面向社交网络的异常传播研究综述

卢昆[1]，张嘉宇[1]，张宏莉[1]，方滨兴[1,2]
（1.哈尔滨工业大学网络空间安全学院，黑龙江 哈尔滨 150001；2.广州大学网络空间先进技术研究院，广东 广州 510006）

摘　要：异常传播是当今在线社交网络中频繁出现的一种非传统的信息传播模式。为了完整地认知社交网络中异常传播的整体过程，将异常传播生命周期系统归纳为潜伏期、扩散期、高潮期和衰退期4个阶段。针对异常传播在不同阶段所存在的科学问题，从微观和宏观视角分别定义和划分了异常信息、异常用户以及异常传播、传播抑制4个当前热门的研究领域，详细综述了当前4个研究领域下的主要研究任务以及相关研究进展，并分析了现有方法存在的问题，对社交网络异常传播领域的未来研究方向进行了展望，为后续研究提供便利。

关键词：异常传播；在线社交网络；异常信息；传播抑制
中图分类号：TP391
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2024045

Survey on anomaly propagation research for social networks

LU Kun[1], ZHANG Jiayu[1], ZHANG Hongli[1], FANG Binxing[1,2]
1. School of Cyberspace Science, Harbin Institute of Technology, Harbin 150001, China
2. Cyberspace Institute of Advanced Technology, Guangzhou University, Guangzhou 510006, China

Abstract: Anomaly propagation is a non-traditional mode of information dissemination that frequently occurs in today's online social networks. To fully comprehend the overall process of anomaly propagation in social networks, the anomaly propagation lifecycle system was classified into incubation period, diffusion period, climax period, and decline period. In response to the scientific problems existing in different stages of anomaly propagation, four popular research fields of anomaly information, anomaly users, anomaly propagation, and propagation containment were defined and divided respectively from micro and macro perspectives. The main research tasks and related research progress in the four current research areas were reviewed and summarized in detail, problems in existing methods were analyzed. The future research directions on anomaly propagation in social networks were prospected, providing convenience for subsequent research.

Keywords: anomaly propagation, online social network, anomaly information, propagation containment

0　引言

随着移动互联网以及智能终端设备的飞速发展，社交网络已经成为世界信息传播的主要载体之一，逐渐在人们的生活中扮演着非常重要的角色，甚至影响人们的思维、认知和决策。据统计，截至2023年4月，全球约48亿人使用社交网络，接近全球人口的60%。全球最大的社交平台Facebook用户达29亿人，我国的微信、抖音等平台用户均超过10亿人。社交网络已成为人们获取信息的主要途径，所承载的信息量呈指数级增长，其信息内容丰富、信息分享便捷、花费成本低等特点，深受人们喜爱。

学术界很早就有针对社交网络中异常传播行为的研究，比如谣言检测[1]、假新闻检测[2]、恶意传

收稿日期：2023–11–10；修回日期：2024–01–17
通信作者：张宏莉，zhanghongli@hit.edu.cn
基金项目：国家重点研发计划基金资助项目（No.2016QY03D0501）
Foundation Item：The National Key Research and Development Program of China (No.2016QY03D0501)

播用户识别[3]等一直以来都是社交网络领域的研究热点。由于社交网络天然的传播属性，无论是谣言等数据组成的网络客体，还是恶意传播群体等用户组成的网络主体，都不是孤立存在的，而是社交网络中异常传播的一部分。但是，以往研究并没有从异常传播的整体过程去归纳总结其中所产生的一系列科学问题。因此，本文系统归纳了社交网络中异常传播生命周期的4个阶段，即潜伏期、扩散期、高潮期和衰退期；然后以生命周期为主线，分别从微观和宏观视角分析总结了在前两个阶段的异常信息、异常用户和后两个阶段的异常传播、传播抑制共4个研究领域的相关内容及方法，并对未来的研究方向进行展望。总体研究路线如图1所示。

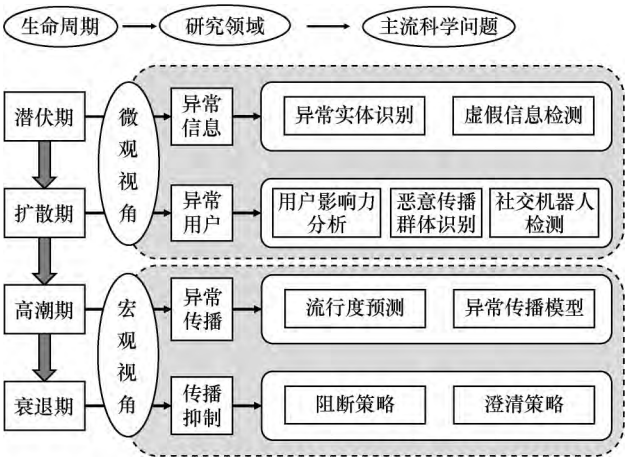

图1　总体研究路线

1　相关概念

本文使用的主要符号及其含义如表1所示。

表1　主要符号及其含义

符号	含义
c_i	字符串序列中的字符
l_i	字符对应的标签
r_j	帖子及其一系列属性，包括文本、图像、评论数等
u_i	用户及其属性组成，包括姓名、注册时间、职业等
a_p	用户u_i对r_j执行的点赞、转发等操作
t_q	用户u_i对r_j执行操作的时间戳
$G=(U,E)$	包含用户集合U和关系集合E的社交网络图
T	用户的语义信息，例如推文和回复
N	用户的邻居信息，即关注者和追随者
h_i	转发量、发布时间等异常信息相关特征

1.1　异常传播定义

异常指的是数据中不符合明确定义的正常行为的模式[4]。以往的研究中并未给出社交网络中异常传播的明确定义，本文尝试从微观和宏观视角阐释异常传播的定义。异常传播是指在社交网络中，由异常用户推动异常信息传播，且与正常传播趋势相异的传播过程，比如谣言扩散过程或者大量恶意传播用户参与的恶意传播过程。

在本文中，微观视角指的是在异常传播的潜伏期和扩散期阶段，异常传播尚未形成显著的传播趋势，此时更应关注实际参与社交网络中异常传播的客体及主体视角，即异常传播内容和异常用户视角；宏观视角指的是在异常传播的高潮期和衰退期阶段，异常信息与真实信息、异常用户与真实用户均已参与异常传播过程，此时的传播规模已较为明显，更应考虑整体异常传播过程，所以基于社交网络的载体视角，即基于社交平台视角。

下面将具体阐述异常传播不同阶段下所存在的相关定义及任务。

1) 异常信息

本文中的异常信息主要是指社交网络中错误信息、虚假信息以及未经证实但带有不良影响的信息的统称，比如谣言、假新闻、“标题陷阱”等。社交网络异常信息领域的主要任务有异常实体识别、虚假信息检测等。

定义1　异常实体识别[5]。给定一段长度为n的社交文本内容，即字符串序列$S=\{c_1,c_2,\cdots,c_n\}$，异常实体识别旨在为该序列的每一个字符标注标签，并根据所输出的标签序列$L=\{l_1,l_2,\cdots,l_n\}$从原始序列S中抽取出异常实体。

定义2　虚假信息检测[6]。给定一个社交网络中特定的声明$s=<R,U,E>$，它一共包含一组相关的n个帖子$R=\{r_1,r_2,\cdots,r_n\}$、一组相关的m个用户$U=\{u_1,u_2,\cdots,u_m\}$以及用户与帖子之间的k个操作集合$E=\{e_1,e_2,\cdots,e_k\}$，其中$e_o=\{r_j,u_i,a_p,t_q\}$。虚假信息检测任务是学习预测函数$F(s)\rightarrow\{0,1\}$来判断声明s是否为虚假信息。

2) 异常用户

异常用户指的是社交网络中参与发布或者传播异常信息的网络主体，其中以恶意传播群体和社交机器人群体为主要组成部分，还有一部分是为谋取非法利益而恶意帮助异常信息传播的高影响力网络主体。与之相关的任务有用户影响力分析、恶意传播群体识别以及社交机器人检测等。

定义 3　用户影响力分析[7]。给定社交网络图 $G=(U,E)$，用户影响力分析任务旨在根据社交平台异常用户 $u_i \in U$ 的特征以及与其他用户的关系 $e_{u_i} \in E$ 来预测用户在社交网络平台的影响力值。

定义 4　恶意传播群体识别[3]。给定一组用户 U，其社交网络图为 G，内容信息为 $X \in \mathbb{R}^{m \times n}$，数据集中部分用户的身份标签信息为 $Y \in \mathbb{R}^{n \times c}$（即训练数据），恶意传播群体识别任务旨在利用上述信息学习分类器 W，以自动为未知用户（即测试数据）分配身份标签，即是否为恶意传播用户。

定义 5　社交机器人检测[8]。给定一组用户 U 及其信息 T、P 和 N，学习社交机器人检测函数 f: $f(U(T,P,N)) \rightarrow \hat{y}$，使得 $\hat{y}$ 接近真实值 y，即真实用户或者社交机器人账号。

3) 异常传播

异常传播是信息传播的一种，是异常信息发送和接收的过程，也可以看作传播者和接收者之间资源共享的过程[9]。异常传播往往是一个事件的动态扩散过程，所以针对异常传播的研究主要是异常传播流行度预测和异常传播规律建模等。

定义 6　异常传播流行度预测[10]。给定异常信息 d_i 的特征集合 $H=\{h_1,h_2,\cdots,h_m\}$，异常传播流行度预测旨在预测该事件未来一段时间内的流行趋势。

定义 7　异常传播规律建模[11]。给定社交网络图 $G=(U,E)$，异常传播规律建模旨在通过节点 $u_i \in U$ 的状态转化规律以及节点之间的连接关系 $e_{u_i} \in E$，描述和预测社交网络中的异常传播过程。

4) 传播抑制

传播抑制指的是在异常信息正在传播的社交网络中，研究一种抑制策略，其能够最大限度地减少传播在社交网络中的影响[12]。

定义 8　传播抑制[13]。给定社交网络图 G、扩散模型 μ、一组恶意节点 MN，且 $|MN| \geqslant 1$，传播抑制任务旨在找到并应用策略 H 来最小化异常信息的影响。该目标通常定义为

$$H^* = \arg\min \varphi_{\mu}^{H}(G,MN) \tag{1}$$

其中，$\varphi_{\mu}^{H}(G,MN)$ 代表在 H 策略下，基于扩散模型 μ 进行传播时社交网络图 G 受到的影响力。

1.2　异常传播周期

社交网络中异常传播是一个周期过程，包括潜伏期、扩散期、高潮期和衰退期，如图 2 所示。

潜伏期是异常信息刚进入社交网络的阶段，异常信息还处在局部发酵过程，并未演变成异常传播事件，该阶段内可通过异常实体检测等技术锁定突然出现的异常实体，也可以通过虚假信息检测技术实现针对谣言、假新闻等异常信息的早期发现，提前预警可能发生的异常传播。

扩散期是异常传播的关键时期，该阶段通常由恶意传播群体及社交机器人群体作为主要参与者。他们通过转发、评论等社交行为加速异常传播，扩大异常传播的影响力和覆盖范围。在他们的引导下，将会有越来越多的真实用户接触并参与此次异常传播。在此阶段，用户影响力分析、恶意传播群体识别和社交机器人检测成为主要研究领域。

高潮期是异常传播被广为人知的重要时期，该阶段真实用户成为异常传播的主要参与者，有的用

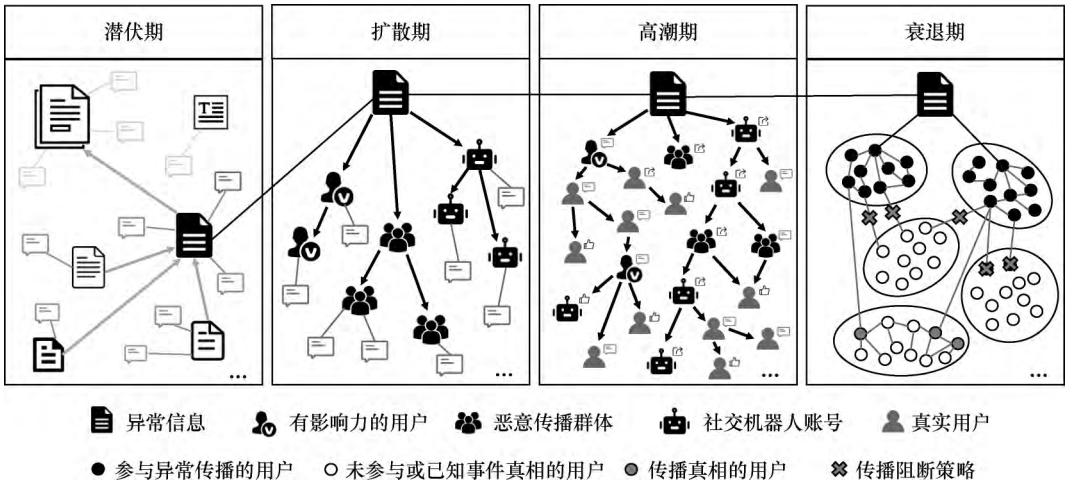

图2　社交网络中异常传播的生命周期

户被异常信息欺骗，成为“感染者”，并促进异常传播；有的用户持反对态度，成为异常传播中的“免疫者”或者“抑制者”。参与用户数量的不断增加，为异常传播流行度预测和异常传播模型的研究提供了大量数据支撑。

衰退期是异常传播过程的最后一个时期，由于异常传播在高潮期给大量真实用户带来了严重的不良影响，针对异常传播的抑制策略开始逐渐控制异常传播的影响力和范围，最终使其消散。常见的抑制策略有两种，一种是阻断传播节点及边的方法，另一种是寻找适合节点传播事件真相以抑制异常信息的传播。

2 面向社交网络的异常信息

在异常传播潜伏阶段，异常信息率先出现在社交平台上，其在发布初期往往混杂在海量网络数据中，影响力较小，极难被识别出来，但如果不能及时做出处理，可能会在短时间内迅速扩散，并在社交网络上引起广泛关注。社交网络中异常信息往往具有强烈的虚假性、误导性和煽动性，通常由恶意传播者策划并发布，旨在误导社交网络用户对社会事件的观点和态度。因此在潜伏期阶段，针对异常信息的早期发现及检测工作就显得尤其重要。

本文将异常传播潜伏期涉及异常信息的任务划分为两个部分，即社交网络异常实体检测和虚假信息检测。现有的社交网络异常实体检测技术主要是挖掘出隐藏在社交网络信息中的异常实体；虚假信息作为异常信息的主要形式，利用相关检测技术能够从网络信息中精准定位异常信息，从而有效地减少异常信息在社交网络中的后续传播。

2.1 异常实体检测

在社交网络平台中，文本是最常见的内容类型，也是用户交流、分享信息和表达观点的主要工具。鉴于文本具有编造成本低、传播速度快、受众群体广的特点，以往的异常信息大多以文本的形式出现[14]。因此社交网络异常实体检测任务旨在识别和分类社交网络文本中可能被用于进行欺诈、虚假信息传播、恶意攻击等其他异常行为的实体。异常实体的检测过程如图3所示，首先，对收集到的网络文本进行格式转换、数据清洗等预处理操作。其次，将多种特征信息输入异常实体检测模型进行计算，以提取潜在的异常实体信息。最后，根据标签从文本中检测并抽取出异常实体。

与新闻、金融等传统领域的实体检测任务相比，社交网络异常实体检测更具挑战性[15]。

1) 上下文信息有限。受限于各大社交网络平台的发文规则，用户发布的文本内容有字数限制，可以利用的上下文语境信息有限。

2) 口语化现象严重。用户发布相关内容时往往不遵循严格的语法规则，口语化表达较多，例如简写、缩写等非正式语言。

3) 噪声干扰较高。社交网络文本中可能包含大量的表情符号、网址等噪声信息，这些噪声信息会严重干扰实体检测的准确性。

针对上述难点，学者们开展了大量相关研究，其方法演化过程分为3个阶段：基于规则的方法、基于特征的统计方法和基于深度学习的方法[5]。

早期的社交网络实体检测方法主要依赖于手工设计的规则和词典[16-18]，利用词性标注、语法解析等手段，结合人工制定的规则和实体词典进行实体

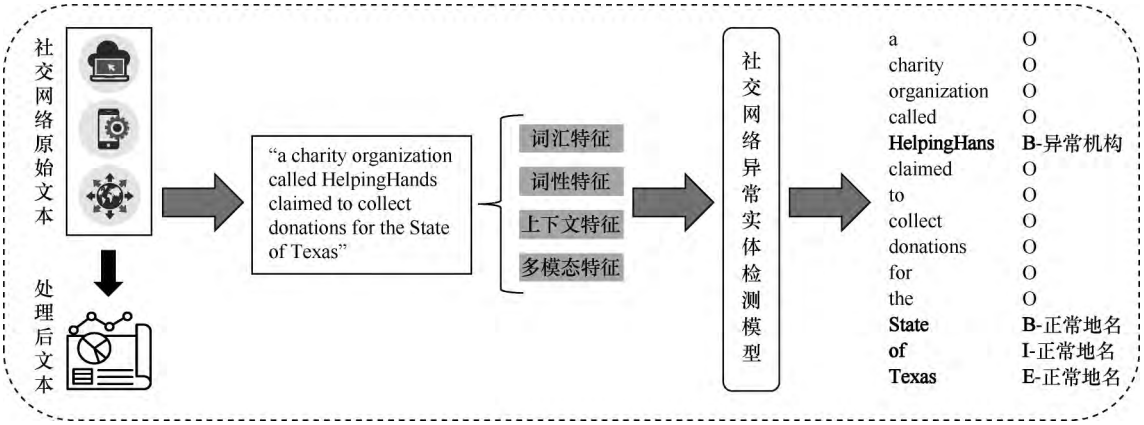

图3 社交网络中异常实体的检测过程

检测。由于社交网络文本的不规范性，传统的词性标注、语法解析等技术难以对其进行准确处理。

为了解决基于规则的方法的局限性，研究者开始尝试使用基于特征的统计方法，如支持向量机（SVM, support vector machine）[19]、条件随机场（CRF, conditional random field）[20]等。此类方法虽然为检测模型提供了一定的泛化能力，但仍然受限于人工设计特征的质量以及覆盖范围。

随着数据量的增加以及深度学习的兴起，基于深度学习的方法开始大量出现。相较于前两种方法，此类方法性能有较大幅度的提升，逐渐成为该任务的主流方法。基于深度学习的方法可以自动提取文本的语义和上下文信息，不需要人工设计特征，提高了模型的泛化能力。并且深度学习检测模型还能够处理大规模的文本数据，具有较高的效率和准确性。

文献[21]向BiLSTM-CRF（bidirectional long short-term memory and conditional random field）[22]模型中引入了文本拼写特征以及字符级单词特征表示，以提高模型对Twitter文本的实体检测效果。文献[23]提出了一种注意力语义增强模块，以解决神经网络文本上下文信息不足所导致的数据稀疏性问题。该模块引入相似词汇作为语义增强信息，通过相似度为其赋予注意力权重，然后使用门控机制将语义增强信息与原始词汇特征进行融合，并使用此融合特征完成实体检测计算。

鉴于中文语言的高度复杂性以及中文社交网络平台所提供的大量社交数据，中文实体检测任务同样备受关注并且引起了众多学者的广泛兴趣。文献[24]针对社交网络中文标注数据稀缺的问题，提出一种综合应用跨域学习和半监督学习的统一模型，分别使用跨域学习模块和半监督学习模块来学习“域外”数据和“域内”未标注数据，从而有效提升实体检测效果。文献[25]提出了一种融入自注意力机制的中文社交网络实体检测模型，在BiLSTM层与CRF层之间加入了自注意力机制，使模型可以更好地理解社交网络文本结构，并充分利用上下文信息。

近年来，图文结合发文也是一种应用广泛的信息发布方式。该方式产生了大量基于图像的视觉模态信息。对于异常实体检测而言，这些信息不仅可以提供丰富的上下文信息，还可以帮助模型解决常见的实体歧义性问题。文献[26]首次定义了多模态社交网络实体检测任务，使用VGGNet（visual geometry group network）[27]将图片划分为49个区域以表示视觉模态信息，提出了一种自适应注意力网络结构，将视觉模态信息引入实体检测的计算过程中，由相关的视觉图像区域丰富或补充文本语义。此外，该研究所构建的大规模多模态实体检测数据集也为后续相关研究提供了数据基础。文献[28]提出了首个基于Transformer[29]的多模态实体检测模型以解决各模态之间交互不充分等问题。该研究将Transformer模型与跨模态注意力机制相结合，创新性地提出了一个可以同时生成词汇感知视觉特征和视觉感知词汇特征的多模态交互模型。同时为了减少视觉信息带来的偏差，作者还在模型中引入了一个实体边界检测器以校正最终的识别结果。为了充分利用图像中的对象信息，从而避免整体图像特征在实体分类的过程产生误差，文献[30]使用Mask RCNN（region-based convolutional neural network）[31]提取图像中的视觉对象（如人物、奖杯等），并提出了一种密集型共同注意力机制将对象级视觉特征与文本特征相结合，从而实现更细粒度的跨模态交互。基于上述思路，文献[32]提出了一种基于FLAT（flat-lattice transformer）[33]结构的多模态实体检测模型。该模型使用平面网格结构以及相对位置编码为不同模态之间提供了新的交互方式，使得对象级视觉特征能够被纳入Transformer模型的计算过程中，为多模态实体检测提供了新的研究思路和解决方案，具有重要的理论研究价值和实际应用价值。

综上所述，社交网络实体检测相关研究目前已经进入了相对成熟的阶段，学者们从文本自身特征挖掘和综合利用相关多模态信息两个角度出发开展工作，以提高实体检测的准确率。这些工作为社交网络实体异常检测的进一步发展和应用提供了坚实的理论基础和工程指导。

2.2　虚假信息检测

作为异常传播的主要内容，虚假信息一直广泛存在于社交网络中。社交媒体上虚假信息的大量传播早已成为全球性问题，严重影响舆论走向并威胁社会稳定[6]。虚假信息更能激发人们的恐惧、厌恶和惊讶等情绪[14]。

因此，虚假信息检测一直是异常传播中最为重

要的研究领域。随着ChatGPT等生成模型技术的不断成熟，在线社交网络中将会出现越来越多的AI生成式虚假信息，针对虚假信息的检测任务更加艰巨。作为虚假信息的主要形式，本节将着重介绍谣言检测和假新闻检测两个当前热门的研究方向。表2总结了虚假信息检测领域的常用数据集[34-40]。

1) 谣言检测

针对谣言的定义，学术界一直没有统一的观点，早期大多数文献将谣言定义为“正在流通中未经证实且间接相关的信息声明”[41]。文献[42]将谣言归类定义为“在发布时其真实性尚未得到验证的流传信息”，同时将谣言分为突发出现的新谣言和一直存在的谣言两类。总体而言，过往的谣言检测研究主要分为基于传统的机器学习方法和基于深度学习方法。表3总结了6类典型的谣言检测方法[1,43-47]。

早期的谣言检测研究以特征工程为主，通过收集大量手工制作的特征，以训练有效的监督分类器。文献[43]使用包括基于消息、用户、主题和传播的四大类特征研究一种自动方法来评估Twitter上的信息可信度，并且后来被用作大多数相关工作的主要参考。文献[44]提供了一组新的语言、结构和时间特征，表明时间特征在虚假信息检测中是有效的。文献[48]系统总结了谣言检测领域的常用特征。文献[49]第一次提出在新浪微博平台上对谣言进行分析与检测研究，并引入了用于新浪微博的客户端程序和位置特征，实现对新浪微博上的谣言分类。

上述方法都是针对历史谣言信息进行检测，文献[50]第一次提出针对Twitter的实时谣言检测算法，将谣言识别为可能包含一个或多个相互冲突的事件，通过将人群的矛盾视为他们对真实性的辩论来提出验证特征。针对新出现的谣言，文献[51]使用“不真实”“未经证实”或“揭穿”等提示术语进行早期谣言检测，以发现质疑和被否认的推文。

与基于特征工程的方法不同，近些年学者们提出了许多基于神经网络的方法来实现谣言检测任务。文献[52]第一次提出使用卷积神经网络（CNN, convolutional neural network）从相关帖子的

表2 虚假信息检测领域的常用数据集

类型	数据集	文献	文本信息	用户信息	时间信息	传播信息	描述
谣言检测	Twitter	文献[34]	√	√	√	—	包含1 101 985条推文所参与的992个事件，其中有498个谣言事件，494个非谣言事件
	新浪微博	文献[34]	√	√	√	—	包含3 805 656条推文所参与的4 664个事件，其中有2 313个谣言事件，2 351个非谣言事件
	PHEME	文献[35]	√	√	√	√	包含9个事件对应的6 425个声明，其中有2 402个谣言声明，4 023个非谣言声明
	Twitter15	文献[36]	√	√	√	√	包含173 487个用户所参与的818个事件，标签分为4类，分别为非谣言、虚假谣言、真实谣言和未经证实的言论
	Twitter16	文献[36]	√	√	√	√	包含276 663个用户所参与的1 490个事件，标签分为4类
假新闻检测	BuzzFeedNews	文献[37]	√	—	—	—	包含2 282条来自Facebook的新闻数据
	Liar	文献[38]	√	—	—	—	包含12 800条来自Politifact网站上的简短陈述
	CRENDBANK	文献[39]	√	√	—	—	包含6 000多万条推文所参与的1 049个事件
	FakeNewsNet	文献[40]	√	√	√	√	包含23 196条来自PolitiFact和GossipCop网站的新闻数据

表3 6类典型的谣言检测方法

类别	文献	方法优劣
基于树形结构	文献[1]	充分考虑了谣言传播结构特性，但未能考虑谣言的社会背景和用户特征
基于文本特征	文献[43]	易实现，但需要手动设计特征，容易被谣言发布者逃避检测
基于时间特征	文献[44]	能够捕获谣言传播时间特性，但忽略了其他影响因素
基于图结构	文献[45]	能够捕获谣言传播及社会背景特性，但对数据集敏感，泛化能力弱
结合立场检测	文献[46]	能够更好地捕获谣言上下文属性信息，但未考虑谣言的传播及时间特性
基于多模态	文献[47]	能够捕获多模态特征，但与上述特征间更有效的结合方式还有待研究

文本内容中提取关键特征进行谣言检测。文献[34]首次引入循环神经网络（RNN, recurrent neural network）进行谣言检测任务，基于时间序列自动学习Twitter内容，同时考虑使用TF-IDF（term frequency-inverse document frequency）对词语进行建模，并结合RNN学习谣言的潜在内容特征。为了利用谣言的传播特性，学者们先后尝试了在树形结构和图结构下研究谣言检测任务。文献[1]将社交网络中谣言的传播结构建模为树形结构，并使用GRU（gate recurrent unit）计算树的每个分支序列以实现谣言检测。文献[53]使用Transformer中的多注意力机制来模拟推文之间的远距离交互，从不同会话线程之间的用户评论中获取依赖关系，能够更准确地检测出谣言。文献[54]引入双向图卷积网络（Bi-GCN, bi-directional graph convolutional network）方法，使用自上而下和自下而上的传播方向来模拟谣言的传播和扩散。文献[55]将对话线程表示为无向交互图并提出一种用于谣言分类的分层图注意网络，该网络结合社会背景特征学习所有评论推文的表示，并从语义上推断目标声明的帖子是否为谣言。文献[56]通过学习转发序列来整合复杂的语义信息，研究使用图注意力网络来捕获由社交网络上的帖子、评论和相关用户构建的异构图中的全局语义信息。文献[45]提出了一种双动态图卷积网络（DDGCN, dual dynamic graph convolutional network），对空间结构、时间结构、外部知识和文本信息进行建模，使用两个耦合的动态GCN来捕获传播中的多视图信息。

近几年，随着社交信息的不断丰富和深度学习技术的不断发展，谣言检测领域出现了许多新的方向。文献[46]通过关联包含相同高频词的帖子来构建层次异构图以促进特征跨主题传播，并将立场和谣言检测联合制定为多阶段分类任务。文献[47]提出一种多模态特征增强注意力网络（MFAN, multi-modal feature-enhanced attention network），这是首次尝试将文本、视觉和社交图特征集成在一个统一的框架中。该框架考虑了不同模态之间的互补关系和对齐关系，以实现更好的融合。

2) 假新闻检测

作为虚假信息的另一种表现形式，假新闻现在被视为对民主、新闻业和言论自由的最大威胁之一[57]。假新闻与谣言的主要区别有如下两个方面：谣言是未经证实的信息，无法知道其真实性，而假新闻一定是虚假的信息；谣言更多时候是任何账号都可以自由发布的声明言论，而假新闻通常是社交媒体或公共人物发布的。

早期基于新闻文本内容的方法主要是对假新闻和真实新闻的内容差异性特征进行分析，提出诸如标点符号数量[58]、负面词比例[59]等手工特征以识别假新闻。文献[60]研究了情感信号在假新闻检测中的作用，并提出一种LSTM模型结合新闻文本中提取的情感信号，以区分新闻的真实性。但假新闻发布者能以很低的成本躲避这些文本内容特征的检测，所以这些基于文本内容特征的方法很难实现对假新闻的长期检测。

基于元数据的假新闻检测方法是通过挖掘文本的上下文多重特征以实现更有效的假新闻检测，比如评论[9,61]、用户档案[62]、传播结构[63]等辅助信息。文献[9]利用新闻内容和用户评论来联合捕获可疑的句子及评论，提升了假新闻检测的可解释性。文献[61]提出了双重情感特征来表示发布者及评论者的双重情感以及它们之间的关系，该方法可以直接接入其他假新闻检测模型。大多数现有的假新闻检测方法侧重于挖掘新闻内容或其上下文信息，而忽略了用户的重要性。文献[62]利用用户档案捕获用户的偏好进行假新闻检测。除了上述方法外，通过捕获传播结构也能够很好地实现假新闻检测。文献[63]通过全局和局部用户的传播行为，在传播的早期捕捉真假新闻的差异。

为了增加新闻的可读性和可信性，新闻发布者更倾向使用图像等多媒体形式发布假新闻，如何充分利用假新闻图像的固有特征是假新闻检测的一个重要但具有挑战性的问题。文献[64]首次在新闻验证任务中系统地探索了有关的图像特征，提出了几种视觉和统计特征来检测假新闻。文献[65]提出一种多域视觉神经网络（MVNN, multi-domain visual neural network）来融合频域和像素域的视觉信息，并利用注意力机制动态融合频域和像素域的特征表示。为了学习对新数据操纵敏感的通用特征和防止对真实图像的误报，文献[66]提出MVSS-Net（multi-view multi-scale supervision）网络，通过多视图特征学习和多尺度监督有效解决上述问题并实现了像素级和图像级的假新闻检测。

上述方法均是从单一模态角度进行假新闻检

测。针对多模态假新闻检测，上述方法都无法检测跨模态的相关性。因此，结合文本、图像等内容形式的多模态假新闻检测技术成为近几年非常热门的研究方向。为了识别新出现事件的假新闻，文献[67]提出了事件对抗神经网络（EANN, event adversarial neural network）的端到端框架，包含多模态特征提取器、假新闻检测器和事件鉴别器。通过挖掘事件的不变特征，有利于实现对新出现事件的假新闻检测。文献[68]提出了一种端到端网络来学习多模态信息的共享表示，即多模态变分自动编码器（MVAE, multimodal variational autoencoder），使用双模态变分自动编码器与二元分类器相结合来执行假新闻检测任务。利用与文本不相关的图像来吸引大家的注意力是假新闻的欺骗策略之一。根据文本、图像间的“不匹配”来识别新闻文章的虚假性是多模态假新闻检测的一种常见方式。文献[69]提出了一种相似性感知的假新闻检测方法。该方法分别提取文本和视觉特征以进行新闻表示，同时进一步研究跨模态提取的特征之间的关系，并利用新闻文本和视觉信息的这种表示及其关系共同学习用于假新闻的检测。文献[70]从信息论的角度提出了一种基于歧义感知的多模态假新闻检测方法。该方法包括一个跨模态对齐模块，用于将异构单模态特征转换为共享语义空间；一个跨模态歧义学习模块，用于估计不同模态之间的歧义；一个跨模态融合模块，用于捕获跨模态的相关性。

总体而言，针对假新闻检测的研究主要分为基于文本内容的方法、基于元数据的方法、基于图像的方法和基于多模态的方法。前3类方法均是从单一模态和社会背景等角度进行假新闻检测，无法利用跨模态的相关性。面对当前日益增多的图像、视频等更加丰富的数据形式，多模态假新闻检测逐渐成为主流方法。同时，由于新闻是一种自带知识属性的信息形式，利用知识图谱等外部知识进行假新闻检测的方法也取得了不错的效果[71-72]。

3 面向社交网络的异常用户

3.1 用户影响力分析

在异常信息传播扩散期，除了恶意传播者以及虚假账号之外，恶意推手所雇佣的高影响力用户也起到了至关重要的作用。如果不能及时发现传播链条中的高影响力用户，会使异常信息得以快速传播，产生以假乱真、混淆视听的恶劣影响，致使国家安全存有很大的潜在风险。因此，异常用户影响力分析可以对传播链条进行全方位的预判，为管理者提供决策建议和科学路径，使其能准确遏制异常信息，非常重要且具有必要性。

经过多年研究，社交网络用户影响力计算方法可以划分为3类：基于网络拓扑结构的方法、基于用户特征的方法和基于深度学习的方法，上述3类方法的详细信息如表4所示。

基于网络拓扑结构的方法主要侧重于社交网络中用户之间的连接关系，认为影响力的传播是由用户之间的关系网络所决定，具体计算方法主要包括度中心度[73]、PageRank[74]、Jaccard相似度[75]、高阶证据中心度[76]等。该方法依赖于用户之间的连接关系，易于理解，并且适用于大规模社交网络。但其仅基于网络拓扑结构，无法考虑用户自身的特征和行为，可能会对某些用户的影响力产生错误判断。

基于用户特征的方法则通过关注用户的账号特征和行为习惯来确定其影响力。这些特征可能包括用户的关注者数量、社交互动频率、发布内容受欢迎程度等。文献[77]面向Github平台数据，从用户

表4 社交网络用户影响力计算方法的详细信息

方法	计算方式	方法优劣
基于网络拓扑结构的方法	度中心度 PageRank Jaccard相似度 高阶证据中心度	容易理解，适用于大规模的社交网络，但没有考虑到用户自身的特征
基于用户特征的方法	粉丝数量 关注行为 转发行为	充分考虑用户个体差异，但特征工程复杂，容易缺乏相关信息
基于深度学习的方法	卷积神经网络 长短期记忆网络 图神经网络	不需要手动设计特征，能捕获复杂的用户关系和非线性影响，但需要较大的计算资源

关注关系、项目受关注程度以及用户活动3个不同角度提取用户特征，随后使用HITS（hyperlink induced topic search）、PageRank和H-index评估用户影响力，并使用波达计数法综合量化评估结果。除了使用关注、转发等显式外部用户特征之外，文献[78]提出利用用户内部因素，包括用户情感和评论可靠性，以衡量用户在特定领域中的影响力，并通过实验发现所提出的用户内部因素可以有效提高评分预测准确性。文献[79]则从用户活动、用户资料、推文特征以及推文互动行为4个角度提取了4组特征，这些特征的综合性优于前期研究，涵盖了用户账号所包含的多方面特征。此外，该研究还引入了时间敏感度特征，考虑了时间因素对影响力的动态影响，从而更加全面地衡量了影响力的变化趋势。上述研究均是对用户影响力进行综合评估，为了进一步识别在特定主题领域中具有显著影响力的用户个体，文献[80]提出了一种基于主题的用户影响力排名模型。该模型通过语义核和情感信息对推文主题和用户特征进行建模，并据此计算出不同主题下的意见领袖。基于用户特征的方法充分考虑到了用户个体差异，可以对不同用户的影响力做出更准确的评估。该类方法可以应用于不同类型的社交网络，根据实际需求添加或调整特征。但该类方法需要复杂的特征工程，增加了模型的复杂度，并且部分用户可能没有足够的特征信息来评估其影响力，尤其对于新用户或不活跃用户而言，此种情况尤其突出。

近年来，基于深度学习的方法在社交网络用户影响力评估方面取得了显著进展。这种方法利用神经网络模型对用户影响力进行建模和预测。文献[81]通过图神经网络对拓扑结构和用户特征信息进行建模，实验结果证明了端到端的深度学习模型的优越性。针对特定主题下高影响力用户检测问题，文献[82]使用语言注意力网络筛选与主题相关的社交内容，然后通过影响力卷积网络模拟社交网络中的影响力传播过程，并输出用户对特定主题的影响力数值。相较于目前已有的主题影响力检测模型，该方法能够更有效地识别在社交网络中涉及罕见主题中出现的高影响力用户。为了降低对标注资源的依赖性，文献[83]提出了首个人机协作影响力预测模型。该模型利用开放式问题进行众包调查，并通过少量标注数据引导模型学习过程，最终通过答案聚合的方式检测高影响力用户。在影响力预测的过程中，用户之间的关系紧密程度被认为是一个重要的影响因素，文献[84]认为先前的大部分研究所使用的账号关注关系并不能准确判断用户之间的关系紧密程度，因此通过交互频率和关注时间长短来计算用户节点间的连接强度，并提出了一种基于图注意网络（GAT, graph attention network）的注意机制和基于图卷积网络（GCN, graph convolutional network）的卷积聚合方法。针对跨组织协作、用户隐私泄露的问题，文献[85]使用联邦学习技术，通过将差分隐私引入本地模型的参数来增强模型聚合过程的隐私性，以保护用户隐私并实现多个本地模型的安全聚合。该模型可以有效平衡模型的预测性能和隐私保护能力。基于深度学习的方法可以自动学习用户影响力的有效特征表示，不需要设计特征，还可以捕捉复杂的用户关系和非线性影响，更适合复杂的社交网络结构。然而，深度学习模型的训练与计算过程对于大规模社交网络而言，需要相当大的计算资源。

异常用户影响力分析能对异常信息在社交网络中的传播链条进行“地毯式”搜索，使其无立足之地，这样极大地控制了舆情的发生。假如有异常传播发生，也可以快速采取各类应对措施，有效控制负面影响，解除社会安全存在的潜在危机，对于控制异常信息的传播具有不可估量的作用。

3.2　恶意传播群体识别

恶意传播群体在国外更多地被称为“垃圾邮件发送者”，这个名词最早来源于垃圾电子邮件传播时期，但与传统的垃圾邮件发送者不同，在社交网络中出现的新型垃圾邮件发送者，其表现形式有舆论操纵、助长人气、宣传垃圾邮件广告、网络钓鱼和恶意软件传播等。

早期的研究人员主要通过构建蜜罐等方式进行数据收集和恶意传播用户检测。文献[86]通过部署社交蜜罐从社交网络社区收集欺骗性配置文件，用于发现在线社交系统中的恶意传播用户。文献[87]提出了最早的基于矩阵分解的社交恶意传播用户检测方法。文献[88]充分利用了消息内容和用户行为以及社交关系信息，通过施加标签信息约束和稀疏性约束对非负矩阵分解进行了改进，以实现恶意传播用户检测。

之后，该领域的大部分研究通过定义并分析内

容特征、用户行为特征等方式，基于机器学习方法进行恶意传播用户检测。文献[89]研究恶意传播用户和普通用户之间的情感差异，用图拉普拉斯算子对情绪信息进行建模，并将其纳入优化公式中，以统一的方式进行检测。文献[90]提出了一个有效的SpamSpotter框架用于区分Facebook上的恶意传播用户与真实用户。文献[91]通过分析恶意传播群体内部的社会关系及其语义信息，提出了一种有效的恶意传播用户推理算法。文献[92]通过将社区特征与元数据、内容和基于交互的特征相结合，建立了一种混合方法来自动发现恶意传播用户。文献[93]定义了粉丝关注比、平均发布微博数、综合质量评价等6个特征用于微博的恶意传播用户识别任务。

由于机器学习方法需要人工标签等限制性规则，既耗时又昂贵，且恶意传播用户可能会改变自己的行为以避免被发现[94]。因此，近几年出现了许多基于图和深度学习的恶意传播用户识别方法。文献[95]将社交网络建模为带时间戳的多关系图，并利用结构特征、序列建模和集体推理以提升恶意传播用户的检测能力。文献[96]提出了一种改进的基于流的信任评估方案，通过迭代推荐解决社交网络中的信任衰减问题。文献[97]开发了一个在有向社交图上结合GCN和MRF（Markov random field）的模型，用于半监督恶意传播用户检测。文献[98]提出了一种社交媒体应用中基于协作神经网络的恶意传播用户检测机制，可以捕获特征空间更全面的表示以提升恶意传播用户检测的准确性。文献[99]提出了一种用于恶意传播用户检测的半监督广泛学习方法，利用少量标记的社交模式和大量未标记的用户信息构建了高精度的恶意传播用户检测模型，同时引入增量学习方法旨在自适应地学习社交特征的变化分布。文献[100]提出了一种基于社交网络中用户对等接受度的无监督恶意传播用户检测方法，其中一个用户对另一个用户的对等接受度是根据两个用户之间多个共享主题的共同兴趣进行计算的。文献[101]为了解决用户真实性模糊问题，提出了一种基于标签平滑的恶意传播用户识别方法，同时引入生成对抗学习，将之前的标签空间转化为分布式形式，提升了识别效率和稳定性。

总体而言，现有的异常传播大多数有恶意传播群体参与，且他们善于逃避基于人工特征的检测手段，导致以往基于特征的方法很难用于检测现在的恶意传播用户。基于图结构挖掘恶意传播用户的文本、元数据及社交行为等多维特征的方式能够有效实现对恶意传播用户的检测，但需要依赖于庞大的标注数据，针对恶意传播用户的数据标注又是极其困难的，所以很多基于半监督、无监督的恶意传播用户检测方法变得更加可行。

3.3 社交机器人检测

在许多社交媒体平台上，用户可能会遇到模拟人类账户和行为的全自动社交媒体账户，这些账户通常被称为社交机器人[102]。近年来，社交机器人早已被用来窃取个人信息和传播错误信息等。因此，社交机器人的检测在异常传播过程中是一项重要的研究工作[103]。表5列出了社交机器人检测领域的相关数据集[8,104-108]。

文献[104]利用关注者和粉丝的数量、推文数量以及创建日期等特征在公开数据集上测试了以往

表5 社交机器人检测领域的相关数据集

数据集	文献	文本信息	用户信息	邻域信息	描述
TwiBot-20	文献[8]	√	√	√	包含229 580个Twitter用户及其发布的33 488 192条推文，其中真实用户5 237个，社交机器人用户6 589个，支持构建图结构
cresci-15	文献[104]	√	√	√	包含5 301个Twitter用户及其发布的2 827 757条推文，其中真实用户1 950个，社交机器人用户3 351个，支持构建图结构
cresci-17	文献[105]	√	√	—	包含14 368个Twitter用户及其发布的6 637 615条推文，其中真实用户3 474个，社交机器人用户10 894个，不支持构建图结构
gilani-17	文献[106]	—	√	—	包含3 062个活跃的Twitter用户，其中真实用户1 758个，社交机器人用户1 304个，不支持构建图结构
midterm-18	文献[107]	—	√	—	包含50 538个参与2018年美国中期选举的Twitter用户，其中真实用户8 092个，社交机器人用户42 446个，不支持构建图结构
TwiBot-22	文献[108]	√	√	√	包含1 000 000个Twitter用户及其发布的86 764 167条推文，其中真实用户860 057个，社交机器人用户139 943个，支持构建图结构

的机器人检测方法，表明基于特征集的分类器相较于原始基于分类规则的算法可以更好地检测社交机器人。BotOrNot 系统通过实验验证随机森林分类器更适合用来评估和检测社交机器人[102]。文献[109]从用户和朋友元数据、推文内容和情绪、网络模式和活动时间序列等类别中提取了 1 000 多个特征，并使用机器学习模型进行检测，最终认为用户元数据和内容特征对检测机器人最为有效。

随着文献[105]提出 Twitter 上社交机器人存在的证据，并公开了用于后续研究的新型社交机器人检测的数据集，社交机器人检测成为热门的研究方向。文献[107]提取关注者数量等原始数据和关注者增长率等衍生特征，提升了社交机器人检测的准确率。文献[110]利用两个在线帖子之间的个人信息相似性作为新机器人检测模型的关键，并首次使用深度上下文词嵌入模型来执行社交媒体机器人检测任务。文献[111]应用 TF-IDF 等技术生成文本特征并引入情感分析特征以进行社交机器人检测任务。文献[112]提出一种基于上下文 LSTM 架构的深度神经网络，该网络利用内容和元数据来检测推文级别的机器人。文献[113]提出了一种注意力感知的深度神经网络模型用于检测社交网络上的社交机器人，使用 BiLSTM 和 CNN 架构联合建模用户的行为、属性、时间和活动信息。

根据最近的一项调查[114]，Twitter 上社交机器人在不断进化，新出现的高级机器人会窃取真实用户的推文并淡化其恶意内容以逃避检测。之前特征的可检测性很容易被社交机器人模仿和逃避，机器人故意与人类进行更多互动，本质上会导致社交机器人的对抗性，即特定信息的影响力增加和故意逃避检测，随着时间的推移，这些特征就会变得无效[115]。为了应对机器人伪装的挑战，基于图的社交机器人检测方法得到了广泛的研究，并且随着包含图信息的基准数据集[8]的出现，在社交机器人检测方面取得了巨大的成功。

文献[116]引入 Graph Hist 用于提取图的潜在局部特征，沿着特征空间的一维横截面将节点分类在一起，并基于该多通道直方图对图进行分类。文献[117]第一次引入基于图卷积神经网络的模型，有效利用 Twitter 账户的图结构和关系实现社交机器人的检测。文献[118]提出一种针对 Twitter 用户的自监督表示学习框架，通过对大量自监督用户进行预训练并针对机器人检测场景进行微调来实现检测任务。文献[119]利用关注关系构建异构图，并利用多模态用户语义和属性信息来避免特征工程，应用关系图卷积网络增强其捕获具有多样化伪装的机器人的能力，解决了现有方法未能解决的社交机器人社区化和伪装性的挑战。后来他们在此基础上应用 Graph Transformer 来更好地自适应聚合来自邻居的信息，并使用语义注意力网络来聚合跨用户和关系的消息，进行异质性感知的 Twitter 机器人检测[120]。文献[121]利用文本图交互和语义一致性来增强 Twitter 机器人检测。文献[122]设计了一种基于 GAN 的联邦知识蒸馏机制，用于在客户端之间有效地传输数据分布的知识，该解决方案实现了跨语言和跨模型机器人检测。以上几种方法的优劣对比如表 6 所示。

在异常传播的扩散阶段，由于人工恶意传播账号的成本日益增加，同时生成式 AI 应用逐渐增多，未来会有越来越多的社交机器人用户出现在社交网络中。针对社交机器人的检测也从以往的基于文本特征、元数据的方式逐渐过渡到基于图模型等方式综合考虑社交机器人的多维信息，并已经取得了更好的检测效果。

4　面向社交网络的异常传播

4.1　流行度预测

与社交网络中的常规信息相比，异常信息往往

表 6　　**社交机器人检测领域代表性方法的优劣对比**

方法	文献	方法优劣
基于元数据的方法	文献[107]	能够捕获用户关注等衍生特征，但容易被逃避检测
	文献[110]	充分考虑用户原始信息间差异性，但未能考虑其行为差异性
基于文本特征的方法	文献[111]	能够捕获用户文本的情感特性，但容易被逃避检测，特征考虑不充分
基于图模型的方法	文献[117]	能够利用关注关系捕获用户社交关系，但考虑的社交关系过于单一化
	文献[120]	能够充分学习多元化关系下用户节点表征，但忽略了社交行为的时间特性
基于生成对抗的方法	文献[122]	实现了跨语言和跨模型的检测，但未考虑多样化实体和关系的数据场景

由恶意推手主导并散布。这些推手通过利用大量网络恶意传播用户和虚假账号，有预谋地分享和转发异常信息，使其在相关社交网络集群中快速传播并高度流行。因此，在异常信息扩散阶段进行流行度预测旨在量化异常信息的传播趋势和潜在风险，从而提前评估这些信息可能在社交网络中引发的恶劣影响。

鉴于当前社交网络平台所具备的用户范围广、传播速度快和信息规模庞大等特征，异常信息流行度预测任务面临着影响因素不可控、数据噪声干扰以及信息级联传播等重大挑战[123]。

随着研究的不断深入，流行度预测任务的相关研究也逐渐趋于完善。从方法模型的角度来看，当前主流方法主要有3类：基于特征提取的方法，基于点过程的方法和基于深度学习的方法[124]。这些方法通过不同的数学模型有效地对信息传播过程中的关键要素进行建模并预测信息流行趋势。从研究资源的角度来看，学者们积极构建了各种类型的数据集以支撑其研究工作，并为进一步的研究提供了坚实的数据基础。表7列举了当前主要公开数据集的相关信息，这些数据集覆盖了多种不同的主题和领域，为相关研究人员提供多样化的参考[125-132]。

在早期研究中，学者们通常使用基于特征提取的方法对信息流行度进行预测。该方法基于对信息相关内容的人工分析，从中提取可能对流行度产生影响的特征，如用户特征[133]、内容特征[134]、时序特征[135]等。随后使用机器学习模型对这些特征进行建模以预测信息未来的流行趋势。基于特征提取的方法可以充分反映不同类型特征对预测性能的影响，但是该方法过于依赖所提取特征的质量，并且在特征提取的过程中需要涉及大量专业领域知识，会显著降低模型的泛化能力。

基于点过程的方法是使用点过程模型对流行度的累积过程进行建模分析。点过程是一种用于描述事件在时空上随机分布的数学模型。文献[126]提出了一种基于强化泊松过程的流行度预测模型，融合了信息吸引力、时间松弛函数以及“富者更富”机制3种流行度演化的关键因素。由于其在时间松弛函数的选择方面更加灵活，该模型具有较强的泛化性，可以适用于多种不同领域的流行度预测任务。除了基于泊松过程的方法之外，基于霍克斯过程的方法也被视为具有代表性的方法。文献[136]对信息流行度演化过程中的内在因素和外在因素之间的关联性进行了深入分析，提出了霍克斯强度过程模型。该模型有效降低了模型的数据依赖以及计算复杂度，并大幅降低了预测平均误差。基于点过程的方法具有较强的可解释性，但受限于建模时所采用的强假设，该类方法的预测能力受到了显著制约。

随着深度学习方法在计算机视觉、自然语言处理、语音识别等多个领域的广泛应用，学者们开始通过对端到端的深度神经网络模型对信息内容、账号特征和传播结构等关键要素进行建模从而实现对信息流行度的预测。文献[137]提出了第一个基于深度学习的流行度预测模型，该模型通过随机游走方法表示信息级联图，并使用GRU网络和注意力机制推测级联规模。在信息传播过程中，不同用户节点的转发时间也会对流行趋势产生重要影响，短时间内获得大量转发的信息可能会具有更高的流行性。为了有效地将上述时间特征与深度学习模型相结合，文献[138]提出了一种显式时间嵌入方法，通过将时间特征转化为时间嵌入向量并将其集成到级联节点特征中，使得深度学习模型能够更加准确

表7 流行度预测任务的公开数据集的相关信息

数据集	来源	年份	介绍
WISE2012	文献[125]	2013年	2011年7月1日—31日发布的1 660万条微博
APS	文献[126]	2014年	1893年—2009年463 348篇科研论文及引用数量
新浪微博	文献[127]	2017年	2016年6月1日119 313条微博信息及转发情况
TPIC17	文献[128]	2017年	Flickr平台上发布的68万条社交媒体内容
SMPD	文献[129]	2019年	16个月内7万Flickr用户发布的48.6万条社交媒体内容
Twitter	文献[130]	2020年	2017年12月23日—2018年3月19日发布的15.2亿推文
ATNNDataSet	文献[131]	2021年	23 107 452条商品信息，400万条用户信息和4 000万条交互信息
TikTok	文献[132]	2022年	2021年9月6日—11日抖音平台20 445条导购视频相关记录

地捕捉信息传播过程中的时间依赖。针对社交网络信息跨平台传播所导致的语法结构差异、流行度序列难以对齐等问题，文献[139]提出了一种跨平台流行度预测模型，该模型通过语义关系量化不同社交平台上的信息流行度，并以此预测相关信息在目标社交平台上的流行度。

社交媒体的内容多样性已被广泛认知，因此在流行度预测任务中，除了信息传播的常规特征外，社交信息中常见的其他模态信息也起到了至关重要的作用[140]。文献[141]从用户生成内容中提取视觉特征、文本特征、用户标签、发布日期和内容类别等多种模态特征，随后通过引入注意力机制计算不同模态特征对流行趋势的影响，从而实现信息流行度的预测。文献[140]分别使用基于过滤器的主题模型和文本感知图像注意力机制过滤文本和图像数据中的噪声信息，然后将这两种模态特征与其他传统特征结合，以提升预测性能。以往研究在处理视觉模态和文本模态这两种基本模态特征时，通常采用较为简单的方法，未能充分挖掘其中所蕴含的丰富信息。针对该问题，文献[10]从视觉模态语义表示、图像质量、图像背景信息3个角度挖掘视觉特征，并从文本信息中提取多重语义嵌入特征，通过融合更加丰富的视觉和文本特征，该模型在流行度预测精度方面取得显著的提升。

基于深度学习的流行度预测方法通过自动学习数据中的各种复杂特征表示，显著增强了其泛化能力和预测效果。然而，该方法的可解释性较为有限，且对数据量有较高的需求。

在异常信息传播阶段，流行度预测是不可或缺的核心环节。当前相关研究从特征提取、点过程以及深度学习3个角度构建数学模型来预测信息在社交网络中的传播规模和影响力，通过预测所得到的精准量化结果使得决策者能够及早地认识到异常信息在社交网络传播中的潜在危机，并及时采取针对性措施进行应对。因此，流行度预测在异常信息传播阶段的重要性不可低估。

4.2　异常传播模型

在网络恶意传播用户和社交机器人群体的参与下，异常信息开始广泛进入大众视野并引起社交平台自然用户的广泛讨论和转发。在这一阶段，针对异常传播模型的研究能够有效洞察其传播模式和特征，为后续的传播抑制工作提供科学支撑。

当前研究中广泛应用的信息传播模型主要包括3种：独立级联模型、线性阈值模型和传染病模型[142]。其中，独立级联模型基于信息传播中节点间相互独立影响的假设[143]，而线性阈值模型也仅仅关注于节点的传播阈值和邻居节点的影响[144]，二者均未考虑时间动力学因素对传播过程的影响，存在一定的局限性。因此这两种模型不适于分析复杂的现实社交网络平台中信息传播的规律[43]。

相比之下，传染病模型则通过模拟传染病在人群中的传播机理对信息传播规律进行建模[145]，不仅考虑了信息在社交网络中的扩散过程，还能够有效模拟信息传播的路径和时间依赖关系。文献[146]面向社交网络中常见的多信息传播场景，在SIR（susceptible infected recovered）模型中引入了“犹豫者”作为双重信息竞争中的中性状态。实验结果表明，具有较高稳定传播速率的优势信息将主导双重信息的总体影响。为了进一步探究积极信息和消极信息对传染病模型的影响，文献[147]提出一种基于双层多重网络的传播模型。其中，上层网络代表具有积极信息和消极信息传播竞争的社交网络，下层网络代表传染病传播网络。该研究有效论证了积极信息和消极信息对传染病流行过程的影响，并为理解传染病传播的复杂性提供了新的思路。当前社交网络用户数量飞速增长，导致社交网络中的复杂关系愈发难以被准确表征。针对该问题，文献[148]提出了一种基于用户和信息属性的在线社交超网络信息传播模型。该模型使用超图的超边来表示用户之间的社区关系，深入研究了用户影响力、置信度、兴趣价值和信息时效性对该过程的影响，为后续复杂网络中信息传播机制的研究奠定了基础。

相较于正常信息在社交网络中的传播过程，异常信息的传播机制呈现更为复杂的特征，具体表现为内容复杂多样、波及范围广泛、常规传播模型难以精准预测以及社会危害严重。因此，对异常信息传播规律建模的过程中不仅需要模拟异常信息在社交网络中不同群体之间的传播过程，还需要综合考虑公众舆论、用户个体行为和抑制策略等其他因素对异常信息传播的影响。为了解决上述问题，学者们对传染病模型进行了进一步完善和优化，以期更加准确地描述异常信息在社交网络中的传播过程。文献[11]基于传染病模型的思想，提出了首个谣言传

播模型。该模型创新性地将传播人群划分为了3种：谣言传播人群、谣言免疫人群和无知谣言人群，并深入探讨了谣言传播与传染病传播之间的差异。在此基础上，文献[149]提出了一种名为IDSRI（ignorance discussant spreader remover ignorance）的谣言传播模型。该模型对传播人群进一步划分，在原有3种人群的基础上，引入了“谣言讨论人群”，即仅参与谣言讨论却不对其进行传播的人群。通过与其他模型相比较，研究结果表明“谣言讨论人群”对谣言传播过程具有重要影响。文献[150]则聚焦于多重辟谣机制对谣言传播过程的影响，分别将新闻媒体的辟谣报道和辟谣者发布的辟谣声明定义为外部辟谣行为和内部辟谣行为，并将这种双重辟谣机制与SIR模型相结合。为了探究复杂网络中所引发的图灵不稳定性对信息传播过程的影响，文献[151]提出了一种具有Allee效应的SI传播模型。该研究发现空间不稳定性会导致在不同网络拓扑结构中出现相同类型的图灵模式，并且扩散系数可以显著改变模式。充分考虑到如社交网络平台二次传播等外部因素对谣言传播的影响，文献[152]提出了一种基于非光滑SIR的谣言传播模型，首先运用上下解理论，对该模型非负解的存在性进行了证明；其次计算了基本再生数并讨论了正平衡的存在性；最后对谣言传播平衡和Hopf分岔的稳定性进行了理论分析。除了谣言传播相关研究之外，文献[153]面向负面舆论信息的传播规律展开研究，通过在线负面舆情信息传播网络和线下社交传播网络继承构建传播模型，并通过连续时间马尔可夫链来模拟负面舆情信息传播过程。实验结果表明，该模型在控制负面舆论规模方面表现出较好的效果，为后续负面舆论信息传播的研究提供了科学方法和研究途径。

通过对异常信息的传播规律展开研究，可以进一步洞察其传播模式和特征，从而深化有关部门对此类现象的认知，以更加有效地应对异常传播带来的负面影响，从而维护网络空间的秩序并促进其健康发展。与此同时，异常信息传播建模研究还可为后续的阻断工作奠定坚实基础，为应对此类传播行为提供科学支撑。

5 面向社交网络的传播抑制

传播抑制问题是面向异常传播过程研究的最后一步，如何抑制异常信息的传播也是网络信息管理面临的一项技术挑战[154]。以往的传播抑制方法大致可以分为基于阻断策略的传播抑制和基于澄清策略的传播抑制[155]。

5.1 基于阻断策略的传播抑制

基于阻断策略的传播抑制指的是在传播过程中阻塞（或删除）一组节点或边，以最大程度地减少网络中异常信息的流动。根据定义可知，常见的阻断方法分为节点阻断和边阻断。

1) 节点阻断

节点阻断又称为节点免疫，通过识别并阻断一组关键节点来最大程度地减少网络中异常信息的扩散。最早的节点阻断方法是使用基于度的方法来查找关键顶点进行阻断[12]。文献[156]在此基础上引入主题感知方法，提出了主题感知介数和主题感知度中心性度量的启发式算法，从主题建模的角度解决了通过阻塞有限数量的节点来最小化网络中异常信息的负面影响问题。文献[157]提出在独立级联模型下使用贪心算法发现并屏蔽未受感染的用户，最大限度地减少最终受感染的用户规模。文献[158]提出一种贪心算法来解决不同扩散模型下的影响力最小化问题。

上述结果验证了贪心算法比传统的基于中心性（例如度中心性、介数中心性和PageRank）的启发式算法更有效。但这些方法都只是在静态网络中去寻找阻断节点，忽略了真实场景下社交网络的时效性和变化性。因此，文献[159]提出了一种基于模拟的方法来估计每个时间戳中阻塞的节点数量，使用启发式方法来计算每个节点的免疫能力（相当于阻塞增益）。在每个时间戳中，确定并封锁免疫能力最高的节点，然后更新剩余节点的免疫能力。文献[160]提出了一种具有用户体验的动态谣言影响最小化模型。为每个节点分配一个容忍时间阈值，一旦用户的阻塞时间超过该阈值，网络的效用就会下降。文献[161]提出了一种新的自适应影响阻塞（AIB, adaptive influence blocking）问题，实现了具有可证明的近似保证和错误边界的可扩展算法，在每轮传播中动态决定阻止节点，并显著改善了时间复杂度。同时，文献[162]考虑到群体的回音室效应，研究了在回音室效应作用下，社交网络中私有群体的解散策略，以最大限度地减少异常信息的传播。文献[163]提出了一种基于整数线性规划的启

发式算法，通过阻止建模为线性阈值模型的复杂社交网络中的节点子集（称为阻止者）来最大程度地减少谣言传播，其性能优于贪心算法和基于中心性的启发式方法。

用于节点阻断的静态方法简单且廉价，但可能会不准确，因为它们不直接处理传播模式。另一方面，自适应动态方法可以通过考虑网络中的传播模式来改善阻塞效果，但由于需要监视和跟踪传播模式，因此需要付出更高的计算成本。

2) 边阻断

在节点阻断方法中，目标是删除关键节点以阻塞传播。然而，由于每个节点可能通过许多边连接到其他节点，这可能会删除大量的边，以至于可能彻底改变网络结构。边阻断方法旨在通过识别一组要阻止的关键边来解决此问题，从而最大限度地减少异常信息的传播。

文献[164]在IC扩散模型下定义了污染度最小化问题，提出了一种基于贪婪启发式的谣言拦截近似解决方案，通过阻止链接（即找到要删除的k条边）来最大程度地减少异常信息的传播。文献[165]提出网络矩阵的特征值是对网络中扩散敏感性的度量；根据特征值计算每条边的分数，目标是识别一组边，将其移除从而最小化矩阵的特征值。文献[166]假设阻塞每条边都有成本，定义了预算约束下的问题，然后提出了贪心算法以有效解决传播抑制问题。由于使用扩散模型计算集合影响力的计算时间较长，文献[167]采用live-edge方法提出了一种高效的迭代贪心方法。文献[168]开发一种启发式算法来阻止候选集的k条边，以最小化网络中节点的激活概率之和。节点的激活概率表示该节点受异常节点影响的概率，即该节点对异常节点传播的异常信息的脆弱程度。同时使用贪心算法，迭代选择边际增益最大的边并更新节点的激活概率。文献[169]通过考虑过去的传播轨迹，以数据驱动方式制定影响限制问题，在控制扩散过程的同时使网络结构中的干扰量最小。此外，他们还考虑了预算约束和矩阵约束这两种类型的边缘移除约束。文献[170]首次研究了社交网络中IC模型下针对目标集的谣言拦截问题，提出了一种新的基于逆向最短路径的采样方法，以实现对目标保护最大化问题目标函数的有效估计。

本节总结了基于节点阻断和边阻断的传播抑制方法，虽然基于阻断的方式是一种很有效的传播抑制策略，但采用阻断方式抑制异常传播，没有考虑真实社交网络中的用户体验，而且在一定程度上会对网络结构造成破坏。

5.2　基于澄清策略的传播抑制

基于澄清策略的传播抑制又称为基于事实真相的传播抑制方法，旨在找到一组节点用于传播真实信息，以最大限度减少异常信息的接受及传播。

文献[12]最早提出引入“竞争活动”的概念，以竞争形式限制恶意活动在社交网络中的传播，并证明了该问题是NP难问题。文献[155]提出一种竞争扩散模型，即单向状态转移线性阈值模型（LT1DT, linear threshold model with one direction state transition），用于对同一网络中两种不同类型的竞争信息传播进行建模，并研究基于扩散动力学的新颖启发式方法来解决LT1DT下的谣言抑制问题。文献[171]研究了竞争线性阈值（CLT, competitive linear threshold）模型下社交网络中竞争影响力的传播，设计了一种比以往的贪婪算法快两个数量级的算法，显著提升了算法效率。文献[172]考虑用户的偏见及其社会邻居意见，从一组揭穿者中找到一组好的种子用户，以最大限度地减少错误信息的影响。文献[173]利用在线社交网络的社区结构来静态选择种子节点，不考虑错误信息节点的分布，以便通过简单的一次性计算更快地遏制错误信息。该方法体现了社区结构在社交网络中异常信息遏制方面的关键作用。文献[174]提出使用节点强度衡量网络中节点重要性，并选取初始正种子集，运行时间比贪心算法快3个数量级。

总体而言，由于社交网络是无标度网络，利用阻断方式来抑制异常传播，用户仍然可以传播来自其他链接的异常信息，因此阻断策略并不是一种很好的选择。澄清策略可以在一定程度上缓解或者抵抗异常信息的传播，但仍需考虑其中的成本及时间衰退等问题。表8总结了社交网络中传播抑制领域的代表性方法。

6　研究展望

面向社交网络中的异常传播过程，现已形成了诸如异常信息识别与发现、异常用户及群体检测、异常事件流行度预测、异常传播模型评估和传播抑

表 8 社交网络中传播抑制领域的代表性方法

类型	技术	算法	文献	扩散模型	方法优劣
阻断策略	点阻断	启发式算法	文献[156]	独立级联模型	充分考虑主题相关性对社交网络中异常信息传播的影响，易实现，但考虑因素过于单一
		贪心算法	文献[158]	独立级联模型	性能优于基于中心性（例如度中心性、介数中心性和 PageRank）的启发式算法，但未考虑节点阻断的成本
		启发式算法	文献[163]	线性阈值模型	性能优于贪心算法和基于中心性的启发式算法，但无法适应超大型网络，未考虑多重网络下的谣言抑制问题
	边阻断	贪心算法	文献[170]	独立级联模型	通过阻断最少的边保护目标用户不受异常信息影响，但性能效果一般，异常信息可通过其他链接继续传播
澄清策略		启发式算法	文献[155]	线性阈值模型	提出一种基于扩散动力学的算法，效果优于 PageRank 且运行速度快于贪心算法，但没有考虑影响力传播的时效性
		启发式算法	文献[174]	竞争独立级联模型	通过标签传播为各个社区分配正种子节点预算并利用节点强度选取抑制种子集，但未考虑抑制成本

制等技术，均取得了优秀且稳定的效果。然而，仍存在需进一步研究或探索的问题。

1) 在异常信息方面

对抗性实体识别和跨语言实体识别。随着社交网络迅速发展，各平台对异常实体的检测技术也日趋完善。异常用户为规避社交网络平台的审核机制，采用拆字、造组合词等恶意手段发布异常信息，从而产生大量对抗性异常实体。因此，如何使异常实体识别模型能够精准识别对抗性实体值得进一步研究。与此同时，社交网络平台的用户群体广泛，涵盖各种语言和文化背景，因为不同语言之间存在显著的语法结构差异，这种多语言特性给异常实体识别带来了新的挑战。

虚假信息的快速早期发现。在社交平台上，虚假信息检测的时效性非常重要。由于虚假信息发布者也在不断躲避检测，大多实时在线检测方法最终还需要专家参与，所以如何更好地利用强化学习等方式实现可学习的实时在线检测技术将是一项有意义的研究内容。同时，如何在海量数据中以更高效的方式获取并筛选虚假信息也是值得研究的方向。

2) 在异常用户方面

动态影响力和群体影响力评估。社交用户的影响力从不是一成不变的，可能会随着一些动态因素的影响而不停改变，例如时间、受众定位、内容质量等。如何在用户影响力分析的过程中引入上述动态因素，值得进一步研究。同时，用户影响力不局限于单个用户，在异常传播的场景下，往往涉及多个异常用户群体的参与。因此，从用户群体的内在特征、群体演化过程等多重维度进行群体影响力分析也是一项重要的研究。

深度伪造的异常用户。随着 GPT-3/4 等大模型技术和 Diffusion Model 等图像生成模型技术的快速发展，深度伪造技术很可能在社交网络中被广泛滥用，最终将导致互联网中出现大批量虚假信息及社交机器人账号以实现各种不法目的。因此，开发反深度伪造的技术和机制是未来异常用户和异常信息领域都需要重点关注的研究方向。

3) 在异常传播方面

跨平台异常传播研究。随着社交平台的不断涌现和发展，用户发布的信息可能被转载到多个平台传播，这使得信息传播变得更加复杂，未来的研究可以聚焦于跨平台的信息流行度预测和传播规律建模，通过考虑不同平台之间的联系和各自特性，提高全域网络空间中信息传播的分析能力。

平台推荐算法的干预。为了吸引用户，社交平台通常会采用推荐算法，通过分析用户的历史行为和兴趣，来预测其喜好并向其推送相关内容。然而，这种做法可能导致“信息茧房”效应，这对信息传播过程产生了重要的影响。因此，有必要深入研究推荐算法在异常信息传播模型中的作用和影响。

信息之间的相互影响。在社交网络中存在海量信息，这些信息之间可能发生相互作用，并且信息传播过程会受到信息之间相互作用力的影响。因此，有必要进行多信息传播机制的研究。

4) 在传播抑制方面

多重网络下的传播抑制。社交网络通常是耦合的，用户之间通常存在多种连接模式，未来可以考虑在重叠社区或高度动态社区结构的网络上进一步研究异常传播抑制最大化问题，还将尝试设计一种

自适应的种子节点选择算法来进行谣言抑制。此外，在研究现实中的异常传播抑制问题时，抑制成本也是不可忽视的因素，因为它限制了可以应用的技术领域。无论是基于阻断策略还是澄清策略，都需要考虑时间、激活用户等因素的成本。低成本的异常传播抑制问题始终是不可忽视的。

意外情况干预和多方介入。许多干预策略仅在理想条件下进行测试，在实际应用中可能会遇到许多不可预测的困难。因此，未来的策略不能仅仅关注干预的实验效果，还需要考虑用户的特征以及其他意外情况，例如现实世界中的激活或免疫失败。现有的抑制策略大多是单一方法，这存在一定的局限性。未来可以考虑将异常信息屏蔽和真相传播等多种策略结合的方法，共同用于异常传播抑制。

7　结束语

随着移动互联网以及人工智能技术的飞速发展，社交网络成为全世界信息传播的重要载体。本文围绕社交网络中的异常传播问题展开，总结归纳了异常传播周期的4个阶段，并在前两个阶段以微观视角选取了异常信息和异常用户两个研究领域；在异常传播的后两个阶段，从宏观视角定义了异常传播和传播抑制两个研究领域。同时，分别探讨和分析了上述4个领域的主要任务及其研究进展，并针对不同领域提出了研究展望，为相关领域研究人员提供参考，促进后续研究。面向社交网络中异常传播研究能够帮助洞察异常传播行为的本质规律、掌握其传播特征、制定有效的识别方法，为有效提升对网络空间的全面认知和面向异常的精准响应能力提供基础理论和技术支撑。

参考文献：

[1] MA J, GAO W, WONG K F. Rumor detection on twitter with tree-structured recursive neural networks[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2018: 1980-1989.

[2] LAZER D M J, BAUM M A, BENKLER Y, et al. The science of fake news[J]. Science, 2018, 359(6380): 1094-1096.

[3] HU X, TANG J L, ZHANG Y C, et al. Social spammer detection in microblogging[C]//Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence. New York: ACM Press, 2013: 2633-2639.

[4] CHANDOLA V, BANERJEE A, KUMAR V. Anomaly detection: a survey[J]. ACM Computing Surveys, 2009, 41(3): 15.

[5] LI J, SUN A X, HAN J L, et al. A survey on deep learning for named entity recognition[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(1): 50-70.

[6] GUO B, DING Y S, YAO L N, et al. The future of false information detection on social media: new perspectives and trends[J]. ACM Computing Surveys, 2021, 53(4): 68.

[7] SHUMOVSKAIA V, KAYAALP M, SAYED A H. Identifying opinion influencers over social networks[C]//Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2023: 1-5.

[8] FENG S B, WAN H R, WANG N N, et al. TwiBot-20: a comprehensive twitter bot detection benchmark[C]//Proceedings of the 30th ACM International Conference on Information & Knowledge Management. New York: ACM Press, 2021: 4485-4494.

[9] BAKSHY E, ROSENN I, MARLOW C, et al. The role of social networks in information diffusion[C]//Proceedings of the 21st International Conference on World Wide Web. New York: ACM Press, 2012: 519-528.

[10] WU J M, ZHAO L M, LI D W, et al. Deeply exploit visual and language information for social media popularity prediction[C]//Proceedings of the 30th ACM International Conference on Multimedia. New York: ACM Press, 2022: 7045-7049.

[11] DALEY D J, KENDALL D G. Epidemics and rumours[J]. Nature, 1964, 204(4963): 1118.

[12] BUDAK C, AGRAWAL D, EL ABBADI A. Limiting the spread of misinformation in social networks[C]//Proceedings of the 20th International Conference on World Wide Web. New York: ACM Press, 2011: 665-674.

[13] ZAREIE A, SAKELLARIOU R. Minimizing the spread of misinformation in online social networks: a survey[J]. Journal of Network and Computer Applications, 2021, 186: 103094.

[14] VOSOUGHI S, ROY D, ARAL S. The spread of true and false news online[J]. Science, 2018, 359(6380): 1146-1151.

[15] PENG N, DREDZE M. Supplementary results for named entityrecog nition on Chinese social media with an updated dataset[R]. 2017.

[16] FININ T, MURNANE W, KARANDIKAR A, et al. Annotating named entities in Twitter data with crowdsourcing[C]//Proceedings of the 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk. New York: ACM Press, 2010: 80-88.

[17] CHITICARIU L, KRISHNAMURTHY R, LI Y Y, et al. Domain adaptation of rule-based annotators for named-entity recognition tasks[C]//Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. New York: ACM Press, 2010: 1002-1012.

[18] LI C L, WENG J S, HE Q, et al. TwiNER: named entity recognition in targeted twitter stream[C]//Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2012: 721-730.

[19] CLAESER D, KENT S, FELSKE D. Multilingual named entity recognition on spanish-english code-switched tweets using support vector machines[C]//Proceedings of the Third Workshop on Computational Approaches to Linguistic Code-Switching. Stroudsburg: Association for Computational Linguistics, 2018: 132-137.

[20] SIKDAR U K, GAMBÄCK B. Feature-rich twitter named entity recognition and classification[C]//Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT). Osaka: The COLING 2016 Organizing Committee, 2016: 164-170.

[21] LIMSOPATHAM N, COLLIER N. Bidirectional LSTM for named entity recognition in twitter messages[C]//Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT). Osaka: The COLING 2016 Organizing Committee, 2016: 145-152.

[22] HUANG Z H, XU W, YU K. Bidirectional LSTM-CRF models for sequence tagging[J]. arXiv Preprint, arXiv: 1508.01991, 2015.

[23] NIE Y Y, TIAN Y H, WAN X, et al. Named entity recognition for social media texts with semantic augmentation[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg: Association for Computational Linguistics, 2020: 1383-1391.

[24] XU J J, HE H F, SUN X, et al. Cross-domain and semisupervised named entity recognition in Chinese social media: a unified model[J]. IEEE/ACM Transactions on Audio, Speech and Language Processing, 2018, 26(11): 2142-2152.

[25] 李明扬, 孔芳. 融入自注意力机制的社交媒体命名实体识别[J]. 清华大学学报(自然科学版), 2019, 59(6): 461-467.
LI M Y, KONG F. Combined self-attention mechanism for named entity recognition in social media[J]. Journal of Tsinghua University (Science and Technology), 2019, 59(6): 461-467.

[26] ZHANG Q, FU J L, LIU X Y, et al. Adaptive co-attention network for named entity recognition in tweets[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1): 5674-5681.

[27] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv Preprint, arXiv: 1409.1556, 2014.

[28] YU J F, JIANG J, YANG L, et al. Improving multimodal named entity recognition via entity span detection with unified multimodal transformer[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2020: 3342-3352.

[29] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. arXiv Preprint, arXiv: 1706.03762, 2017.

[30] WU Z W, ZHENG C M, CAI Y, et al. Multimodal representation with embedded visual guiding objects for named entity recognition in social media posts[C]//Proceedings of the 28th ACM International Conference on Multimedia. New York: ACM Press, 2020: 1038-1046.

[31] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 2980-2988.

[32] LU J Y, ZHANG D X, ZHANG P J. Flat multi-modal interaction transformer for named entity recognition[C]//Proceedings of the 29th International Conference on Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2022: 2055-2064.

[33] LI X N, YAN H, QIU X P, et al. FLAT: Chinese NER using flat-lattice transformer[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2020: 6836-6842.

[34] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks[C]//Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. New York: ACM Press, 2016: 3818-3824.

[35] KOCHKINA E, LIAKATA M, ZUBIAGA A. All-in-one: multi-task learning for rumour verification[J]. arXiv Preprint, arXiv: 1806.03713, 2018.

[36] MA J, GAO W, WONG K F. Detect rumors in microblog posts using propagation structure via kernel learning[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2017: 708-717.

[37] POTTHAST M, KIESEL J, REINARTZ K, et al. A stylometric inquiry into hyperpartisan and fake news[J]. arXiv Preprint, arXiv: 1702.05638, 2017.

[38] WANG W Y. "Liar, liar pants on fire": a new benchmark dataset for fake news detection[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2017: 422-426.

[39] MITRA T, GILBERT E. CREDBANK: a large-scale social media corpus with associated credibility annotations[J]. Proceedings of the International AAAI Conference on Web and Social Media, 2021, 9(1): 258-267.

[40] SHU K, MAHUDESWARAN D, WANG S H, et al. FakeNewsNet: a data repository with news content, social context, and spatiotemporal information for studying fake news on social media[J]. Big Data, 2020, 8(3): 171-188.

[41] DIFONZO N, BORDIA P. Rumor, gossip and urban legends[J]. Diogenes, 2007, 54(1): 19-35.

[42] ZUBIAGA A, AKER A, BONTCHEVA K, et al. Detection and resolution of rumours in social media: a survey[J]. ACM Computing Surveys, 51(2): 1-36.

[43] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on twitter[C]//Proceedings of the 20th International Conference on World Wide Web. New York: ACM Press, 2011: 675-684.

[44] KWON S, CHA M, JUNG K, et al. Prominent features of rumor propagation in online social media[C]//Proceedings of the 2013 IEEE 13th International Conference on Data Mining. Piscataway: IEEE Press, 2013: 1103-1108.

[45] SUN M Z, ZHANG X, ZHENG J Q, et al. DDGCN: dual dynamic graph convolutional networks for rumor detection on social media[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(4): 4611-4619.

[46] LI C, PENG H, LI J X, et al. Joint stance and rumor detection in hierarchical heterogeneous graph[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(6): 2530-2542.

[47] ZHENG J Q, ZHANG X, GUO S C, et al. MFAN: multi-modal feature-enhanced attention networks for rumor detection[C]//Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2022: 2413-2419.

[48] 张志勇, 荆军昌, 李斐, 等. 人工智能视角下的在线社交网络虚假信息检测、传播与控制研究综述[J]. 计算机学报, 2021, 44(11): 2261-2282.
ZHANG Z Y, JING J C, LI F, et al. Survey on fake information detection, propagation and control in online social networks from the perspective of artificial intelligence[J]. Chinese Journal of Computers, 2021, 44(11): 2261-2282.

[49] YANG F, LIU Y, YU X H, et al. Automatic detection of rumor on Sina Weibo[C]//Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics. New York: ACM Press, 2012: 1-7.

[50] LIU X M, NOURBAKHSH A, LI Q Z, et al. Real-time rumor debunk-

ing on twitter[C]//Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. New York: ACM Press, 2015: 1867-1870.

[51] ZHAO Z, RESNICK P, MEI Q Z. Enquiring minds: early detection of rumors in social media from enquiry posts[C]//Proceedings of the 24th International Conference on World Wide Web. New York: ACM Press, 2015: 1395-1405.

[52] YU F, LIU Q, WU S, et al. A convolutional approach for misinformation identification[C]//Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2017: 3901-3907.

[53] KHOO L M S, CHIEU H L, QIAN Z, et al. Interpretable rumor detection in microblogs by attending to user interactions[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(5): 8783-8790.

[54] BIAN T, XIAO X, XU T Y, et al. Rumor detection on social media with Bi-directional graph convolutional networks[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1): 549-556.

[55] LIN H Z, MA J, CHENG M F, et al. Rumor detection on twitter with claim-guided hierarchical graph attention networks[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2021: 10035-10047.

[56] YUAN C Y, MA Q W, ZHOU W, et al. Jointly embedding the local and global relations of heterogeneous graph for rumor detection[C]// Proceedings of the 2019 IEEE International Conference on Data Mining (ICDM). Piscataway: IEEE Press, 2019: 796-805.

[57] ZHOU X, ZAFARANI R. A survey of fake news: fundamental theories, detection methods, and opportunities[J]. ACM Computing Surveys (CSUR), 2020, 53(5): 1-40.

[58] PARIKH S B, ATREY P K. Media-rich fake news detection: a survey[C]// Proceedings of the 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). Piscataway: IEEE Press, 2018: 436-441.

[59] GUO C, CAO J, ZHANG X, et al. Exploiting emotions for fake news detection on social media[J]. arXiv Preprint, arXiv: 1903.01728, 2019.

[60] GIACHANOU A, ROSSO P, CRESTANI F. Leveraging emotional signals for credibility detection[C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2019: 877-880.

[61] ZHANG X Y, CAO J, LI X R, et al. Mining dual emotion for fake news detection[C]//Proceedings of the Web Conference 2021. New York: ACM Press, 2021: 3465-3476.

[62] DOU Y T, SHU K, XIA C Y, et al. User preference-aware fake news detection[C]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2021: 2051-2055.

[63] SUN L, RAO Y, LAN Y Q, et al. HG-SL: jointly learning of global and local user spreading behavior for fake news early detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(4): 5248-5256.

[64] JIN Z W, CAO J, ZHANG Y D, et al. Novel visual and statistical image features for microblogs news verification[J]. IEEE Transactions on Multimedia, 2017, 19(3): 598-608.

[65] QI P, CAO J, YANG T Y, et al. Exploiting multi-domain visual information for fake news detection[C]//Proceedings of the 2019 IEEE International Conference on Data Mining (ICDM). Piscataway: IEEE Press, 2019: 518-527.

[66] CHEN X R, DONG C B, JI J Q, et al. Image manipulation detection by multi-view multi-scale supervision[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2021: 14165-14173.

[67] WANG Y Q, MA F L, JIN Z W, et al. EANN: event adversarial neural networks for multi-modal fake news detection[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2018: 849-857.

[68] KHATTAR D, GOUD J S, GUPTA M, et al. MVAE: multimodal variational autoencoder for fake news detection[C]//Proceedings of the World Wide Web Conference. New York: ACM Press, 2019: 2915-2921.

[69] ZHOU X Y, WU J D, ZAFARANI R. Safe: similarity-aware multi-modal fake news detection[J]. arXiv Preprint, arXiv: 200304981, 2020.

[70] CHEN Y X, LI D S, ZHANG P, et al. Cross-modal ambiguity learning for multimodal fake news detection[C]//Proceedings of the ACM Web Conference 2022. New York: ACM Press, 2022: 2897-2905.

[71] WANG Y Z, QIAN S S, HU J, et al. Fake news detection via knowledge-driven multimodal graph convolutional networks[C]//Proceedings of the 2020 International Conference on Multimedia Retrieval. New York: ACM Press, 2020: 540-547.

[72] DUN Y Q, TU K F, CHEN C, et al. KAN: knowledge-aware attention network for fake news detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(1): 81-89.

[73] FREEMAN L C. Centrality in social networks conceptual clarification[J]. Social Networks, 1978, 1(3): 215-239.

[74] BRIN S, PAGE L. The anatomy of a large-scale hypertextual Web search engine[J]. Computer Networks and ISDN Systems, 1998, 30(1-7): 107-117.

[75] JACCARD P. Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines[J]. Bulletin De La Societe Vaudoise Des Sciences Naturelles, 1901, 37(140): 241-272.

[76] 闫光辉, 张萌, 罗浩, 等. 融合高阶信息的社交网络重要节点识别算法[J]. 通信学报, 2019, 40(10): 109-118.
YAN G H, ZHANG M, LUO H, et al. Identifying vital nodes algorithm in social networks fusing higher-order information[J]. Journal on Communications, 2019, 40(10): 109-118.

[77] HU Y, WANG S S, REN Y Z, et al. User influence analysis for Github developer social networks[J]. Expert Systems with Applications, 2018, 108: 108-118.

[78] ZHAO G S, LEI X J, QIAN X M, et al. Exploring users' internal influence from reviews for social recommendation[J]. IEEE Transactions on Multimedia, 2019, 21(3): 771-781.

[79] REZAIE B, ZAHEDI M, MASHAYEKHI H. Measuring time-sensitive user influence in Twitter[J]. Knowledge and Information Systems, 2020, 62(9): 3481-3508.

[80] BERNA A G. Ranking influencers of social networks by semantic kernels and sentiment information[J]. Expert Systems with Applications, 2021, 171: 114599.

[81] QIU J Z, TANG J, MA H, et al. DeepInf: social influence prediction with deep learning[C]//Proceedings of the 24th ACM SIGKDD Interna-

tional Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2018: 2110-2119.

[82] ZHENG C, ZHANG Q, YOUNG S, et al. On-demand influencer discovery on social media[C]//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. New York: ACM Press, 2020: 2337-2340.

[83] AROUS I, YANG J, KHAYATI M, et al. OpenCrowd: a human-AI collaborative approach for finding social influencers via open-ended answers aggregation[C]//Proceedings of the Web Conference 2020. New York: ACM Press, 2020: 1851-1862.

[84] ZHUANG H W, ZHOU B, XI W, et al. Modeling connection strength in graph neural networks for social influence prediction[C]//Proceedings of the 2021 IEEE Sixth International Conference on Data Science in Cyberspace (DSC). Piscataway: IEEE Press, 2021: 8-15.

[85] SONG L, WANG H B, ZHANG G Y, et al. FedInf: social influence prediction with federated learning[J]. Neurocomputing, 2023, 548(1): 1-11.

[86] LEE K, CAVERLEE J, WEBB S. Uncovering social spammers: social honeypots + machine learning[C]//Proceedings of the 33rd international ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2010: 435-442.

[87] ZHU Y, WANG X, ZHONG E H, et al. Discovering spammers in social networks[J]. Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence, 2012, 26(1): 171-177.

[88] YU D G, CHEN N, JIANG F, et al. Constrained NMF-based semi-supervised learning for social media spammer detection[J]. Knowledge-Based Systems, 2017, 125: 64-73.

[89] HU X, TANG J L, GAO H J, et al. Social spammer detection with sentiment information[C]//Proceedings of the 2014 IEEE International Conference on Data Mining. Piscataway: IEEE Press, 2014: 180-189.

[90] RATHORE S, LOIA V, PARK J H. SpamSpotter: an efficient spammer detection framework based on intelligent decision support system on Facebook[J]. Applied Soft Computing, 2018, 67: 920-932.

[91] YANG C, HARKREADER R, ZHANG J L, et al. Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on twitter[C]//Proceedings of the 21st International Conference on World Wide Web. New York: ACM Press, 2012: 71-80.

[92] FAZIL M, ABULAISH M. A hybrid approach for detecting automated spammers in twitter[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(11): 2707-2719.

[93] 张艳梅, 黄莹莹, 甘世杰, 等. 基于贝叶斯模型的微博网络水军识别算法研究[J]. 通信学报, 2017, 38(1): 44-53.
ZHANG Y M, HUANG Y Y, GAN S J, et al. Weibo spammers' identification algorithm based on Bayesian model[J]. Journal on Communications, 2017, 38(1): 44-53.

[94] CHEN C, ZHANG J, XIE Y, et al. A performance evaluation of machine learning-based streaming Spam tweets detection[J]. IEEE Transactions on Computational Social Systems, 2015, 2(3): 65-76.

[95] FAKHRAEI S, FOULDS J, SHASHANKA M, et al. Collective spammer detection in evolving multi-relational social networks[C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2015: 1769-1778.

[96] JIANG W J, WU J, LI F, et al. Trust evaluation in online social networks using generalized network flow[J]. IEEE Transactions on Computers, 2016, 65(3): 952-963.

[97] WU Y J, LIAN D F, XU Y H, et al. Graph convolutional networks with Markov random field reasoning for social spammer detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1): 1054-1061.

[98] GUO Z W, SHEN Y, BASHIR A K, et al. Robust spammer detection using collaborative neural network in Internet-of-things applications[J]. IEEE Internet of Things Journal, 2021, 8(12): 9549-9558.

[99] QIU T, LIU X Z, ZHOU X B, et al. An adaptive social spammer detection model with semi-supervised broad learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(10): 4622-4635.

[100] KOGGALAHEWA D, XU Y, FOO E. An unsupervised method for social network spammer detection based on user information interests[J]. Journal of Big Data, 2022, 9(1): 1-35.

[101] GUO Z W, YU K P, JOLFAEI A, et al. Fuz-spam: label smoothing-based fuzzy detection of spammers in Internet of things[J]. IEEE Transactions on Fuzzy Systems, 2022, 30(11): 4543-4554.

[102] 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. New York: ACM Press, 2016: 273-274.

[103] FERRARA E, VAROL O, DAVIS C, et al. The rise of social bots[J]. Communications of the ACM, 2016, 59(7): 96-104.

[104] CRESCI S, DI PIETRO R, PETROCCHI M, et al. Fame for sale: efficient detection of fake Twitter followers[J]. Decision Support Systems, 2015, 80: 56-71.

[105] CRESCI S, DI PIETRO R, PETROCCHI M, et al. The paradigm-shift of social spambots: evidence, theories, and tools for the arms race[C]//Proceedings of the 26th International Conference on World Wide Web Companion. New York: ACM Press, 2017: 963-972.

[106] GILANI Z, FARAHBAKHSH R, TYSON G, et al. Of bots and humans (on Twitter)[C]//Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. New York: ACM Press, 2017: 349-354.

[107] YANG K C, VAROL O, HUI P M, et al. Scalable and generalizable social bot detection through data selection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1): 1096-1103.

[108] FENG S B, TAN Z X, WAN H R, et al. TwiBot-22: towards graph-based twitter bot detection[J]. Advances in Neural Information Processing Systems, 2022, 35: 35254-35269.

[109] VAROL O, FERRARA E, DAVIS C, et al. Online human-bot interactions: detection, estimation, and characterization[J]. Proceedings of the International AAAI Conference on Web and Social Media, 2017, 11(1): 280-289.

[110] HEIDARI M, JONES J H, UZUNER O. Deep contextualized word embedding for text-based online user profiling to detect social bots on twitter[C]//Proceedings of the 2020 International Conference on Data Mining Workshops (ICDMW). Piscataway: IEEE Press, 2020: 480-487.

[111] RODRIGUES A P, FERNANDES R, A A, et al. Real-time twitter spam detection and sentiment analysis using machine learning and deep learning techniques[J]. Computational Intelligence and Neuroscience, 2022, 2022: 5211949.

[112] KUDUGUNTA S, FERRARA E. Deep neural networks for bot detection[J]. Information Sciences, 2018, 467: 312-322.

[113] FAZIL M, SAH A K, ABULAISH M. DeepSBD: a deep neural network model with attention mechanism for SocialBot detection[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 4211-4223.

[114] CRESCI S. A decade of social bot detection[J]. Communications of the ACM, 2020, 63(10): 72-83.

[115] LE T, TRAN-THANH L, LEE D. Socialbots on fire: modeling adversarial behaviors of socialbots via multi-agent hierarchical reinforcement learning[C]//Proceedings of the ACM Web Conference 2022. New York: ACM Press, 2022: 545-554.

[116] MAGELINSKI T, BESKOW D, CARLEY K M. Graph-hist: graph classification from latent feature histograms with application to bot detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(4): 5134-5141.

[117] ALI ALHOSSEINI S, BIN TAREAF R, NAJAFI P, et al. Detect me if you can: Spam bot detection using inductive representation learning[C]//Proceedings of the 2019 World Wide Web Conference. New York: ACM Press, 2019: 148-153.

[118] FENG S B, WAN H R, WANG N N, et al. SATAR: a self-supervised approach to twitter account representation learning and its application in bot detection[C]//Proceedings of the 30th ACM International Conference on Information & Knowledge Management. New York: ACM Press, 2021: 3808-3817.

[119] FENG S B, WAN H R, WANG N N, et al. BotRGCN: twitter bot detection with relational graph convolutional networks[C]//Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. New York: ACM Press, 2021: 236-239.

[120] FENG S B, TAN Z X, LI R, et al. Heterogeneity-aware twitter bot detection with relational graph transformers[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(4): 3977-3985.

[121] LEI Z Y, WAN H R, ZHANG W Q, et al. BIC: twitter bot detection with text-graph interaction and semantic consistency[J]. arXiv Preprint, arXiv: 2208.08320, 2022.

[122] YANG Y G, YANG R Y, PENG H, et al. FedACK: federated adversarial contrastive knowledge distillation for cross-lingual and cross-model social bot detection[C]//Proceedings of the ACM Web Conference 2023. New York: ACM Press, 2023: 1314-1323.

[123] MARTIN T, HOFMAN J M, SHARMA A, et al. Exploring limits to prediction in complex social systems[C]//Proceedings of the 25th International Conference on World Wide Web. New York: ACM Press, 2016: 683-694.

[124] 曹綺, 沈华伟, 高金华, 等. 基于深度学习的流行度预测研究综述[J]. 中文信息学报, 2021, 35(2): 1-18, 32.
CAO Q, SHEN H W, GAO J H, et al. Survey on deep learning based popularity prediction[J]. Journal of Chinese Information Processing, 2021, 35(2): 1-18, 32.

[125] BAO P, SHEN H W, HUANG J M, et al. Popularity prediction in microblogging network: a case study on sina weibo[C]//Proceedings of the 22nd International Conference on World Wide Web. New York: ACM Press, 2013: 177-178.

[126] SHEN H W, WANG D S, SONG C M, et al. Modeling and predicting popularity dynamics via reinforced Poisson processes[C]//Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2014: 291-297.

[127] CAO Q, SHEN H W, CEN K T, et al. DeepHawkes: bridging the gap between prediction and understanding of information cascades[C]//Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. New York: ACM Press, 2017: 1149-1158.

[128] WU B, CHENG W H, ZHANG Y D, et al. Sequential prediction of social media popularity with deep temporal context networks[C]//Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2017: 3062-3068.

[129] WU B, CHENG W H, LIU P Y, et al. SMP challenge: an overview of social media prediction challenge 2019[C]//Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM Press, 2019: 2667-2671.

[130] WANG C, GONG W Z, GAO X F, et al. An attention-based neural model for popularity prediction in social service[C]//Proceedings of the 2020 IEEE International Conference on Web Services (ICWS). Piscataway: IEEE Press, 2020: 516-523.

[131] XIN S, LI Z, ZOU P C, et al. ATNN: adversarial two-tower neural network for new item's popularity prediction in E-commerce[C]//Proceedings of the 2021 IEEE 37th International Conference on Data Engineering (ICDE). Piscataway: IEEE Press, 2021: 2499-2510.

[132] OU N R, YU L, LI H Y, et al. MTAF: shopping guide micro-videos popularity prediction using multimodal and temporal attention fusion approach[C]//Proceedings of the 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2022: 4543-4547.

[133] BAKSHY E, HOFMAN J M, MASON W A, et al. Everyone's an influencer: quantifying influence on twitter[C]//Proceedings of the Fourth ACM International Conference on Web Search and Data Mining. New York: ACM Press, 2011: 65-74.

[134] TSUR O, RAPPOPORT A. What's in a hashtag? Content based prediction of the spread of ideas in microblogging communities[C]//Proceedings of the Fifth ACM International Conference on Web Search and Data Mining. New York: ACM Press, 2012: 643-652.

[135] GAO S, MA J, CHEN Z M. Effective and effortless features for popularity prediction in microblogging network[C]//Proceedings of the 23rd International Conference on World Wide Web. New York: ACM Press, 2014: 269-270.

[136] RIZOIU M A, XIE L X, SANNER S, et al. Expecting to be HIP: hawkes intensity processes for social media popularity[C]//Proceedings of the 26th International Conference on World Wide Web. New York: ACM Press, 2017: 735-744.

[137] LI C, MA J Q, GUO X X, et al. DeepCas: an end-to-end predictor of information cascades[C]//Proceedings of the 26th International Conference on World Wide Web. New York: ACM Press, 2017: 577-586.

[138] SUN X G, ZHOU J Y, LIU L, et al. Explicit time embedding based cascade attention network for information popularity prediction[J]. Information Processing & Management, 2023, 60(3): 103278.

[139] GAO X F, XU W Y, ZHANG Z X, et al. Cross-platform event popularity analysis via dynamic time warping and neural prediction[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(2): 1337-1350.

[140] QIAN Y, XU W, LIU X, et al. Popularity prediction for marketer-generated content: a text-guided attention neural network for multi-modal feature fusion[J]. Information Processing & Management,

2022, 59: 102984.

[141] XU K L, LIN Z M, ZHAO J Q, et al. Multimodal deep learning for social media popularity prediction with attention mechanism[C]//Proceedings of the 28th ACM International Conference on Multimedia. New York: ACM Press, 2020: 4580-4584.

[142] CHANG B, XU T, LIU Q, et al. Study on information diffusion analysis in social networks and its applications[J]. International Journal of Automation and Computing, 2018, 15(4): 377-401.

[143] GOLDENBERG J, LIBAI B. Using complex systems analysis to advance marketing theory development: modeling heterogeneity effects on new product growth through stochastic cellular automata[J]. Academy of Marketing Science Review, 2001, 9(3): 1-18.

[144] LITTLESTONE N. Learning quickly when irrelevant attributes abound: a new linear-threshold algorithm[J]. Machine Learning, 1988, 2(4): 285-318.

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

[146] LIU Y, DIAO S M, ZHU Y X, et al. SHIR competitive information diffusion model for online social media[J]. Physica A: Statistical Mechanics and Its Applications, 2016, 461: 543-553.

[147] WANG Z S, XIA C Y, CHEN Z Q, et al. Epidemic propagation with positive and negative preventive information in multiplex networks[J]. IEEE Transactions on Cybernetics, 2021, 51(3): 1454-1462.

[148] GONG Y C, WANG M, LIANG W, et al. UHIR: an effective information dissemination model of online social hypernetworks based on user and information attributes[J]. Information Sciences, 2023, 644: 119284.

[149] YU Z H, LU S, WANG D, et al. Modeling and analysis of rumor propagation in social networks[J]. Information Sciences: an International Journal, 2021, 580: 857-873.

[150] GUO H M, YAN X F. Dynamic modeling and simulation of rumor propagation based on the double refutation mechanism[J]. Information Sciences: an International Journal, 2023, 630: 385-402.

[151] HE L, ZHU L H, ZHANG Z D. Turing instability induced by complex networks in a reaction-diffusion information propagation model[J]. Information Sciences: an International Journal, 2021, 578: 762-794.

[152] MA X R, SHEN S L, ZHU L H. Complex dynamic analysis of a reaction-diffusion network information propagation model with nonsmooth control[J]. Information Sciences: an International Journal, 2023, 622: 1141-1161.

[153] LIU X Y, TANG T, HE D B. Double-layer network negative public opinion information propagation modeling based on continuous-time Markov chain[J]. The Computer Journal, 2021, 64(9): 1315-1325.

[154] 曹玖新, 高庆清, 夏蓉清, 等. 社交网络信息传播预测与特定信息抑制[J]. 计算机研究与发展, 2021, 58(7): 1490-1503.
CAO J X, GAO Q Q, XIA R Q, et al. Information propagation prediction and specific information suppression in social networks[J]. Journal of Computer Research and Development, 2021, 58(7): 1490-1503.

[155] YANG L, LI Z W, GIUA A. Containment of rumor spread in complex social networks[J]. Information Sciences: an International Journal, 2020, 506: 113-130.

[156] YAO Q P, SHI R S, ZHOU C, et al. Topic-aware social influence minimization[C]//Proceedings of the 24th International Conference on World Wide Web. New York: ACM Press, 2015: 139-140.

[157] WANG S Z, ZHAO X J, CHEN Y, et al. Negative influence minimizing by blocking nodes in social networks[C]//Proceedings of the 17th AAAI Conference on Late-Breaking Developments in the Field of Artificial Intelligence. New York: ACM Press, 2013: 134-136.

[158] YAN R D, LI D Y, WU W L, et al. Minimizing influence of rumors by blockers on social networks: algorithms and analysis[J]. IEEE Transactions on Network Science and Engineering, 2020, 7(3): 1067-1078.

[159] SONG C G, HSU W, LEE M L. Node immunization over infectious period[C]//Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. New York: ACM Press, 2015: 831-840.

[160] WANG B, CHEN G, FU L Y, et al. DRIMUX: dynamic rumor influence minimization with user experience in social networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(10): 2168-2181.

[161] SHI Q H, WANG C, YE D S, et al. Adaptive influence blocking: minimizing the negative spread by observation-based policies[C]//Proceedings of the 2019 IEEE 35th International Conference on Data Engineering (ICDE). Piscataway: IEEE Press, 2019: 1502-1513.

[162] ZHU J M, NI P K, WANG G Q, et al. Misinformation influence minimization problem based on group disbanded in social networks[J]. Information Sciences: an International Journal, 2021, 572: 1-15.

[163] YANG L, MA Z Y, LI Z W, et al. Rumor containment by blocking nodes in social networks[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2023, 53(7): 3990-4002.

[164] KIMURA M, SAITO K, MOTODA H. Minimizing the spread of contamination by blocking links in a network[C]//Proceedings of the 23rd National Conference on Artificial intelligence. Palo Alto: AAAI Press, 2008: 1175-1180.

[165] TONG H H, PRAKASH B A, ELIASSI-RAD T, et al. Gelling, and melting, large graphs by edge manipulation[C]//Proceedings of the 21st ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2012: 245-254.

[166] KUHLMAN C J, TULI G, SWARUP S, et al. Blocking simple and complex contagion by edge removal[C]//Proceedings of the 2013 IEEE 13th International Conference on Data Mining. Piscataway: IEEE Press, 2013: 399-408.

[167] KHALIL E B, DILKINA B, SONG L. Scalable diffusion-aware optimization of network topology[C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2014: 1226-1235.

[168] YAN R D, LI Y, WU W L, et al. Rumor blocking through online link deletion on social networks[J]. ACM Transactions on Knowledge Discovery from Data, 2019, 13(2): 1-26.

[169] MEDYA S, SILVA A, SINGH A. Approximate algorithms for data-driven influence limitation[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(6): 2641-2652.

[170] GUO J X, LI Y, WU W L. Targeted protection maximization in social networks[J]. IEEE Transactions on Network Science and Engineering, 2019, 7(3): 1645-1655.

[171] HE X R, SONG G J, CHEN W, et al. Influence blocking maximization in social networks under the competitive linear threshold model[C]//Proceedings of the 2012 SIAM International Conference on Data

Mining. Philadelphia: Society for Industrial and Applied Mathematics, 2012: 463-474.

[172] SAXENA A, HSU W, LEE M L, et al. Mitigating misinformation in online social network with top-k debunkers and evolving user opinions[C]//Proceedings of the Web Conference 2020. New York: ACM Press, 2020: 363-370.

[173] GHOSHAL A K, DAS N, DAS S. Influence of community structure on misinformation containment in online social networks[J]. Knowledge-Based Systems, 2021, 213: 106693.

[174] 刘维, 杜宁宁, 陈峻, 等. CIC 模型下基于社区检测的谣言抑制最大化方法[J]. 南京大学学报(自然科学版), 2023, 59(2): 282-294.
LIU W, DU N N, CHEN L, et al. Rumor blocking maximization method based on community detection under the CIC model[J]. Journal of Nanjing University (Natural Science), 2023, 59(2): 282-294.

[作者简介]

卢昆（1996– ），男，江苏徐州人，哈尔滨工业大学博士生，主要研究方向为社交网络分析、数据挖掘等。

张嘉宇（1997– ），男，山西太原人，哈尔滨工业大学博士生，主要研究方向为社交网络分析。

张宏莉（1973– ），女，吉林榆树人，博士，哈尔滨工业大学教授、博士生导师，主要研究方向为社交网络分析、网络与信息安全等。

方滨兴（1960– ），男，江西万年人，博士，中国工程院院士，哈尔滨工业大学教授，主要研究方向为计算机体系结构、计算机网络、信息安全。