络社团结构特征, 以及外部因素的影响 (例如政府信息公开的速度以及应急处理机制等^[24]), 这些研究工作成为了目前一个新的研究热点^[93,94].

6 接受者偏好与群体作用对信息传播的影响

一般地, 凡是符合或者能够迎合人们的主观愿望、主观印象或者主观偏见的谣言往往最容易使人相信并乐于传播^[95]. 网络上的社会化群体事件不仅与区域经济发展和人员素质相关外, 而且与参与者对诱因的认知、参与机动以及决策行为关系紧密相关, 特别是用户在决策时容易受到外界特殊线索与信息的诱导, 倾向于接受符合个人需要的解释, 而拒绝现实的客观且合理的理由^[96]. 但在传统的信息传播模型中, 例如传染病传播模型等, 信息接收者个体往往具有相同的感染概率, 从而弱化了个体偏好、差异性以及对周围环境的影响. 但是在实际的信息传播过程中, 通过个体和个体之间、个体与群体之间以及不同群体之间的相互作用, 传播者与接收者的行为偏好与影响力特征、信息内容的敏感度以及传播媒介的品牌与信用定位等, 对于传播的效果均存在着直接的影响^[97]. 特别是个体对传播信息的认知、情感倾向、兴趣意向和行为活动等, 除了受到个体因素的影响外, 也会受到他人或者群体的影响, 并表现出从众、众从和利他等群体性的行为特征, 这些行为使得群体中个体对信息内容的倾向和行为趋向一致, 并达成“群体精神的统一性”^[15]. 因此, 在舆情信息传播并被接受的过程中, 如何根据信息接收者的个人兴趣和群体行为特征所产生的影响来判断信息的接受程度一直是人们关注的焦点.

6.1 接受者的兴趣偏好对信息传播的影响

研究表明^[91]: Twitter 和新浪微博中分别有 31% 和 47.8% 的微博为转发内容, 转发已成为了社交网络中不同节点之间进行信息传播、分享和获取信息的重要形式, 并且这种转发增长的模式具有明显的时效性和话题性, 信息传播的速度随时间而快速衰减, 但也有可能由于公共性话题性引起了用户的关注与转发, 并导致了二次或多次的快速传播, 从而表现为信息传播过程中的波动性与突发性. Huberman 等^[98]进一步针对 Twitter 中的话题传播发现: 热门话题的传播与用户的活跃度以及粉丝数量无必然关系, 相反, 用户之间对传播内容所达成的共鸣会导致话题的快速传播.

Berger 和 Milkman^[99]针对纽约时报中的文章进行研究后发现, 信息内容情感特征与传播能力相关, 具有积极和正能量的文章容易被分享, 特别是那些能够刺激读者产生忿怒或惊愕情感的文章则会更加易于被传播, 这从侧面反映出了在舆情信息被接收的过程中用户个体兴趣倾向对传播行为的影响.

Yuan 等^[100] 则进一步指出在接收传播的信息时, 首先应该确保内容的主题是用户所关心和有兴趣的, 其次才需要考虑信息的权威性与可信度. 特别是, 相对于传播信息的精确性, 人们更希望提高信任的效率, 这一方面倾向于选择一个可信的平台作为信息提供的来源; 另一方面则倾向于那些花费时间较少的、包括短文或者图片的碎片化信息. 此外, 除了个人名誉、认可度、一致性、违反预期和启发性等特征外, 信息本身的有用性、完整性、准确性、实时性以及重要性等也是接收信息的重要判据. 因此, 为了促进信息的转发行为, 深入研究接受者兴趣偏好对信息传播过程中的影响作用, 则成为了目前广告传播、电子商务营销、舆情分析等领域中的一个重要的研究热点.

6.2 网络群体行为对信息接受者的影响

群体行为由于受到传染性、社会性与随机性的影响会产生不确定性, 并表现为群体行为的涌现性特点, 即当一些非线性的微小局部变化积累达到一个特定的阈值时所引起整个系统产生临界相变, 并导致了网络群体、网络结构、群体行为以及信息扩散数量上的涌现^[101]. 这种涌现现象是通过自组织的方式, 将具有共同关注点或兴趣的人进行聚合, 并通过相互作用、相互影响而达成集体行为的动态过程, 这种集体行为则是一种从个人行为出现、叠加到相互传染和互动, 形成了一种具有不稳定性 and 演化特征的动态行为. 而网络中的群体行为特征本质上反映了网民观点从分散到一致的同步行为^[102], 即通过网络信息的传播、讨论与演化而达成群体心理一致的复杂非线性的同步涌现过程, 该过程不仅包含了信息传播网络拓扑结构的宏观动态演化, 也包括了网民个体观点的微观交互传播, 研究表明信息传播网络存在从众性阈值, 当网络中的从众性指标大于该阈值时, 则网民观点具有明显的一致性同步收敛过程, 否则将会形成并保持多种舆论方向共存. 上述研究与经典的传播效果心理一致性模型^[15] 有一定的相似性, 该模型认为信息传播的效果往往取决于传播内容对信息接受者的固有信仰、观点、态度的威胁或强化程度. 可见, 在舆情信息的传播过程中, 群体性行为会对其成员接受信息的程度和能力产生直接作用与影响. 为了针对网络群体行为进行评价, Wang 等^[101] 进一步提出了一个网络群体行为分析评价的模型框架, 该框架如图 2 所示.

综上所述, 针对信息接受者的个体兴趣和倾向特征以及与群体行为特征进行研究, 不仅为信息在传播和接受过程中的个人特征与行为的作用提供分析视角, 同时通过对群体行为模式的研究, 来理解舆情传播与演化过程中存在的同步现象, 这对于舆情态势与传播趋势的预测研究具有重要的作用.

7 信息传播的评价模式

网络舆情信息的传播形态、内容以及方式等明显地影响到社会结构、群体行为与社会心理、社会经济形态以及社会的和谐程度, 分析信息传播绩效的影响因素、作用关系、评估模型与方法能够更加有效地促进舆情传播的质量与机制研究. 信息传播效果是指信息通过各种网络媒介传播的过程中, 使接受信息的用户在认知、情感、态度和行为等方面产生的反应与影响^[103], 它不仅依赖于传播者的行为特征、信息传播的特征, 同时也强烈地依赖于传播内容的质量与传播媒介的特征. 例如, 从传播者行为特征的角度, 信息传播的效果可以通过传播内容对信息接受者固有信仰、观点、态度的威胁或强化程度, 特别是同步行为来度量. Wang 等^[104] 根据信息传播过程中存在的谣言的产生、传递、判断和影响结果 4 个阶段, 抽象出了产生诱因、地点、时间、转评量、传播渠道和范围、传播生命周期等核心指标并进行了量化分级的定义, 并通过这些指标的度量反映出信息的传播对社会秩序、经济损失、政府公信力的影响. Lin^[16] 在对微博信息传播进行评价时, 提出了将信息传播广度、传播深度以及传播速

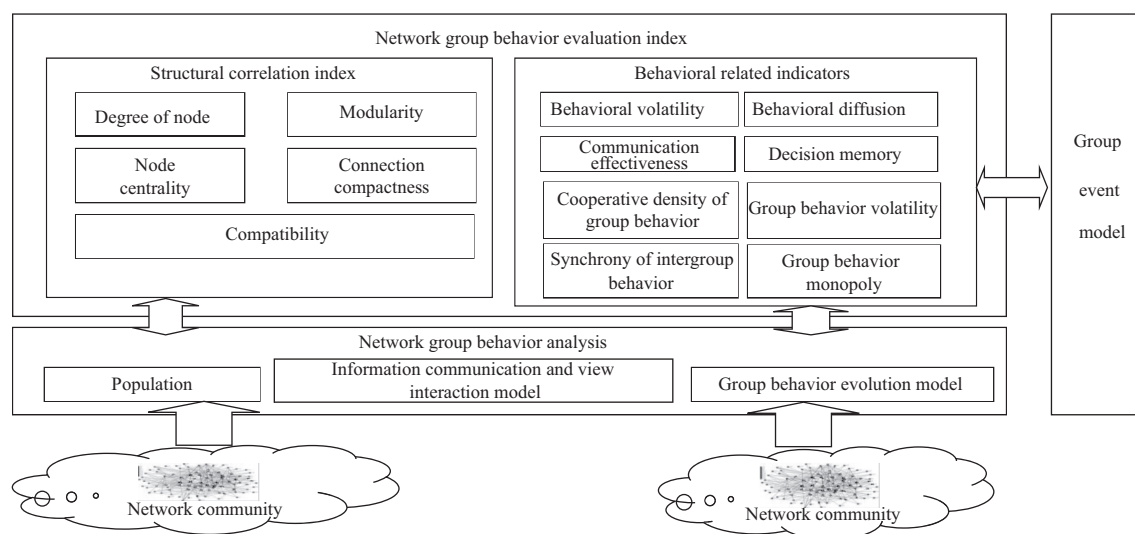


图 2 网络群体行为分析评价的模型框架

Figure 2 The analysis and evaluation framework for network group behavior

度相结合的评价模型, 并且将信息的真实性、可靠性、准确性和完整性作为信息内容质量的衡量标准. Metzger^[53] 则从信息内容的判别角度进行了研究, 提出了信息内容的精确性、权威性、目标性、实时性和覆盖率 5 个评价指标, 但是网络用户往往可能会忽略掉实时性, 甚至在时间和能力等实际情况约束下, 更关注于信息的精确性与覆盖率.

除了上述评估策略外, 在信息传播过程中的物理特征与控制特征也对评估效果产生影响, 如何将影响评估效果的各种参数和特征进行融合, 形成一个有效的舆情传播效果的评估体系, 则是下一步研究的关键. 本文在 PRCMA 信息传播模型的基础上, 对信息传播媒介、信息内容、传播者动机与影响力、接受者偏好等不同维度下评估指标与特征进行了一个初步的梳理, 如表 2 所示.

8 总结

网络的个性化发展使得用户的个性化行为与特征也得到更加充分的表达, 这也使得舆情信息的分析与传播无论是在理论上, 还是在工程实践上均存在一些急待解决的问题. 特别是在大数据的应用场景下^[105], 舆情信息的处理不仅需要融合复杂网络分析的关键技术, 同时也正在向跨媒介、跨媒体以及跨语言的方向快速发展与演化^[106]. 除了本文重点分析的跨媒介舆情传播的相关技术挑战外, 目前的舆情信息传播领域还存在着以下两个重大的研究挑战.

(1) 在跨媒体条件下, 舆情信息内容的跨媒体表达使得舆情事件信息可以通过文本、语音、图像、视频等不同类型的数据格式进行传播, 如何将舆情传播过程中的异构数据内容进行融合和处理, 并为不同类型的数据内容提供关联、融合以及统一检索则成为了目前最具挑战研究领域之一^[107, 108].

(2) 在跨语言条件下, 舆情传播过程中存在的多语言交叉传播, 如何实现跨语言之间的自动翻译和映射关联, 为监测更大范围内的舆情动态信息提供了技术支持, 并成为了目前舆情信息传播理论与技术研究的又一个重要的前沿^[109, 110].

本文针对舆情信息传播过程中存在的 5 个核心资源对象建立了 PRCMA 模型, 并针对该模型下

表 2 社会网络信息传播模型评估指标
Table 2 Information propaganda model evaluation index in social network

Criterion		Related definitions
The distribution characteristics of media	Information media	Communication node node interaction is strong, the power index will be higher, the node of heavy tailed phenomenon is more significant, which rely on information in interactive media communication is more strong in the “opinion leaders” is more significant.
	The community of propaganda nodes	The higher the clustering coefficient, the more obvious the characteristics of the network structure community, namely, the lower the speed of information transmission on the medium, but the higher the credibility of the information dissemination.
	Authority	The weaker interaction between media nodes, the more stringent information dissemination channels, the higher the authority, such as: government media
	Network propagation speed and delay	The speed and influence of information in different media: the propagation delay of information in the media network.
	Ability to interact on-line and offline	The degree of network user participation and the influence of network on offline activities.
Information content perspective	Accuracy	Information provides a true description; information provides a true result of subsequent events; information can accurately describe reality.
	Authority	Information source authority; information based on scientific discovery.
	Goal	The presentation of information presents a description of real goals; information provides an inseparable and unbiased description of reality.
	Real-time	The information is timely; the information provides multiple sides of the real-time state; the information remains sufficiently fresh.
	Coverage	Information covers a set of facts and perspectives; information can meet the needs of individuals or groups; information is beneficial to individuals or groups.
The spread of motivation and influence	Preference characteristics	Communicators (forwarding comments and browsing frequency); forwarding (comments and browsing) topic types; spread interest tags; user type propagation researchers and fans of user types.
	Emotional characteristics	Positive features: share happiness, enhance influence, reward motivation, help companies and help others motivation. Negative characteristics: negative emotional venting, psychological comfort, warning others, seeking compensation and revenge. Neutral characteristics: participation, interaction, support, community development, habitual motivation, etc.
	Reciprocal exchange capability	The purpose of sharing and disseminating information is to better access to equal information resources, to reflect the influence and value of communicators, and to value the equivalence of the content of communication.
	Influence characteristics	The physical attributes of network nodes, network topology attributes and network link characteristics.
Receiver preference	User perception feature	Semantic content category and emotional tag of information content.
	Characteristics of interest preferences	Receiver forwarding (comments and browsing), topic classification and interest tags; receiver focuses on user type and fan user type.
	Comment/Forwarding features	Forwarding (comments) content syntax and semantic features; comments and content information and negative comparison; whether or not the comment supports content.

每一个资源对象的研究过程中所存在的关键性技术问题与挑战进行了深入分析与综述. 其中, 在多媒体条件下, 不同媒介所拥有的用户网络拓扑结构具有小世界、无标度以及涌现等特征, 并在跨媒介信

息交叉传播过程中,通过相互影响使得舆情信息的传播动力学机制更为复杂,并成为了目前与跨媒体、跨语言并列的 3 个最具挑战性的研究任务之一.此外,本文针对传播者影响力的识别与度量,以及对信息传播影响作用问题、不可信信息的识别以及可信度的实时度量问题、信息接收者个体兴趣偏好与群体行为对信息接收的影响程度的度量问题以及跨媒介信息传播效果的评价等问题与技术挑战进行了理论探讨、实际应用以及发展趋势的分析,为解决跨媒介舆情信息的传播机制提供了新的研究线索与方向.相信在未来的研究中,通过与大数据技术和复杂网络技术的深度融合,必然会对舆情分析领域产生重大且深远的影响.

参考文献

- 1 Borchers C. Here's why Trump's attacks on 'fake news' succeed. Washington Post. 2017.02. http://www.washingtonpost.com/news/the-fix/wp/2017/02/17/heres-why-trumps-attacks-on-fake-news-succeed/?utm_term=.91a3d6e3c8fa
- 2 CNN Reports. Did fake news help elect donald trump?. 2016.11.19. <http://money.cnn.com/2016/11/19/technology/mark-zuckerberg-facebook-fake-news-election/index.html>
- 3 Bessi A, Coletto M, Davidescu G A, et al. Science vs. conspiracy: collective narratives in the age of misinformation. PLOS One, 2015, 10: 1–17
- 4 Zhou D H, Han W B, Wang Y J. Social network information dissemination model based on node and information characteristics. Comput Res Dev, 2015, 52: 156–166 [周东浩, 韩文报, 王勇军. 基于节点和信息特征的社会网络信息传播模型. 计算机研究与发展, 2015, 52: 156–166]
- 5 Lazarsfeld P, Berelson B, Gaudet H. The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign. New York: ACM Press, 1968
- 6 Schifferes S. Technological transformations of news: a long term perspective. In: Proceedings of the 25th International Conference Companion on World Wide Web, Montréal, 2016. 743–744
- 7 Li B. Portrait sketch of micro-blog opinion leaders — an example of opinion leaders in 40 micro-blog events. Journalism Rev, 2012, 1: 19–25 [李彪. 微博意见领袖群体“肖像素描”——以 40 个微博事件中的意见领袖为例. 新闻记者, 2012, 1: 19–25]
- 8 Wang S H, Huang Q M. Research progress of heterogeneous media analysis technology. Integrat Tech, 2015, 4: 7–21 [王树徽, 黄庆明. 异质媒体分析技术研究进展. 集成技术, 2015, 4: 7–21]
- 9 Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks. Nature, 1998, 393: 440–442
- 10 Barabasi A L, Albert R. Emergence of scaling in random networks. Science, 1999, 286: 509–512
- 11 Du H F, Li S Z, Marcus W F, et al. Community structure of small world networks and scale-free networks. Acta Phys Sin, 2007, 56: 6886–6893 [杜海峰, 李树苗, Marcus W F, 等. 小世界网络与无标度网络的社区结构研究. 物理学报, 2007, 56: 6886–6893]
- 12 Zhang L. Research on some algorithms in network public opinion propagation. Dissertation for Ph.D. Degree. Beijing: Beijing Jiaotong University, 2009 [张立. 网络舆论传播中若干算法的研究. 博士学位论文. 北京: 北京交通大学, 2009]
- 13 Hu H B, Wang L. A brief history of power law distribution. Physical, 2005, 34: 889–896 [胡海波, 王林. 幂律分布研究简史. 物理, 2005, 34: 889–896]
- 14 Kou Z B, Zhang C S. Reply networks on a bulletin board system. Phys Rev E, 2003, 67: 328–335
- 15 Li W J, Bao H Y. Analysis of characteristics of information dissemination model in BBS. Comput Eng App, 2010, 46: 18–22 [李维杰, 包红云. BBS 中信息传播模式的特征分析. 计算机工程与应用, 2010, 46: 18–22]
- 16 Lin S. Research on evaluation model of micro-blog individual information communication influence. Modn Lib Inf Tech, 2014, 243: 79–85 [林琛. 微博个体信息传播影响力评价模型研究. 现代图书情报技术, 2014, 243: 79–85]
- 17 Weng J, Lim E P, Jiang J, et al. Twitterank: finding topic sensitive influential twitters. In: Proceedings of the 3rd ACM International Conference on Web Search and Data Mining (WSDM 2010), New York, 2010. 261–270
- 18 Pastorsatorras R, Vespignani A. Epidemic spreading in scale-free networks. Phys Rev Lett, 2001, 86: 3200–3203
- 19 Zhang F, Si G Y, Luo P. A survey of rumor propagation model. Complex Syst Complex Sci, 2009, 6: 1–11 [张芳, 司光亚, 罗批. 谣言传播模型研究综述. 复杂系统与复杂性科学, 2009, 6: 1–11]

- 20 Wu X Y, Liu Z H. How community structure influences epidemic spread in social network. *Phys A Stat Mech Appl*, 2008, 387: 623–630
- 21 Liu Z H. Signal detection and transmission in complex networks (in Chinese). *Sci Sin-Phys Mech Astron*, 2014, 44: 1334–1343 [刘宗华. 基于复杂网络的信号检测与传递. *中国科学: 物理学 力学 天文学*, 2014, 44: 1334–1343]
- 22 Yang J, Counts S. Predicting the speed, scale and rang of information diffusion in Twitter. In: *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, Menlo Park, 2010. 355–358
- 23 Lagnier C, Denoyer L, Gaussier E, et al. Predicting information diffusion in social networks using content and user's profiles. In: *Proceedings of the 35th European Conference on Advance in Information Retrieval*, Moscow, 2013. 74–85
- 24 Zhang Y W, Qi J Y, Ma J, et al. The interaction mechanism of online public opinion and unconventional emergencies — based on system dynamics modeling analysis. *Intell Mag*, 2010, 29: 1–6 [张一文, 齐佳音, 马君, 等. 网络舆情与非常规突发事件作用机制 —— 基于系统动力学建模分析. *情报杂志*, 2010, 29: 1–6]
- 25 Di G Q, Zeng H Y, Le Z J, et al. System dynamics model and simulation of network public opinion events. *Intell Mag*, 2012, 31: 12–19 [狄国强, 曾华艺, 勒中坚, 等. 网络舆情事件的系统动力学模型与仿真. *情报杂志*, 2012, 31: 12–19]
- 26 Spiro E S, Irvine C A, Dubois C L, et al. Waiting for a retweet: modeling waiting time in information propagation. In: *Proceedings of Workshop on Social Network and Social Media Analysis: Methods, Models and Applications*, Lake Tahoe, 2013
- 27 Dredze M, Kambadur P, Kazantsev G, et al. How Twitter is changing the nature of financial news discovery. In: *Proceedings of the 2nd International Workshop on Data Science for Macro-Modeling*, San Francisco, 2016
- 28 Leskovec J, Backstrom L, Kleinberg J. Meme-tracking and the dynamics of the news cycle. In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, Paris, 2009
- 29 Heimbach I, Hinz O, Schiller B, et al. Content virality on online social networks: empirical evidence from Twitter, Facebook, and Google+ on german news websites. In: *Proceedings of the 26th ACM International Conference on Hypertext & Social Media*, Guzelyurt, 2015. 39–47
- 30 Yu K, Rong L L, Guo W Q, et al. Research on public opinion communication model based on online and online network. *Manag Rev*, 2015, 27: 200–212 [于凯, 荣莉莉, 郭文强, 等. 基于线上线上网络的舆情传播模型研究. *管理评论*, 2015, 27: 200–212]
- 31 Du R, Yu Z W, Liu Z L, et al. Research on social impact of online and offline activities based on watercress city. *Chin J Comput*, 2014, 37: 238–245 [杜蓉, 於志文, 刘振鲁, 等. 基于豆瓣同城活动的线上线下社交影响研究. *计算机学报*, 2014, 37: 238–245]
- 32 Bauch C T, Galvani A P. Social factors in epidemiology. *Science*, 2013, 342: 47–49
- 33 Howell L. Digital wildfires in a hyperconnected world. *World Economic Forum Report*, 2013. <http://reports.weforum.org/global-risks-2013/risk-case-1/digital-wildfires-in-a-hyperconnected-world/>
- 34 Shen Y, Yu J, Dong K, et al. Automatic fake followers detection in chinese micro-blogging system. In: *Proceedings of Advances in Knowledge Discovery and Data Mining*. Berlin: Springer, 2014. 596–607
- 35 Hardalov M, Koychev I, Nakov P. In search of credible news. In: *Proceedings of International Conference on Artificial Intelligence: Methodology, Systems, and Applications*. Berlin: Springer, 2016. 172–180
- 36 Gupta A, Kumaraguru P, Castillo C, et al. TweetCred: real time credibility assessment of content on Twitter. In: *Proceedings of International Conference on Social Informatics*. Berlin: Springer, 2014. 228–243
- 37 Papadopoulos S, Bontcheva K, Jaho E, et al. Overview of the special issue on trust and veracity of information in social media. *ACM Trans Inf Syst*, 2016, 34: 1–5
- 38 Frakes W B. Information and misinformation: an investigation of the notions of information, misinformation, informing, and misinforming. *Soc Sci Inf Stud*, 1985, 5: 147–149
- 39 Hamidian S, Diab M T. Rumor identification and belief investigation on Twitter. In: *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, San Diego, 2016. 3–8
- 40 Ratkiewicz J, Conover M, Meiss M, et al. Detecting and tracking the spread of astroturf memes in microblog streams. In: *Proceedings of the 20th International Conference Companion on World Wide Web*, Hyderabad, 2011. 249–252
- 41 Qazvinian V, Rosengren E, Radev D R, et al. Rumor has it: identifying misinformation in microblogs. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, 2011. 1589–1599
- 42 Karlova N A, Lee J H. Notes from the underground city of disinformation: a conceptual investigation. *Proc Assoc Inf Sci Technol*, 2011, 48: 1–9

- 43 Takahashi T, Igata N. Rumor detection on Twitter. In: Proceedings of the 6th International Conference on Soft Computing and Intelligent Systems, Kobe, 2012. 452–457
- 44 Wang A H. Don't follow me: spam detection in Twitter. In: Proceedings of the 2010 International Conference on Security and Cryptography (SECRYPT), Athens, 2010. 1–10
- 45 Kumar K P K, Geethakumari G. Detecting misinformation in online social networks using cognitive psychology. *Human-Centric Comput Inf Sci*, 2014, 4: 14–36
- 46 Karlova N A, Fisher K E. A social diffusion model of misinformation and disinformation for understanding human information behaviour. *Inform Res*, 2013, 18: 1–17
- 47 Kumar S, West R, Leskovec J. Disinformation on the web: impact, characteristics, and detection of wikipedia hoaxes. In: Proceedings of the 25th International Conference on World Wide Web, Montréal, 2016. 591–602
- 48 Zhang H L, Alim M A, Li X, et al. Misinformation in online social networks: detect them all with a limited budget. *ACM Trans Inf Syst*, 2016, 34: 1–24
- 49 Abdulla R, Garrison B, Salwen M, et al. The credibility of newspapers, television news, and online news. In: Proceedings of the Association for Education in Journalism and Mass Communication Annual Convention, Miami, 2002. 97–106
- 50 Li L, Sun J H. Group characteristics based on social network conflict information dissemination. *Syst Eng Theory Prac*, 2014, 34: 207–214 [李林, 孙军华. 基于社会网络冲突信息传播的群体特征. *系统工程理论与实践*, 2014, 34: 207–214]
- 51 Schwarz J, Ringel M M. Augmenting web pages and search results to support credibility assessment. In: Proceedings of the CHI 2011 Conference on Human Factors in Computing Systems, Vancouver, 2011. 1245–1254
- 52 Odonovan J, Kang B, Meyer G, et al. Credibility in context: an analysis of feature distributions in Twitter. In: Proceedings of International Conference on Privacy, Security, Risk and Trust, Amsterdam, 2012. 293–301
- 53 Metzger M J, Flanagin A J. Credibility and trust of information in online environments: the use of cognitive heuristics. *J Pragmatics*, 2013, 59: 210–220
- 54 Gill A J, Nowson S, Oberlander J. What are they blogging about? personality, topic and motivation in blogs. In: Proceedings of the 3rd International ICWSM Conference, San Jose, 2009
- 55 Xu J, Yang X P, Liu Z. Research on web information credibility verification method based on content trust. *J Beijing Ins of Tech*, 2014, 34: 710–715 [徐静, 杨小平, 柳增. 基于内容信任的 Web 信息可信度验证方法研究. *北京理工大学学报*, 2014, 34: 710–715]
- 56 Fang B X, Guo Y C, Zhou Y. ICCON control model and evaluation of internet information content security. *Sci China Ser F-Inf Sci*, 2009, 39: 951–965 [方滨兴, 郭云川, 周渊. 互联网信息内容安全的 ICCON 控制模型及评价. *中国科学 F 辑: 信息科学*, 2009, 39: 951–965]
- 57 Castillo C, Mendoza M, Poblete B. Predicting information credibility in time-sensitive social media. *Internet Res*, 2013, 23: 560–588
- 58 Zubiaga A, Hoi G W S, Liakata M, et al. Analyzing how people orient to and spread rumors in social media by looking at conversational threads. *PLOS One*, 2016, 11: e0150989
- 59 Zubiaga A, Ji H. Tweet, but verify: epistemic study of information verification on Twitter. *Soc Netw Anal Min*, 2014, 4: 1–12
- 60 Ratkiewicz J, Conover M, Meiss M, et al. Truthy: mapping the spread of astroturf in microblog streams. In: Proceedings of the 20th International Conference Companion on World Wide Web, Hyderabad, 2011. 249–252
- 61 Castillo C, Mendoza M, Poblete B. Information credibility on Twitter. In: Proceedings of the 20th International Conference on World Wide Web, Hyderabad, 2011. 675–684
- 62 Xia X, Yang X, Wu C, et al. Information credibility on Twitter in emergency situation. In: Proceedings of Pacific Asia Conference on Intelligence and Security Informatics, Kuala Lumpur, 2012. 45–59
- 63 Zhao B, Rubinstein B I, Gemmell J, et al. A Bayesian approach to discovering truth from conflicting sources for data integration. *Proc VLDB Endowment*, 2012, 5: 550–561
- 64 Liu W J. Research on big data method and rumor. *Folklore Stud*, 2016, 127: 85–91 [刘文江. 大数据方法与谣言研究. *民俗研究*, 2016, 127: 85–91]
- 65 Liu J G, Ren S M, Guo Q, et al. Research progress of node importance ordering in complex networks. *Acta Phys Sin*, 2013, 62: 178901 [刘建国, 任单明, 郭强, 等. 复杂网络中节点重要性排序的研究进展. *物理学报*, 2013, 62: 178901]

- 66 Granovetter M S. The strength of weak ties. *Am J Soc*, 1973, 78: 1360–1380
- 67 Qian Y X. An empirical study on motivation of internet word of mouth in China's virtual brand community. Dissertation for Master Degree. Hefei: University of Science and Technology China, 2014 [钱玉霞. 国内虚拟品牌社区中网络口碑传播动机实证研究. 硕士学位论文. 合肥: 中国科学技术大学, 2014]
- 68 Rao Y, Feng N. Data Analysis — Forum Oriented Network Content Analysis Report (2015). Beijing: Electronic Industry Press, 2015 [饶元, 冯妮. 数据分析 —— 面向论坛的网络内容分析报告 (2015). 北京: 电子工业出版社, 2015]
- 69 Linderholm O. 10 people you won't believe have fake followers on Twitter. 2013. <https://www.yahoo.com/news/blogs/profit-minded/10-people-won-t-believe-fake-followers-twitter-215539518.html?ref=gs>
- 70 Chen D B, Lü J Y, Shang M S, et al. Identifying influential nodes in complex networks. *Phys A Stat Mech Appl*, 2012, 391: 1777–1787
- 71 Centola D. The spread of behavior in an online social network experiment. *Science*, 2010, 329: 1194–1197
- 72 Ugander J, Backstrom L, Marlow C, et al. Structural diversity in social contagion. *Proc Natl Acad Sci USA*, 2012, 109: 5962–5966
- 73 Kitsak M, Gallos L K, Havlin S, et al. Identification of influential spreaders in complex networks. *Nat Phys*, 2010, 6: 888–893
- 74 Zeng A, Zhang C J. Ranking spreaders by decomposing complex networks. *Phys Lett A*, 2012, 377: 1031–1035
- 75 Fang B X, Jia Y, Han Y. Social networking analysis, core scientific issues, current research and future perspectives. *Bull Chin Acad Sci*, 2015, 30: 187–199 [方滨兴, 贾焰, 韩毅. 社交网络分析核心科学问题、研究现状及未来展望. 中国科学院院刊, 2015, 30: 187–199]
- 76 Ren Z M, Shao F, Liu J G, et al. Research on network node importance measurement method based on degree and aggregation coefficient. *Acta Phys Sin*, 2013, 62: 128901 [任卓明, 邵凤, 刘建国, 等. 基于度与集聚系数的网络节点重要性度量方法研究. 物理学报, 2013, 62: 128901]
- 77 Tan Y J, Wu J, Deng H Z, et al. Review of invulnerability research on complex networks. *Syst Eng*, 2006, 24: 1–5 [谭跃进, 吴俊, 邓宏钟, 等. 复杂网络抗毁性研究综述. 系统工程, 2006, 24: 1–5]
- 78 Li P X, Ren Y Q, Xi Y M. A measure of the importance of a network node (set). *Syst Eng*, 2004, 22: 13–20 [李鹏翔, 任玉晴, 席西民. 网络节点 (集) 重要性的一种度量指标. 系统工程, 2004, 22: 13–20]
- 79 Yu X, Li Y H, Zheng X P, et al. Evaluation method of node importance in communication networks based on network performance gradient. *J Tsinghua Univ (NAT SCI EDIT)*, 2008, 48: 541–544 [余新, 李艳和, 郑小平, 等. 基于网络性能变化梯度的通信网络节点重要程度评价方法. 清华大学学报 (自然科学版), 2008, 48: 541–544]
- 80 Huang X, Vodenska I, Wang F, et al. Identifying influential directors in the United States corporate governance network. *Phys Rev E Stat Nonlinear Soft Matter Phys*, 2011, 84: 046101
- 81 Goyal A, Bonchi F, Lakshmanan L V. Learning influence probabilities in social networks. In: *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining*, New York, 2010. 241–250
- 82 Zhang Y L, Li C P, Chen H. An efficient link based similarity measure in information networks. *J Softw*, 2014, 25: 2602–2615 [张应龙, 李翠平, 陈红. 信息网络中一个有效的基于链接的结点相似度量. 软件学报, 2014, 25: 2602–2615]
- 83 Song J, Guo C P, Zhang Y C, et al. Research and implementation of incremental iterative computation model. *Chin J Comput*, 2016, 39: 109–125 [宋杰, 郭朝鹏, 张一川, 等. 增量式迭代计算模型研究与实现. 计算机学报, 2016, 39: 109–125]
- 84 Xu W T, Liu F, Zhu E Z. Research on new micro-blog user influence ranking algorithm based on MapReduce. *Comput Sci*, 2016, 43: 66–70 [徐文涛, 刘锋, 朱二周. 基于 MapReduce 的新型微博用户影响力排名算法研究. 计算机科学, 2016, 43: 66–70]
- 85 Qi X M, Sun W X. A PageRank algorithm for fusing multiple feature factors. *Comput Eng Appl*, 2015, 53: 97–103 [齐向明, 孙文心. 一种多特征因子融合的 PageRank 算法研究. 计算机工程与应用, 2015, 53: 97–103]
- 86 Lü L, Zhang Y C, Chi H Y, et al. Leaders in social networks, the delicious case. *PLOS One*, 2011, 6: e21202
- 87 Li Q, Zhou T, Lü L, et al. Identifying influential spreaders by weighted LeaderRank. *Phys A Stat Mech Appl*, 2014, 404: 47–55
- 88 Xu J M, Zhu F X, Liu S C, et al. Improved LeaderRank algorithm for opinion leader mining. *Comput Eng Appl*, 2015, 51: 110–114 [徐郡明, 朱福喜, 刘世超, 等. 改进 LeaderRank 算法的意见领袖挖掘. 计算机工程与应用, 2015,

- 51: 110–114]
- 89 Guo C, Shen Q, Yang Y, et al. User rank: a user influence-based data distribution optimization method for privacy protection in cloud storage system. In: Proceedings of the 39th Annual Computer Software and Applications Conference, Taichung, 2015. 104–109
 - 90 Huang J M, Shen H W, Chen X Q. Explore the spread of information using the backbone of social networking. J Chin Inf Process, 2016, 30: 74–82 [黄俊铭, 沈华伟, 程学旗. 利用社交网络的影响力骨架探索信息传播. 中文信息学报, 2016, 30: 74–82]
 - 91 Wang C X, Guan X H, Qin T, et al. Research on influence modeling of opinion leaders in micro-blog news communication. J Softw, 2015, 26: 1473–1485 [王晨旭, 管晓宏, 秦涛, 等. 微博消息传播中意见领袖影响力建模研究. 软件学报, 2015, 26: 1473–1485]
 - 92 Bakshy E, Hofman J M, Watts D J, et al. Identifying “influencers” on Twitter. In: Proceeding of the 4th International Conference on Web Search and Data Mining. New York: ACM Press, 2011. 1–10
 - 93 Wang Z J, Wang S H, Zhang W G, et al. Potential influence propagation model based on social content. Chin J Comput, 2016, 39: 1528–1540 [王祯骏, 王树徽, 张维刚, 等. 基于社交内容的潜在影响力传播模型. 计算机学报, 2016, 39: 1528–1540]
 - 94 Zhao Z Y, Yu H, Wang X F, et al. Influence analysis of node propagation based on network community structure. Chin J Comput, 2014, 37: 753–766 [赵之滢, 于海, 汪小帆, 等. 基于网络社团结构的节点传播影响力分析. 计算机学报, 2014, 37: 753–766]
 - 95 Zhao L J, Xie W L, Gao H O, et al. A rumor spreading model with variable forgetting rate. Phys A: Stat Mech Appl, 2013, 392: 6146–6154
 - 96 Wang X C, Lin Z Y, Sun J. 10 judgments about the regularity of group events — based on the analysis of participants’ behavioral characteristics. J Nation Sch Admin, 2011, 1: 13–16 [王新超, 林泽炎, Sun J. 关于群体性事件发生规律的 10 个判断 —— 基于参与者行为特征的分析. 国家行政学院学报, 2011, 1: 13–16]
 - 97 Zhu G H, Jiang G P, Xia L L. The influence of bandwagon phenomenon on rumor propagation in social networks. Comput Sci, 2016, 43: 135–139 [朱冠桦, 蒋国平, 夏玲玲. 社交网络上从众现象对谣言传播影响的研究. 计算机科学, 2016, 43: 135–139]
 - 98 Asur S, Huberman B A, Szabo G, et al. Trends in social media: persistence and decay. In: Proceedings of the 5th International AAAI Conference on Weblogs and Social Media. Menlo Park: The AAAI Press, 2011. 434–437
 - 99 Berger J, Milkman K L. What makes online content viral? J Marketing Res, 2012, 49: 192–205
 - 100 Yuan Q, Gao Q. The analysis of online news information credibility assessment on weibo based on analyzing content. In: Proceedings of International Conference on Engineering Psychology and Cognitive Ergonomics. Berlin: Springer, 2016. 125–135
 - 101 Wang Y Z, Yu J Y, Qiu W, et al. Evolutionary game model and analysis method of network group behavior. Chin J Comput, 2015, 38: 282–300 [王元卓, 于建业, 邱雯, 等. 网络群体行为的演化博弈模型与分析方法. 计算机学报, 2015, 38: 282–300]
 - 102 Zhang Y F, Xiao R B. The emergence mechanism of network group event public opinion synchronization based on cellular automata. Syst Eng Theor Pract, 2014, 34: 2600–2608 [张耀峰, 肖人彬. 基于元胞自动机的网络群体事件舆论同步的涌现机制. 系统工程理论与实践, 2014, 34: 2600–2608]
 - 103 Niu P Y. Research on performance evaluation of network information dissemination. Dissertation for Ph.D. Degree. Wuhan: Wuhan University, 2009 [牛培源. 网络信息传播绩效评估研究. 博士毕业论文. 武汉: 武汉大学, 2009]
 - 104 Wang J Y, Zhang P, Liu L W, et al. Research on traffic evaluation of network rumor propagation. Intell Mag, 2016, 35: 105–109 [王靖元, 张鹏, 刘立文, 等. 网络谣言传播交通评价研究. 情报杂志, 2016, 35: 105–109]
 - 105 Meng X F, Du Z J. Big data integration research: problems and challenges. Comput Res Dev, 2016, 53: 231–246 [孟小峰, 杜治娟. 大数据整合研究: 问题与挑战. 计算机研究与发展, 2016, 53: 231–246]
 - 106 Gao J, Buldyrev S V, Stanley H E, et al. Networks formed from interdependent networks. Nat Phys, 2012, 8: 40–48
 - 107 Chen C, Chen F, Cao D, et al. A cross-media sentiment analytics platform for microblog. In: Proceedings of the 23rd ACM International Conference on Multimedia, Brisbane, 2015. 767–769
 - 108 Cao D, Ji R, Lin D, et al. A cross-media public sentiment analysis system for microblog. Multimed Syst, 2016, 22: 479–486
 - 109 Zhou H W, Chen L, Shi F L, et al. Learning bilingual sentiment word embeddings for cross-language sentiment

- classification. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics, Beijing, 2015. 430–440
- 110 Zhou G, Zeng Z, Huang J X, et al. Transfer learning for cross-lingual sentiment classification with weakly shared deep neural networks. In: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, Pisa, 2016. 245–254

A survey of information propaganda mechanism under the cross-medium

Yuan RAO*, Lianwei WU & Junyi ZHANG

Laboratory of Social Intelligence and Complex Data Processing, School of software, Xi'an Jiaotong University, Xi'an 710049, China

* Corresponding author. E-mail: yuanrao@163.com

Abstract With the development of mobile networking, cloud computing, and big data technology, propaganda through public opinion and sentiment on the Internet has shown some new features, such as cross-medium, cross-media, and cross-language characteristics, which indicates exciting challenges for next-generation research and application of public opinion analysis. In this paper, the information propaganda mechanism of public sentiment under social networking is explored under a cross-medium environment, which means that the information can be spread across different web sites and social networks. A multi-element information propaganda model, called the PRCMA model, is then proposed, which includes five critical elements during the entire process of information propaganda, such as the information publisher (P), information receiver (R), content (C), medium (M), and result assessment (A). Furthermore, the critical challenges and vital research problems are illustrated, and some up-to-date research results are analyzed and reviewed based on each dimension of the PRCMA model, particularly on how to identify and measure the influence of nodes, identify fake information in real time, measure the personal preference and group behaviors, and measure the results of propaganda in the process of information spread under a cross-medium environment. Finally, some new technologies, theoretical methods, applications, and future trends are discussed, which provide new clues and research fields for further research on public opinion and sentiment on the Internet under a cross-medium environment.

Keywords cross-media, sentiment network, information propaganda, big data, assessment mechanism



gence and complex data processing.

Yuan RAO is an associate professor and Ph.D. director at the Software School of Xi'an Jiaotong University. He obtained his Ph.D. degree from Xi'an Jiaotong University in 2005. He was a postdoctoral fellow at Tsinghua University from 2005 to 2007 and a visiting scholar at Carnegie Mellon University (CMU), USA from 2015 to 2016. His research interests include social intelligence and complex data processing.



Lianwei WU was born in August 1992. He is currently working toward his Ph.D. degree at the Laboratory of Social Intelligence and Complex Data Processing. His current research interests include information detection, information credibility evaluation, and natural language processing.



Junyi ZHANG was born in 1983. He received his BS degree from North East Normal University in 2006, and his MS degree from Xi'an Micro Electronic Institute in 2009. He is currently working toward his Ph.D. degree at the Software Academy of Xi'an JiaoTong University. As a Ph.D. candidate, his research interests include information detection and social intelligence.