社交媒体上虚假信息的识别与控制
——基于 Web of Science 研究文献的分析

迟妍玮，张增一
（中国科学院大学人文学院，北京 100049）

摘 要：社交媒体上虚假信息泛滥导致越发严重的负面影响，引起公众担忧和恐慌，如何高效、精准地识别虚假信息并控制其传播和影响受到国际学术界的关注。本文基于 Web of Science 数据库，从识别与控制虚假信息传播存在的困难、社交媒体平台所发挥的作用、纠正虚假信息的策略、基于计算机技术手段的检测和识别、基于心理学的检测和识别以及阻止和控制虚假信息的策略等方面，对国外权威期刊近几年发表的相关研究论文进行了分析，揭示了国际学术界关于虚假信息的识别与控制的研究现状、热点和趋势，希望为我国在识别与控制虚假信息传播的理论研究和政策实践方面提供参考。

关键词：虚假信息；社交媒体；信息传播
中图分类号：F49；G202；C939 **文献标识码**：A DOI：10.3772/j.issn.1009-8623.2020.10.008

如今，社交媒体已成为人们获得信息、沟通交流的重要渠道，极大地促进了社会经济发展，影响了人类社会生活的各个方面。然而，应对和治理社交媒体上虚假信息的泛滥，已成为世界性难题[1]，引起国际学术界的普遍关注。相比于真实的信息，虚假信息传播得更快[2]，造成广泛的负面影响[3]。社交媒体在放大了其影响[4]的同时，为骗局的大规模实施提供了便利[5]，也成为有关健康问题的虚假信息的主要传播渠道[6]。公众对有关政治、科学等方面的虚假信息的担忧与日俱增[7]。因此，从理论和实践上对虚假信息进行有效的自动识别，并控制其基于社会关系网络[8]的病毒式[9]的传播是必要的，也是亟待解决的问题。

本文关注的是在社交媒体上被有意或无意的传播[10]的虚假信息，无论传播者是否有误导的动机[11]。那些虚假信息可能内容不完整、不准确、混淆、过时、存在偏见、缺乏科学证据等[12]。特别关注那些有关虚假信息的识别和控制的研究，试图梳理这一方向的研究概况、研究热点以及未来趋势。

1 国际上有关虚假信息识别和控制的研究文献的基本情况

在 Web of Science 数据库中，以 misinformation 为关键词进行检索，截至 2020 年 7 月 15 日，共有标题中含有关键词 misinformation 的文章 1 125 篇。本研究主要关注那些有关虚假信息的识别与控制的研究，在上述结果中检索有关识别和控制的研究文献，关键词包括 testing、detect、check、distinguish、control、identify 等，去掉重复文献、专利文章、编者按等不相关文献后，共 339 篇。

有关虚假信息的识别和控制受到多个学科领域的关注，从发表渠道来看，涉及 181 本期刊杂志，以及美国电气和电子工程师协会（IEEE）、社交

第一作者简介：迟妍玮（1988—），女，在读博士研究生，主要研究方向为科学传播。
通讯作者简介：张增一（1963—），男，教授，主要研究方向为科学传播、科技与社会。邮箱：zhzy@ucas.ac.cn
收稿日期：2020-08-19

计算与社交媒体国际会议（SCSM）、国际万维网大会（WWW）等国际会议。这些期刊分布于心理学、传播学、医学、计算机科学等多个学科，暂无有关虚假信息研究的专门期刊。

从研究对象来看，结合文献标题（见图1）和关键词（见图2）的词云图，研究多涉及虚假信息的影响效果、对虚假信息的记忆研究、目击者研究，假新闻、伪造信息及虚假信息来源、引申受到关注，研究素材多来自社交媒体、媒体及重大事件，如何发现、减少、纠正虚假信息被关注。

从学科分布来看，跨学科研究较多，38.64%（131篇）的文献涉及两个及以上研究方向。这些文献共涉及88个学科研究方向，其中心理学数量最多，其后依次为计算机科学、行为科学、传播学、公共环境和职业健康、神经科学与神经学、政府与法律、普通内科、工程学、信息科学与图书馆学等。各学科发表相关论文的数量和时间跨度存在差异，具体的学科发表文章数量、时间跨度和主要研究内容可见表1。第一，上述学科近两年均有发表有关虚假信息的文章，且涉及虚假信息的识别与控制；第二，计算机科学和工程学的论文的时间跨度较短，均为10年内开始发表的相关论文；第三，传播学的论文的时间跨度最长，最早可追溯到1957年，但除此之外其他论文均为2008年之后发表；第四，其他7个学科，均为20世纪80年代至90年代初期开始发表相关论文，结合总体文章年代分布，可推测多数学科在该时间段开始关注虚假信息的识别和控制研究并陆续发表文章。结合文章的关键词和摘要，发现不同学科的研究内容存在差异，第一，心理学、行为科学、神经科学与神经学偏重对人的心理、行为、大脑活动等的实验研究，特别是与记忆相关的研究；第二，计算机科学、工程学侧重于对基于程序算法的自动化监控的探索；第三，传播学侧重实证性研究，研究对象多为当下热点；第四，政府与法律关注党派与选举中的虚假信息；第五，普通内科、公共环境和职业健康更关注与公共卫生健康相关的虚假消息研究。

对上述学科近年来发表与虚假信息相关的论文数量进行分析发现：第一，计算机科学、工程学和传播学学科近5年发表有关虚假信息的论文数量明显增多；第二，行为科学和普通内科则近5年发表的有关虚假信息方面的论文数量明显减少；第三，心理学、神经科学与神经学、政府与法律学、信息科学与图书馆学、公共环境和职业健康学等持续关注虚假信息相关议题，其中心理学领域发表的相关论文在近5年持续增多，政府与法律学、公共环境和职业健康学则在2020年明显增加，其余学科发表相关论文数量变化不明显。

总而言之，越来越多的学科和研究关注到虚假信息的识别和控制，但各学科的关注点和发文量不同。在20世纪80年代之后许多学科开始关注并陆续发表相关研究成果，计算机相关学科在近10年来重视虚假信息的识别和控制问题，发表相关论文的数量不断增多。近年来，媒体、网络、社交媒体上的虚假信息研究，公共卫生相关的虚假消息研究，以及基于计算机技术手段对虚假信息的研究、检测、判断和控制等研究方向受到普遍关注。

2　识别和控制虚假信息传播的困境

公众识别虚假信息是阻止虚假信息广泛传播的重要手段之一，但由于多方面的因素，仅依靠公众识别停止传播虚假信息存在难度。

第一，被误导。公众面对社交媒体上的信息是容易被误导的。故意散布者会进行伪装，且内容和传播的过程可能被操纵[13]。公众媒体素养低被认为是被误导的原因之一，即在新媒体环境中获取、分析、评估和以各种形式创造信息的能力低[14]。公众被虚假信息误导，不仅关乎个体识别错误、谎言的功能，也关乎群体和社会因素[15]。

第二，不知情。虚假信息可能是被掺入了一些半真半假的信息内容创造出的难以辨认的混合信息[16]，人们可能在不知情的情况下成为传播虚假信息的角色[17]，或是不加区别地在社交媒体上分享虚假信息[18]。

第三，错误的认知。网络上的虚假信息影响信息本质及用户对信息的认知[19]，纠正虚假信息并不能纠正人们对虚假信息的认知[20]。虽然强化自我肯定能够让人们对外部来源的虚假信息产生抵抗力[21]，但人们的认知来自个人对世界的体验，需要从外部获取信息，以从社交媒体关

图 1　文献标题词云图

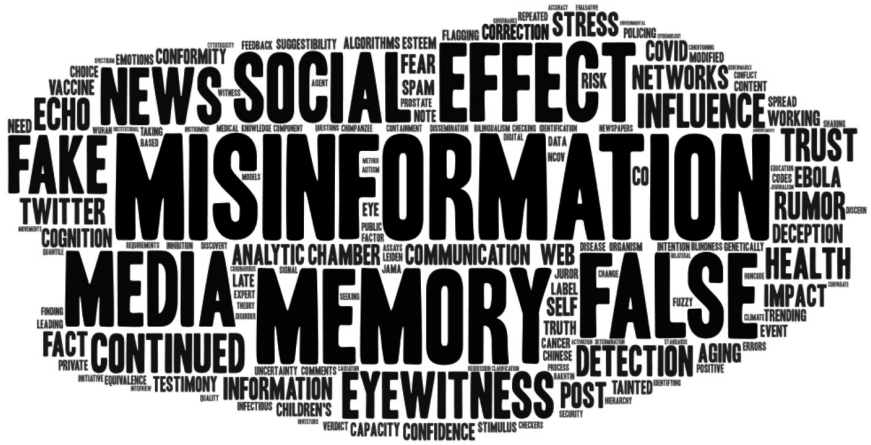

图 2　文献关键词词云图

表 1　排名前 10 的学科分布、时间跨度和主要研究内容

序号	学科	文章数量	占比（%）	文章时间跨度	主要研究内容
1	心理学	152	44.84	1984—2020	事件后信息的误导效应，目击者研究相关，记忆相关研究，虚假记忆的制造和判断等
2	计算机科学	46	13.57	2012—2020	利用模型、算法来识别、检测、控制虚假信息传播，判断信息可信度，研究对象包括社交媒体、网络、报纸等
3	行为科学	29	8.55	1986—2020	虚假记忆信息的影响因素，识别测试，信息对人类的判断、决策的影响等
4	传播学	27	7.96	1957—2020	社交媒体、网络上的虚假信息研究，可信度、持续影响、态度改变等研究
5	神经科学与神经学	23	6.78	1991—2019	对虚假记忆、隐藏信息、回忆、欺骗误导等的研究，对眼球、大脑反应等的测试

续表

序号	学科	文章数量	占比（%）	文章时间跨度	主要研究内容
6	政府与法律	21	6.19	1992—2020	假新闻、党派偏见、持续营销效果、误导暗示、媒体信息曝光等
7	公众、环境和职业健康	16	4.72	1992—2019	公共卫生（病毒、疫苗、烟草标签等）中的虚假信息，信息检测研究
8	普通内科	15	4.42	1987—2020	关注病毒、疫苗、健康、烟草等的虚假信息研究
9	工程学	13	3.83	2013—2019	信息业务流程，虚假信息检测、遏制传播等
10	信息科学与图书馆科学	13	3.83	1989—2019	信息检索和需求确定，对误导、人际欺骗、社交媒体的虚假信息研究

系网上获得的信息作为判断依据，有时是难以区分真伪的，或者说正确和错误的信息可能源自同一地方[22]。

第四，信息来源有限。虽然社交媒体上的信息访问量几乎是没有限制的，但用户的选择性阅读以及社交媒体基于阅读偏好的推送程序造成了他们的信息来源和观点的有限,或与他们当前的信念一致。一个人的选择依赖他潜在的社会关系网络，也就是说一个人接受和传播什么样的信息取决于临近人的选择、行为分布以及数量[23]。

3 社交媒体平台的责任与作用

因为存在和传播大量的虚假信息，社交媒体曾被指责[24]，但社交媒体在纠正虚假信息方面可以发挥重要和积极的作用。一些有效的纠正虚假信息的方法能够减少社交媒体用户对科学信息的误解，如进行内容管理、内容选择性暴露。虽然在阅读纠正信息后，一些用户否认他们的信仰发生改变，或产生矛盾的理解，但他们对于相关科学问题的态度有所改变[25]。

社交媒体的程序能快速检测正在传播中的虚假信息，在推特中，已经实现了可以检测可疑信息的监督学习技术，这些技术都基于机器学习算法或启发式算法，用以检测社交媒体发布的消息中的虚假信息[26]。社交媒体服务商可以直接过滤掉那些被识别的虚假信息。基于社交网络的信息过滤器，可以通过组建可信赖的在线社交朋友圈、由管理员控制流入的信息并为信息加上标签加速观点的永久化，形成对虚假信息的过滤[27]。

但社交媒体上每时每刻都有海量信息被生产和传播，且虚假信息可以隐藏于丰富的上下文中，从现有研究成果来看，现阶段没有能够完全检测、彻底解决社交媒体上虚假信息的方式。

而且，社交媒体的算法和解决方案有时候会导致用户看到更多能够加强他们原本意识形态的内容，一些社交媒体积极阻止虚假信息传播的技术方案反而促进了虚假信息的传播，如脸书为了解决向用户展示的内容中存在大量虚假信息而被批评的问题，利用算法让用户减少看到病毒视频和媒体新闻文章的机会，突出显示朋友之间的帖子，这种策略在某种程度上促进了虚假信息的传播，因为虚假信息可能就隐藏在朋友的帖子中，且他们不认为是错误的[28]。

4 纠正虚假信息的效果

对最主要或影响最大的虚假信息的关注、基于事实的富有逻辑的阐述、权威的信息发布来源被认为是纠正虚假信息的重点。

虚假信息存在持续影响效应被广泛论证，对信息接收者来说可信来源、重复出现、纠正延迟等会降低纠正效果[29]。对于专家而言，纠正社交媒体上的虚假信息的效果与公众和媒体的新闻素养有

关[30]。与科学相关的虚假信息可能导致错误的健康教育内容、健康消费趋势和公共政策，需要有针对性地制定纠正策略[31]。

为了确定在传染病暴发期间如何提高公众对虚假信息的认识以及纠正信息的策略，一项 700 人的在线实验在美国公众中进行，包含了不同的性别、年龄、种族、地区。在最初的虚假信息暴露之后，一组参与者接收到了不同类型的纠正信息（简单的反驳和阐述事实）和来源（政府卫生机构、新闻媒体和社会同行），另一组对照组没有接收到纠正信息。结果表明：第一，纠正信息的存在可以揭穿虚假信息。当被扩散的虚假信息是有关新出现的危机时，反驳相关说法可以显著地改变个人对危机严重性的看法；第二，对于简单的反驳和事实阐述这两种纠正虚假信息的方式，与个人对危机严重性的认知无关，仅仅是虚假信息的存在就足以改变人们的看法。相对而言，与简单的反驳相比，基于事实的阐释效果更好；第三，比较三个有影响力的信息源 (政府卫生机构、新闻媒体和社会同行) 的传播效果，以了解个人对不同纠正信息来源的反应，与社会同行相比，政府机构和新闻媒体作为信息来源在提高可信度方面更为成功[32]。

一些研究发现，当人们缺乏专业知识和技能时，就容易受到虚假信息的误导。但这也为控制虚假信息的传播和影响提供了方法。研究者建议通过解释虚假信息中的谬误推理来预防虚假信息的传播，并提出基于批判性思维方法的策略来分析和检测虚假信息中的不良推理。策略包括详细地描述论点结构、确定信息的真实性、检查信息的有效性、隐藏前提假设或使用模糊的语言。这种方式的优势在于，能够让那些缺乏专业科学知识的人接受[33]。

更多的研究表明，如果虚假信息造成了危机，那么纠正虚假消息更有效的方式是披露与当下危机有关的正确的信息，而不是试图改变人们原有的态度。在这种情况下，纠正虚假信息可能更需要关注当前的危机情况[34]。

5　检测和识别虚假信息的方法

如何识别虚假信息，并控制其基于社会关系网络的病毒式的传播，越来越受到学者的关注。特别是使用计算机网络技术，如机器学习和自然语言处理技术，基于内容的文本特征或者信息来源特征，自动识别和监控虚假信息，并使用修改网络拓扑结构等方式限制虚假信息传播，如暂停账号、删除链接等，取得了理论和实践上的成效。

第一类技术基于文本特征实现虚假信息的识别和控制。通过提取符合特定的纠正模式的文本段落，将这些文本段落聚类成不同的虚假信息的主题，通过选择主题达到有效控制的目的[35]。基于机器学习和文本分类模型的框架，可通过区分那些可靠和不可靠的信息来实现虚假信息的早期检测[36]。针对网络新闻可信度的指标被提出，包括内容（标题代表性、点击链接、专家引用、研究引用、置信度校准、逻辑谬误、语气、推断）和语境（独创性、事实核实、代表性引用、引用的声誉、广告数量、电话数量、垃圾广告、广告和社交电话的位置）两个指标[37]。

基于网络数据深度分析，检测社交媒体中可能出现的语义攻击①类型，并对其进行精准的分类。研究者使用自动化的网络分析算法和人工的方式标注可能存在的语义攻击类型：造谣攻击和错误攻击。绘制包含用户节点和消息节点的 Retweet 图，对社交媒体中语义攻击进行分类。然后，使用基于模块化的社区检测算法（Gephi 软件），检测信息传播情况。基于此，构建传播图，利用 K-Core 分解，分离出其中可能的虚假信息和涉及的用户节点。此方法可以在可接受的时间框架内，有效地识别与控制虚假信息的传播[38]。

第二类技术是基于社会关系节点的模型。测试社交媒体上社会关系节点之间的影响概率和强弱关系的不同学习模型被提出，还可以用于预测用户是否会执行某一项操作，以及受邻居节点影响后可能执行操作的时间，其中包括：静态模型（伯努利分布、雅卡指数、部分信用）、连续时间（CT）模型、离散时间（DT）模型，使用 Flickr 数据集验证[39]。

社交媒体中竞争活动的概念的提出解决了如何找到从网络中的某个节点开始传播的虚假信息，并使用限制活动的概念来抵消虚假信息的影响的问题。Budak 等[40] 比较了贪婪算法与各种启发式算

法的性能，认为在大多数情况下，启发式算法比贪婪算法性能更好，并提出了一种基于随机生成树的预测算法，在识别虚假信息时可以容忍一定比例的数据缺少，其性能能够满足需求。

为了解决虚假信息的误报、预防问题，特别是显著独立级联模型下的虚假信息预防问题，Tong 等 [41] 提出了一种新的采样方法，与传统的对所有节点一视同仁、对节点均匀采样的逆向采样方法不同，该方法采用混合采样的方法，其算法核心是设计有效的采样方法来估计函数值，对易受虚假信息影响的用户赋予较高的权重。因此，这种新的采样方法在生成有效样本方面具有更强的能力。他们还在收集大量数据的基础上对所提出的方法进行了实验评估，结果显示其性能与传统逆向采样方法相比具有显著的优越性。

基于系统网络体系结构（SNA）的度量，一个自动化的解决方案被提出，可以用来识别在危机期间传播虚假信息的用户。通过开发可以实时评估内容的可信度的技术，来阻止虚假信息在推特上的传播，这可以用来过滤数据中令人难以置信的虚假信息。使用有监督的机器学习和相关反馈方法，研究者发现了基于推特特性（主题和来源）的推特排名，有助于评估事件消息中信息的可信度。这些算法可以帮助用户对推特的可信度做出判断 [42]。

在检测和识别虚假信息的过程中，除了计算机算法外，认知心理学的方法也被广泛使用。

用户在社交媒体上转发的虚假信息，受到含有欺骗性的信息或符号的暗示 [43]。认知心理学的理论可以被用于评估社交媒体上的虚假信息和宣传信息的传播，通过分析信息的一致性、信息的相关性、来源的可信度、信息的普遍可接受性，可推断出那些隐藏在信息中的误导和欺骗内容的线索，进而制定控制虚假信息传播的解决方案 [44]。

使用可以引导人点击、转发的方式，成为在社交媒体上传播虚假信息的高效方式。针对这种情况，通过自动检索，识别社交媒体上的文本、图片中的相关线索，是一种有效途径 [45]。

6 虚假信息的预防与控制对策

在控制虚假信息传播方面，可以先建构虚假信息的传播模型，用以识别最有影响力的节点，在这些节点上对虚假信息进行净化（删除或修正），以阻止该节点上虚假信息的传播。再通过使用多个有影响力的节点发起反击的方式，控制虚假信息传播 [46]。采用节点保护的方式，将虚假信息的传播限定在一个预先确定的速度和时间段内，在具有高度影响力的节点中，找到最小的一组，控制和清除这一组节点上的虚假信息，也是一种有效方法 [47]。

有的研究还提出了通过改变扩散网络的拓扑结构来控制虚假信息传播的策略 [48]。通过独立级联或线性阈值模型的技术手段切断社交媒体上用户之间的链接，或者对其进行封号处理，以达到控制虚假信息传播的目的 [49]。

但是，这些扩散模型中的参数很难在现实数据中被提取出来。因此，Song 等 [50] 的研究利用网络中发生的实际级联，基于场景为虚假信息设置不同的信息传播概率模型，选择最优的链接子集，将链路去除问题转化为混合整数规划问题，从而最大限度地消除虚假信息和谣言的传播。

Lewandowsky 等人 [51] 认为意识形态和个人的世界观可能是降低控制和纠正虚假信息成本的主要障碍，但仍然有许多有效的方法可以减少虚假信息的影响，如适当的控制设计、控制结构和应用方式，可以最大限度地提高控制的效果。他们随后提出了更具体的应对虚假信息传播的对策：第一，为虚假信息提供有理有据的纠正解释；第二，删除虚假信息，以降低持续影响；第三，在纠正的过程中如果需要可以提及虚假信息，但要同时给出明确的说明，以免混淆；第四，注意受众的世界观，在执行过程中很可能会有所偏向；第五，纠正信息需简洁有力，且比之前的虚假信息更有吸引力。

公共危机（如自然灾害、公共卫生等）事件是缺乏科学证据的虚假信息泛滥的主要情景之一，它可能会引起媒体、公众的过度反应和恐慌，甚至会

① 语义攻击：Semantic Attacks，一种利用网络应用缺陷插入虚假的数据包破坏合法通信的网络攻击手段。

削弱可靠信息或官方公告的传播效果[52]。如果公共卫生专家在公共卫生危机期间积极应对并及时地纠正虚假信息，将有助于阻止危机时期虚假信息的传播[53]。面向公众的公共卫生专家或医护人员应具备最新的、准确的研究成果和信息，同时，应基于自然语言处理或大数据挖掘技术对所有社交媒体上的没有科学依据的内容进行检测和删除[54]。

也有研究模拟公众卫生危机的情况，如通过假设爆发传染病评估公众的反应，预判相关虚假信息的传播路径、效果和范围，并以此作为预防危机或做好应对的准备，成为危机管理实践的重要手段[55]。

还有的学者研究了应对虚假信息的时机。面对公共卫生事件，等待正确的时机或更严谨完整的证据链也许不是最优选择，越快纠正或反驳已经发现的虚假信息（特别是在公众对此的信念和记忆根深蒂固之前），越有可能成功[56]。

7 总结与讨论

社交媒体上虚假信息被有意或无意地传播，可以说是人类追求获得和交换信息的便利性、即时性和互动性带来的必然结果，特别是随着社交媒体使用量和用户量的剧增，在信息发布和传播更容易的同时，有海量的、鱼龙混杂的信息广泛传播。因此，能否在虚假信息泛滥或在社会中造成严重的负面影响前，有效地识别并控制其传播，是虚假信息相关研究关注的核心问题。上述研究呈现出以下特点：

第一，有关虚假信息的研究属于跨学科研究，多个学科领域关注到虚假信息的问题，但各领域关注的重点和采用的研究方法等有所不同。传播学视角的研究一般比较关注社交媒体平台的性质和虚假信息的传播机制，如发布信息的门槛低、追求吸引眼球和流量、媒介平台监管困难和缺失，以及如何有效地纠正虚假信息及其效果等；心理学视角的研究一般关注虚假信息传播者的认知、动机、行为和社会影响等；计算机科学领域的研究则侧重于虚假信息的自动识别与控制技术等方面，各类研究都在不断深入和拓展。

第二，就虚假信息的识别与控制这一主题，研究者从理论和技术实践上进行了较为广泛的讨论，特别是基于计算机算法的解决方案较为丰富。而且对于一些特殊领域、典型事件中的虚假信息的控制进行了深入探讨，如公共卫生事件、群体性事件等。但是，有关如何基于公众认知、传播和人际关系的特质来利用算法识别和控制社交媒体上虚假信息的传播，以发挥及时和正面的作用，仍需要深入讨论和在实践中检验。

第三，关于个体在社交媒体虚假信息的识别与控制中的作用，学者们进行了大量的探讨，也是近几年的研究热点。很多研究提及了个体的错误信念的形成、影响，以及由科学素养、信息素养、经济条件、教育水平等多方面因素导致的认知和识别能力的差异。但是，影响或改变个体形成的错误信念的有效手段以及社交媒体上用户节点的可操作性策略等方面还有待深入探讨。

第四，现阶段社交媒体采用的解决方案，以及基于优先数据检验的理论上适用的算法模型，是否可在现实中长期有效控制虚假信息的传播，是否会带来其他负面影响，仍需要进一步检验和发展。

第五，多数研究对象为已产生影响的被普遍承认的虚假信息，但是未来可能出现的各种类型的虚假信息，或隐藏于丰富上下文的、暗示性的、引导性的虚假信息，能否建立起一些标准或模型发挥预警或预防作用，将会是未来研究的重点发展方向。■

参考文献：

[1] Berenbaum R, Scheufele D, Hallman W K, et al. Advancing the science and practice of science communication: Misinformation about science in the public sphere[EB/OL]. (2019-04-03)[2020-06-30]. http://www.nasonline.org/programs/nas-colloquia/completed_colloquia/advancing-the-science.html.

[2] Vosoughi S, Roy D, Aral S. The spread of true and false news online[J]. Science, 2018，359(6380): 1 146-1 151.

[3] Martinez T. The Effects of Cognitive Engagement while Learning about Misinformation on Social Media[D]. Arizona: Arizona State University, 2019.

[4] Wang Y X, McKee M, Torbica A, et al. Systematic literature review on the spread of health-related

misinformation on social media[EB/OL]. [2020-07-26]. https://www.sciencedirect.com/science/article/pii/S0277953619305465?via%3Dihub.

[5] Borchard E, Anderson M A. Web of deceit: Misinformation and manipulation in the age of social media[J]. Online Information Review, 2013, 37(1): 155-156.

[6] Bode L, Vraga E K. See something, say something: Correction of global health misinformation on social media[J]. Health Communication, 2017, 33(9): 1-10.

[7] Dietram A S, Nicole M K. Science audiences, misinformation, and fake news[J]. PNAS, 116(16): 7 662-7 669.

[8] Jackson M. The diffusion of behavior and equilibrium properties in network games[J]. American Economic Review, 2007, 97(2): 92-98.

[9] Goffman W, Newill V A. Generalization of epidemic theory. An application to the transmission of ideas[J]. Nature, 1964, 204: 225-228.

[10] Antoniadis S, Litou I, Kalogeraki V. A model for identifying misinformation in online social networks[C]. Debruyne C, Panetto H, Meersman R, et al. Confederated International Conferences. On the Move to Meaningful Internet Systems: OTM 2015 Conferences. Switzerland: Springer International Publishing, 2015 (9 415): 473-482.

[11] NBC. "MisinformationIs" Dictionary.com's Word of the Year[N/OL]. (2018-12-26)[2020-06-30]. https://www.nbcmiami.com/news/national-international/misinformation-is-dictionarycom-word-of-the-year/2028407/.

[12] Fitzgerald M A. Misinformation on the internet: Applying evaluation skills to online information[J]. Emerg Libr, 1997, 24(3): 9-14.

[13] Wu L. Misinformation Detection in Social Media[D]. Arizona: Arizona State University, 2019.

[14] Lazer D M J, Baum M A, Benkler Y, et al. The science of fake news[J]. Science, 2018, 359 (6 380): 1 094-1 096.

[15] National Science Board. Science and engineering Indicators 2018[R]. Alexandria, VA: National Science Foundation, 2018.

[16] Rojecki A, Meraz S. Rumors and factitious informational blends: The role of the web in speculative politics[J]. New Media & Society, 2016, 18(1): 25-43.

[17] Kuklinski J H, Quirk P J, Jerit J, et al. Misinformation and the currency of democratic citizenship[J]. Journal of Politics, 2000, 62(3): 790-816.

[18] Chen X, Sin S C J, Theng Y L, et al. Why students share misinformation on social media: Motivation, gender, and study-level differences[J]. The Journal of Academic Librarianship, 2015, 41(5): 583-592.

[19] Stahl B C. On the difference or equality of information, misinformation, and disinformation: A critical research perspective[J]. Information Sciences, 2006(9): 83-96.

[20] Nyhan B, Reifler J. When Corrections Fail: The Persistence of Political Misperceptions[J]. Political Behavior, 2010, 32(2): 303-330.

[21] Szpitalak M, Polczyk R. How to induce resistance to the misinformation effect? Characteristics of positive feedback in the reinforced self-affirmation procedure[J]. Psychology Crime & Law, 2019, 25(7): 771-791.

[22] Connor C O, Weatherall J O. The Misinformation Age: How False Beliefs Spread[M]. New Haven, CT: Yale University Press, 2019: 12-18.

[23] Jackson M. The diffusion of behavior and equilibrium properties in network games[J]. American Economic Review, 2007, 97(2): 92-98.

[24] Vishwanath A. Habitual Facebook use and its impact on getting deceived on social media[J]. Journal of Computer-Mediated Communication, 2014, 20(1): 83-98.

[25] Bode L, Vraga E K. In related news, that was wrong: The correction of misinformation through related stories functionality in social media[J]. Journal of Communication, 2015, 65(4): 619-638.

[26] Ratkiewicz J, Conover M, Miess M, et al. Truthy: Mapping the spread of astroturf in microblog streams[C]. Sadagopan S, Ramamritham K, Kumar A, et al. WWW '11: Proceedings of the 20th international conference companion on World Wide Web. New York: Association for Computing Machinery, 2011: 249–252.

[27] Kanekar A S, Thombre A. Fake medical news: avoiding pitfalls and perils[J]. Family Medicine and Community Health, 2019, 7(4): e000142.

[28] Isaac M. Facebook overhauls news feed to focus on what friends and family share[N/OL]. (2018-01-11) [2020-05-28]. https://www.seattletimes.com/business/facebook-overhauls-news-feed-to-focus-on-what-friends-and-family-share/.

[29] Walter N, Tukachinsky R. A meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop it?[J]. Communication Research, 2020, 47(2): 155-177.

[30] Vraga E K, Bode L, Tully M. Creating news literacy messages to Enhance expert corrections of misinformation on Twitter[DB/OL]. (2020-01-30)[2020-06-28].https://journals.sagepub.com/doi/10.1177/0093650219898094.

[31] Smith C N, Seitz H H. Correcting misinformation about neuroscience via social media[J]. Science Communication, 2019, 41(6): 790-819.

[32] Toni G L A, Jin Y. Seeking formula for misinformation treatment in public health crises: The effects of corrective information type and source[J]. Health Communication, 2019, 35(5): 1-16.

[33] Cook J, Ellerton P, Kinkead D. Deconstructing climate misinformation to identify reasoning errors[J]. Environmental Research Letters, 2018, 13(2): 24 018.

[34] Ecker U K H, Lewandowsky S, Fenton O, et al. Do people keep believing because they want to? Preexisting attitudes and the continued influence of misinformation[J]. Memory & Cognition, 2014, 42(2): 292-304.

[35] Nabeshima K, Mizuno J, Okazaki N, et al. Mining false information on Twitter for a major disaster situation[C]. Slezak D, Schaefer G, Vuong ST, et al. Active Media Technology. Heidelberger Platz: Springer-Verlag Berlin, 2014: 96-109.

[36] Liu Y, Yu K, Wu X F, et al. Analysis and detection of health-related misinformation on Chinese social media[J]. IEEE ACCESS, 2019(7): 154 480-154 489.

[37] Zhang A X, Ranganathan A, Metz S E, et al. A structured response to misinformation: Defining and annotating credibility indicators in news articles[DB/OL]. (2018-04-27)[2020-06-24]. https://dl.acm.org/doi/abs/10.1145/3184558.3188731.

[38] Kumar K P K, Geethakumari G. Analysis of semantic attacks in online social networks[J]. International Journal of Trust Management in Computing and Communications, 2014, 2(3): 207-228.

[39] Goyal A, Bonchi F, Lakshmanan L V S. Learning influence probabilities in social networks[C]. Davison B D, Suel T. WSDM'10: Proceedings of the Third International Conference on Web Search and Web Data Mining. New York: Association for Computing Machinery, 2010: 241-250.

[40] Budak C , Agrawal D , Abbadi A E . Limiting the spread of misinformation in social networks[C]. Sadagopan S, Ramamritham K, Kumar A, et al. WWW'11: Proceedings of the 20th International Conference on World Wide Web. New York: Association for Computing Machinery, 2011: 665-674.

[41] Tong G A, Du D Z. Beyond Uniform Reverse Sampling: A Hybrid Sampling Technique for Misinformation Prevention[DB/OL]. (2019-12-20)[2020-06-28]. https://arxiv.org/abs/1901.05149v2.

[42] Gupta A, Kumaraguru P. Misinformation in Social Networks, Analyzing Twitter During Crisis Events[M]. New York: Springer New York, 2014: 922-924.

[43] Karlova N A, Fisher K E. A social diffusion model of misinformation and disinformation for understanding human information behaviour[J]. Information Research, 2013, 18(1): 1-17.

[44] Kumar K P K, Geethakumari G. Detecting misinformation in online social networks using cognitive psychology[J]. Human-centric Computing and Information Sciences, 2014, 4(1): 14.

[45] Chen Y M, Conroy N J, Rubin L. Misleading Online Content: Recognizing Clickbait as “False News”[C]. Abouelenien M, Burzo M. Proceedings of the 2015 ACM On Workshop on Multimodal Deception Detection. New York: Association for Computing Machinery, 2015: 15-19.

[46] Budak C, Agrawal D, Abbadi AE. Limiting the spread of misinformation in social networks[C]. Sadagopan S, Ramamritham K, Kumar A, et al. WWW'11: Proceedings of the 20th International Conference on World Wide Web. New York: Association for Computing Machinery, 2010:

665–674.

[47] Nguyen N P, Yan G, Thai M T, et al. Containment of misinformation spread in online social networks[C]. Contractor N, Uzzi B. Proceedings of the 3rd Annual ACM Web Science Conference. New York: Association for Computing Machinery, 2012: 213-222.

[48] Kempe D, Kleinberg J M, ÉvaTardos. Influential Nodes in a Diffusion Model for Social Networks[DB/OL]. (2005-07-15)[2020-06-24]. https://doi.org/10.1007/11523468_91.

[49] Hemmati M, Smith J C, Thai M T. A cutting-plane algorithm for solving a weighted influence interdiction problem[J]. Computational Optimization and Applications, 2014, 57(1): 71–104.

[50] Song Y, Dinh TN. Optimal containment of misinformation in social media: A scenario-based approach[DB/OL]. [2020-09-01]. https://link.springer.com/chapter/10.1007%2F978-3-319-12691-3_40, 2014-11-13.

[51] Lewandowsky S, Ecker U K, Seifert C M, et al. Misinformation and its correction: Continued influence and successful debiasing[J]. Psychological Science in the Public Interest, 2012, 13(3): 106-131.

[52] Merino J G. Response to ebola in the US: misinformation, fear, and new opportunities[J]. BMJ, 2014, 349(13): 6 712.

[53] Tan A S L, Lee C J, Chae J. Exposure to health (Mis) information: Lagged effects on young adults' health behaviors and potential pathways[J]. Journal of Communication, 2015, 65(4): 674-698.

[54] Tasnim S, Hossain M M, Mazumder H. Impact of rumors and misinformation on COVID-19 in social media[J]. Journal of Preventive Medicine and Public Health, 2020, 53(3): 171-174.

[55] Youngblood S. Ongoing crisis communication: Planning, managing, and responding, 2nd Edition (Coombs, W. T.) and Handbook of risk and crisis communication (Heath, R. L. and O'Hair, H. D. Eds.) [Book reviews][J]. IEEE Transactions on Professional Communication, 2010, 53(2): 174-178.

[56] Leticia B, Vraga E K. In related news, that was wrong: The correction of misinformation through related stories functionality in social media[J]. Journal of Communication, 65(4): 619-638.

Identification and Control of Misinformation on Social Media: Analysis of Literatures Based on Web of Science

CHI Yan-wei, ZHANG Zeng-yi

(School of Humanities, University of Chinese Academy of Sciences, Beijing 100049)

Abstract: The spread of misinformation on social media has led to serious negative effects, causing public concern and panic. How to effectively and accurately identify misinformation and control its spread and impact has been widely studied in the world. This paper reviews the literature on misinformation on social media of Web of Science. This paper analyzes relevant research papers published in foreign authoritative journals in recent years from the following aspects: difficulty to identify and control the spread of misinformation, the role of social media platform, the misinformation correct strategy, the computer techniques of detection and identification, the psychology means of detection and recognition, and the strategies for prevention and control of misinformation. This paper reveals the research status, hotspots and trends of the international academic circles on the identification and control of misinformation, hoping to provide reference for the theoretical research and policy practice of the identification and control of false information transmission in China.

Key words: misinformation; social media; information dissemination