

# 科学推文的文本属性特征及其对 科学论文的级联演化趋势分析

曹仁猛<sup>1,2</sup>, 许小可<sup>1,2</sup>, 王贤文<sup>3</sup>

(1. 北京师范大学(珠海校区)计算传播学研究中心, 珠海 519087; 2. 北京师范大学新闻传播学院, 北京 100875; 3. 大连理工大学公共管理学院科学学与科技管理研究所暨 WISE 实验室, 大连 116024)

**摘要** 科学推文是科学论文在社交媒体上的重要传播载体, 揭示科学推文的文本属性特征对科学论文传播效果的影响, 可以帮助科学传播者优化传播策略, 提高科学信息的传播范围, 促进学术交流与公众参与。本文基于 5 万余篇论文和 40 万条科学推文的研究样本, 从推文文本内容、多媒体信息和表情符号 3 个维度, 分析了不同文本属性特征的科学推文的级联演化趋势。研究发现, 将论文亮点的内容作为推文文本, 以及在推文中加入图片、视频和表情符号等视觉元素, 能够显著增强科学论文的传播范围。这种效果不仅在传播初期得到体现, 而且在后续传播中得到进一步增强, 形成“强者愈强”的传播趋势, 即级联马太效应。本文将计算传播视角引入 altmetrics 研究中, 能够帮助研究者更深入、全面地理解科学论文在社交媒体上的传播过程和模式, 从而揭示科学论文传播背后的机制。

**关键词** altmetrics; 级联传播; 科学推文; 传播效果

## Analysis of Textual Attributes of Scientific Tweets and Their Impact on the Cascading Evolution Trends of Scientific Papers

Cao Renmeng<sup>1,2</sup>, Xu Xiaoke<sup>1,2</sup> and Wang Xianwen<sup>3</sup>

(1. Computational Communication Research Center, Beijing Normal University, Zhuhai 519087; 2. School of Journalism and Communication, Beijing Normal University, Beijing 100875; 3. WISE Lab, Institute of Science of Science and S&T Management, School of Public Administration and Policy, Dalian University of Technology, Dalian 116024)

**Abstract:** Scientific tweets are an important medium for the diffusion of scientific papers on social media. The understanding of the impact of textual attributes of scientific tweets on the diffusion effects of scientific papers helps science communicators in optimizing their strategies, broadening the reach of scientific information, and facilitating academic communication and public engagement. In this study, based on a dataset of over 50,000 papers and 400,000 scientific tweets, we explore the influence of various textual attributes on the cascade propagation dynamics of scientific tweets across three dimensions: tweet content, multimedia elements, and emojis. The results indicate that the use of the highlights of the paper as tweet content, along with incorporation of visual elements such as images, videos, and emojis into the tweets, can significantly enhance the diffusion scope of scientific papers. This effect not only is evident in the initial stages of propagation but also becomes more pronounced in subsequent stages, resulting in a “rich-get-richer” trend, i.e., Matthew effect in cascade diffusion. By introducing a computational communication perspective into altmetrics research, this study provides a

收稿日期: 2025-01-01; 修回日期: 2025-03-28

基金项目: 国家自然科学基金项目“科学文献全景大数据下的研究热点及研究前沿探测”(71974029); 中央高校基本科研业务费“科学如何促进技术创新: 影响因素及作用模式研究”(DUT24RW303)。

作者简介: 曹仁猛, 1994 年生, 博士, 讲师, 硕士生导师, 主要研究方向为补充计量学和计算传播学; 许小可, 1979 年生, 博士, 教授, 博士生导师, 主要研究方向为计算传播学和网络科学; 王贤文, 通信作者, 1982 年生, 博士, 教授, 博士生导师, 主要研究方向为科技情报、科学学与科技政策, E-mail: xianwenwang@dlut.edu.cn。

deeper and more comprehensive understanding of the diffusion processes and patterns of scientific papers on social media, and thereby uncovers the mechanisms that drive the diffusion of scientific papers.

**Keywords:** altmetrics; cascade diffusion; scientific tweet; diffusion effects

## 0 引言

随着全民科学素养的提升,科学信息逐渐成为广大社会公众消费的对象。社交媒体作为现代生活获取和传播信息的主要工具,为科学传播提供了便利条件<sup>[1-3]</sup>。科学信息来源于科学研究成果,包含研究过程、实验数据、分析结果和理论框架等内容。这些信息通过科学论文、技术报告、新闻报道、社交媒体等多种渠道传播。然而,科学信息通常具有较高的专业性和复杂性,可能不容易为广大非学术性受众所理解。因此,如何通过简洁、易懂且吸引人的方式将这些科学信息传递给公众,成为现代科学传播面临的重大挑战<sup>[4-6]</sup>。

帖子(post)是社交媒体平台上常见的主要信息发布形式之一,它通常以简短的文字形式呈现,并通过社交媒体平台快速传播。帖子不仅限于文字,还可以包含图片、视频、链接表情符号和话题标签等,这些元素共同增强了信息的表现力和传播效果。在社交媒体传播中,帖子承担着重要的沟通和互动功能,尤其在科学传播领域<sup>[3,7-8]</sup>。帖子为科研人员提供了一种简洁的方式,将复杂的学术内容提炼成容易理解且能够吸引公众注意的核心信息。尽管帖子内容的简洁性是其优势,却也带来了挑战:如何在有限的字符空间内有效传递科学信息,并同时吸引用户的关注、激发其参与,已成为科学传播领域亟待解决的关键问题。

在众多全球性社交媒体平台当中, Twitter以其庞大的活跃用户数量、广泛的学术使用程度、丰富的互动交流功能,为观测科学信息的社交媒体传播贡献了海量数字印记,成为了目前社交媒体计量学最重要的数据来源。推文(tweet)是Twitter平台上的帖子,作为平台的核心信息传播形式,每条推文最多可以包含280个字符。科学推文是指包含科学论文链接的推文,如DOI(digital object identifier)。它通过简洁的文字和多媒体元素,将科学研究成果、学术进展等信息传递给公众,附带的论文链接便于受众进一步阅读。

然而,社交媒体信息的传播并非仅限于首次发帖,其传播过程通常呈现级联效应——信息在社交媒体平台上被用户不断转发、分享和再传播,形成

多轮的连续传播过程<sup>[9-11]</sup>。在这一过程中,每次传播不仅延续了信息的流动,还能触及不同的受众群体,从而扩展信息的传播范围和影响力。目前,级联传播已在多个传播场景中得到广泛研究,包括网络博客<sup>[11]</sup>、电子邮件<sup>[12]</sup>、产品推荐<sup>[13]</sup>以及社交媒体平台<sup>[14-16]</sup>。这些研究揭示了一个重要事实:信息的传播效果不仅依赖于初次传播,还受到后续互动与扩散的影响。

在先前的研究中,我们基于计算传播学视角,首次揭示了科学论文在社交媒体传播中的级联传播现象,并分析了发表在不同期刊层次上的科学论文的级联结构特征<sup>[14]</sup>。本文聚焦于科学推文的文本属性特征,从文本内容、多媒体信息和表情符号3个方面,通过级联传播视角,分析具有不同文本属性特征的科学推文的级联演化趋势。通过这一研究,我们不仅能够深入理解科学推文文本特征如何塑造其传播路径和传播效果,还能为优化科学传播策略、提升科学推文的传播效能提供重要的理论支撑和实践参考。

## 1 相关研究工作

### 1.1 社交媒体信息的文本属性特征

#### (1) 文本内容

社交媒体上的信息通常以短消息的形式呈现,其字符数量受到平台的限制,例如,推文最多为280个字符。这一限制有助于提高内容简洁性,但也可能削弱信息表达的完整性。为此,社交媒体引入了话题标签(#hashtag)、网页链接、图片、视频等多媒体元素,用于增强信息的表现力和传播效果。研究表明,包含话题标签、超链接或嵌入图片和视频的推文更有可能被用户转发<sup>[17-24]</sup>。例如,推文中的视觉特征(如颜色分布、全局场景等)与转发数量存在显著相关性<sup>[25]</sup>,而带有问号或感叹号的推文语调也能够有效提升转发率<sup>[26]</sup>。此外,话题标签的长度和是否使用热门关键词都会显著影响推文的传播效果<sup>[23]</sup>。

#### (2) 情绪化内容

情感化内容在社交媒体上具有更强的吸引力和传播潜力。研究发现,推文中的道德情感词汇能够

显著提升传播效果，而情感强度与转发频率、传播速度呈现正相关关系<sup>[26-30]</sup>。不仅如此，情感的细粒度类别（如愤怒、喜悦、悲伤）也会对传播模式产生不同的影响。例如，负面情感推文通常比正面情感推文获得更多的关注与转发<sup>[31]</sup>，但高唤醒程度的情绪（如敬畏、焦虑和愤怒）比低唤醒程度的情绪（如悲伤）更有利于信息的病毒式传播<sup>[32]</sup>。在具体情感类型中，愤怒情绪比喜悦或悲伤情绪更具传播性，并且这种传播性会随着社交距离的增加而减弱<sup>[28]</sup>。

### （3）主题内容

不同主题的信息传播模式差异显著。政治类信息通常引发更广泛的传播与讨论，其级联率和持续性指标均高于娱乐、运动等其他话题<sup>[33-34]</sup>。此外，政治类信息中的假新闻比其他类别的信息传播速度更快、病毒性更强<sup>[33]</sup>。与之相比，娱乐类信息虽然对用户情感的即时吸引力较强，但其传播过程更多依赖多个核心用户的参与，通过间接转发实现广泛传播<sup>[35]</sup>。还有研究发现，娱乐类信息通常在传播规模、传播速度和病毒性上表现优于公共信息<sup>[36]</sup>。

## 1.2 科学推文的文本属性特征

标签是科学推文中用于关联话题的重要工具。通过添加与研究相关的标签，推文得以融入更广泛的讨论群体，从而扩大信息传播的覆盖范围。研究表明，带有标签的推文通常能吸引更多的关注。例如，有研究分析了 Twitter 上的学术推文，发现当推文附加至少一个标签时，其平均提及次数显著高于不使用标签的推文，表明标签在推动信息传播中具有重要作用<sup>[37]</sup>。

用户提及（@username）为推文提供了直接面向特定用户的交流通道。这种定向的互动形式增强了推文的传播效率和社交属性。一项基于心理学论文的研究显示，提及特定用户的推文，其平均提及次数是未提及用户推文的两倍以上<sup>[38]</sup>。此外，用户提及的使用还被证明有助于提升推文的受欢迎程度，尤其在学术界，通过提及相关学者或机构，推文更容易吸引目标受众的注意。

跨平台的研究也验证了标签和用户提及对传播效果的正向影响。例如，在整形外科期刊的 Instagram 账号研究中发现，带有更多标签或标记用户的帖子往往获得更高的 altmetrics 得分，而仅包含文章链接或作者简介的帖子则相对关注度较低<sup>[39]</sup>。

## 1.3 研究评述

通过对现有文献的梳理，本文发现，现有研究主要集中在推文文本属性特征对转发概率的影响，而对信息传播中的级联传播现象关注较少。在级联传播过程中，一条推文的传播效果不仅取决于自身的初始影响力，还受到后续转发和互动的累积效应的影响。因此，研究不同文本属性如何影响这一连续传播过程具有重要的理论和实践意义。基于此，本文将研究问题进一步细化，具体如下。

（1）用户在发布科学推文时通常会为科学论文配备什么样的文字内容？具有不同文本内容的科学推文在级联演化趋势上是否存在差异？

（2）何种多媒体信息式推文更容易吸引用户的关注和转发？相较于纯文本式推文，包含多媒体信息的科学推文在级联演化趋势上呈现哪些不同的特征？

（3）科学推文中常用的表情符号的有哪些？相较于纯文本式推文，包含表情符号的科学推文在级联演化趋势中体现出哪些独有的特点？

## 2 研究设计

### 2.1 数据获取

本文选取大型开放获取出版社 Frontiers 旗下期刊于 2021 年发表的科学论文作为研究对象。目前，该出版社已发行多种不同学科领域的学术期刊，覆盖了自然科学、医学、工程、社会科学、人文学科等科学领域。此外，该出版社一直致力于追踪科学论文在不同网络站点的访问次数，如 Twitter（t.co）、Google（www.google.com）、NCBI 数据网站（www.ncbi.nlm.nih.gov）等，这些网络站点的访问统计为本文提供了科学论文传播效果的直接测度。

#### （1）科学论文的 Twitter 传播数据

基于荷兰莱顿大学科学与技术研究中心的内部数据库查询每篇论文的 Twitter 传播数据，共匹配到 77762 篇科学论文（每篇论文的 Twitter 提及次数至少大于 1）。科学论文传播数据主要包含以下字段：原创推文 ID（identity）、原创推文提及的论文 ID、发帖用户 ID、互动用户 ID、发帖时间、互动时间（如转发、回复）、推文文本内容、推文媒体类型（如包含图片、视频和动图）等。用户信息数据主要包含用户的粉丝数、认证状态、账号创建时间、地理位置等字段。

## (2) 科学论文的推文链接访问数据

对于科学论文的 Twitter 站点访问数据, 本文利用该期刊官方网站提供的站点访问数 API (application programming interface) <sup>①</sup> 和 DOI 字段来获取论文的 Twitter 站点访问数。利用 Python 爬虫工具包 Scrapy 编写爬虫程序, 匹配到 52219 篇论文。这些论文共涉及 154778 条原创推文、377267 条互动推文 (转发、引用和回复) 以及 148935 名参与传播用户。

## 2.2 科学论文的传播级联构建

级联传播 (cascade diffusion) 是指信息在社交网络中通过用户之间的相互影响逐步扩散的过程<sup>[9-11]</sup>。这种传播现象广泛存在于多个传播场景中, 如网络博客<sup>[11]</sup>、电子邮件<sup>[12]</sup>、产品推荐<sup>[13]</sup>以及社交媒体平台<sup>[14-16]</sup>。具体而言, 当一个用户分享某条信息时, 其他用户可能会受到该用户行为的影响, 继而产生相似的分享行为, 从而推动信息在网络中的扩散。随着越来越多的用户重复这一过程, 信息逐步传播到更广泛的用户群体当中。这一现象给予本文一个重要的启示: 信息的传播效果是一个多轮积累的过程, 不仅依赖于首次传播, 还受到后续互动的影响。

### (1) 论文层面的传播级联

本文采用复杂网络方法对科学论文的级联传播

过程进行建模。首先, 从数据集中匹配出提及该篇论文所有推文记录, 按照发帖时间和互动时间对其进行升序排列。其次, 基于提及用户之间的互动关系 (转发、回复、引用), 构建一个有向的、以论文为根节点的时变网络  $G_i = (V, E, T)$ 。其中,  $u_i, v_j \in V$  分别表示发帖用户  $u_i$  和互动用户  $v_j$ 。如果用户  $v_j$  转发/回复/引用了用户  $u_i$  发布的推文, 那么这两个用户之间就会形成一条  $u_i$  指向  $v_j$  的连边  $e(u_i, v_j) \in E$ 。连边的权重为  $\Delta t_i \in T$ , 表示论文的传播时间, 其定义为用户围绕论文的第  $i$  次互动时间  $r_i^i$  与首次发帖时间  $p_i^{1st}$  之间的时间间隔  $\Delta t_i = r_i^i - p_i^{1st}$ , 即在论文被用户首次提及后, 经历了多长时间后获得了第  $i$  次提及。最终, 本文构建了 52219 个论文级联。

### (2) 推文层面的传播级联

为了揭示具有不同文本属性特征的科学推文的传播效果, 本文聚焦于由用户发布不同推文所引导的推文级联。具体来说, 首先, 基于论文级联确定分享科学论文的首批发帖用户。其次, 提取由这些用户发布推文后衍生出的传播链, 进而得到推文级联。图1展示了推文级联识别过程。左图展示了示例论文级联结构拓扑图。该论文在出版后, 被3个用户发帖分享, 并引发了多轮的连续传播过程; 右图展示了示例论文的2个推文级联。按照上述步骤, 本文共识别了 49550 个推文级联。

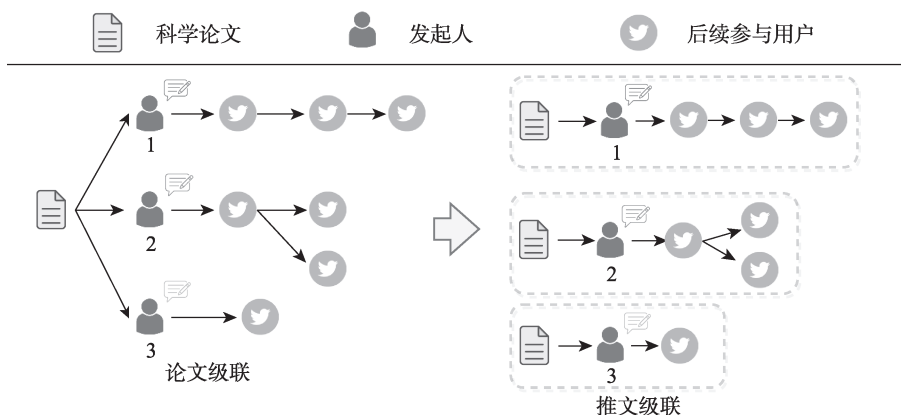


图1 科学论文级联和科学推文级联

## 2.3 科学推文的文本属性特征

### (1) 文本内容

对于每条提及科学论文的原创推文, 本文利用 Python 中的正则表达式对推文内容进行处理, 删除

其中的论文链接、标点符号、表情符号以及用户提及标识 (@用户名), 仅保留纯文字内容。随后, 对推文内容和论文标题进行分词 (以空格为分隔符), 并通过余弦相似度计算两者的文本相似度。分析结果如图2所示, 推文内容与论文标题之间的

<sup>①</sup> [https://impact-api.frontiersin.org/v2/Publications/\(论文DOI\)/ReferrerTo](https://impact-api.frontiersin.org/v2/Publications/(论文DOI)/ReferrerTo)

文本相似度分布极不均匀，其中约 32% 的推文文本相似度位于 0~25% 区间，34% 的推文文本相似度分布在 25%~75% 区间，另有 34% 的推文相似度处于 75%~100% 区间。据此，本文将推文文本分为两种

主要类型：“标题帖”，即直接使用论文标题，或推文中包含论文标题的主要词汇；“非标题帖”，即推文内容与论文标题的词汇重叠度较低，采用与标题完全不同的表达方式。

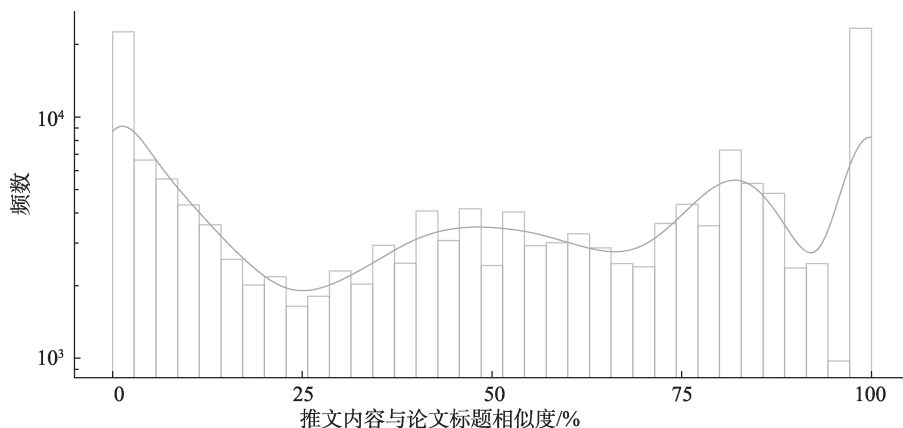


图2 推文内容与论文标题之间的文本相似度

为了进一步揭示不同文本内容的科学推文在级联传播效果上的差异，本文按照用户是否发布标题帖，对这些推文级联进行分组。同时，为控制用户粉丝规模对传播效果的潜在影响，本文将研究样本限定为粉丝数在 10~100 万的发帖用户。选取该区间的理由在于，粉丝数过低（少于 10）的用户传播范围有限，难以形成规模化的推文级联，而粉丝数过高（超过 100 万）的用户通常受限于其强影响力带来的名人效应，会削弱推文文本特征的实际作用。此外，为了确保分析的公平性，本文在相同粉丝数下对两组样本进行匹配，确保每个具体粉丝数下，对照组和实验组的样本数量相同。通过这种方式，尽可能消除粉丝数差异带来的潜在偏差，确保分析结果的公平性和可比性。最终，本文识别出 5225 个首条推文为标题帖的推文级联和 27423 个首条推文为非标题帖的推文级联。

### (2) 多媒体信息类型

本文根据数据集中的多媒体信息类型字段，将原创推文划分为 4 种类型，包含图片、视频、动图和纯文本。其中，纯文本式推文占比约为 88% (136133 条)，包含图片元素的推文占比约为 12% (18342 条)，包含动图和视频元素的推文占比均不足 0.1% (分别为 155 条和 148 条)。为了研究多媒体信息对推文级联传播效果的影响，本文根据发起人发布的首条推文是否包含多媒体信息进行分组，并将发帖用户限定为粉丝数在 10~100 万区间内，以控

制粉丝规模对传播效果的潜在影响。最终，本文识别出 6231 个首条推文为多媒体式推文的推文级联和 29901 个首条推文为纯文本式推文的推文级联。

### (3) 表情符号

表情符号被视为一种社交互动的手段，它们在有限的字符空间中能为文本增添情感和视觉效果。本文利用 Python 工具包 emoji 来提取每条推文内容中的表情符号。emoji 是一个用于处理和操作表情符号的 Python 工具包<sup>①</sup>，能够识别多种表情符号格式，包括标准的 Unicode 表情符号和自定义的短代码表情符号。根据推文是否包含表情符号，本文将原创推文分为“表情帖”和“非表情帖”两类，并识别出 6116 条表情帖和 130017 条非表情帖。为了控制粉丝规模对传播效果的潜在影响，本文进一步将发帖用户限定为粉丝数在 10~100 万区间内，并根据首条推文是否为表情帖对推文级联进行分组。最终，本文识别出 6231 个首条推文为表情帖的推文级联和 29901 个首条推文为非表情帖的推文级联。

## 2.4 级联结构特征

为了揭示具有不同文本属性的科学推文在级联演化趋势上的差异，本文参考 Vosoughi 等<sup>[33]</sup>提出的信息级联结构分析框架，该框架成功揭示了真实新闻和虚假新闻在网络传播上的差异。具体来说，该框架共涉及两类指标：静态指标和动态指标。其中，静态指标主要反映论文级联的基本结构特征；

① <https://github.com/carpedm20/emoji/>

动态指标分为基于深度的动态指标和基于时间的动态指标, 基于深度的动态指标反映了论文级联的层次结构特征, 基于时间的动态指标主要反映论文的级联结构演化趋势。各项指标定义如下。

### (1) 静态指标

**深度:** 级联的深度是指某信息级联中节点深度的最大值。节点的深度定义为该节点与根节点之间的最短路径的边数。对于一个节点数为  $N$  的级联来说, 其深度  $D$  被定义为

$$D = \max(d_i), 0 \leq i \leq N \quad (1)$$

其中,  $d_i$  表示网络中节点  $i$  的深度。图 3a 展示了传播深度的测量方法。在示例级联中, 该级联的深度为 3, 表示该信息引发了 3 轮的连续互动过程。

**规模:** 级联规模是指参与该级联的唯一用户数量。图 3b 展示了不同深度下的参与用户数量变化。图中示例论文的级联规模为 7, 表示有 7 位不同用户参与了示例论文的传播。

**广度:** 广度是指级联在所有传播深度中触及用户数量的最大值。对于一个深度为  $d$  的级联, 其最大广度  $B$  被定义为

$$B = \max(b_i), 0 \leq i \leq d \quad (2)$$

其中,  $b_i$  表示深度为  $i$  时的广度。图 3c 展示了示例网络中不同深度下的广度。在该网络中, 最大广度为 4。

**病毒性:** 该指标主要用来刻画信息传播网络以广播式 (类似于大众传播) 或病毒式 (类似于人际传播) 方式扩散的程度。“病毒性 (virality)” 一词最早可追溯到疾病传播领域。传染病通过人与人接触的方式在短时间内感染大量人群。社交媒体信息传播与传染病传播有相似之处。当一条视频、图片或新闻文章等社交媒体信息“走红 (go viral)” 时, 不仅意味着其能够广泛传播, 也意味着该信息以人际传播的方式进行扩散。根据 Zhang 等<sup>[40]</sup>的定义, 一个网络的级联病毒性是指该网络中节点与其后代节点之间的平均路径长度。对于一个节点数目  $N \geq 2$  的传播网络  $T$  来说, 其级联病毒性被定义为

$$V_T = \sum_{i=1}^N \bar{d}_i, \bar{d}_i = \sum_{j \in D_i} d_{ij} / |D_i| \quad (3)$$

其中,  $d_{ij}$  表示节点  $i$  与子节点  $j$  之间的距离;  $D_i$  表示节点  $i$  的可达节点的集合;  $|\cdot|$  表示集合的大小。图 3d 展示了级联病毒性指标的计算方式, 图中示例网络的级联病毒指标数值为 2.5。

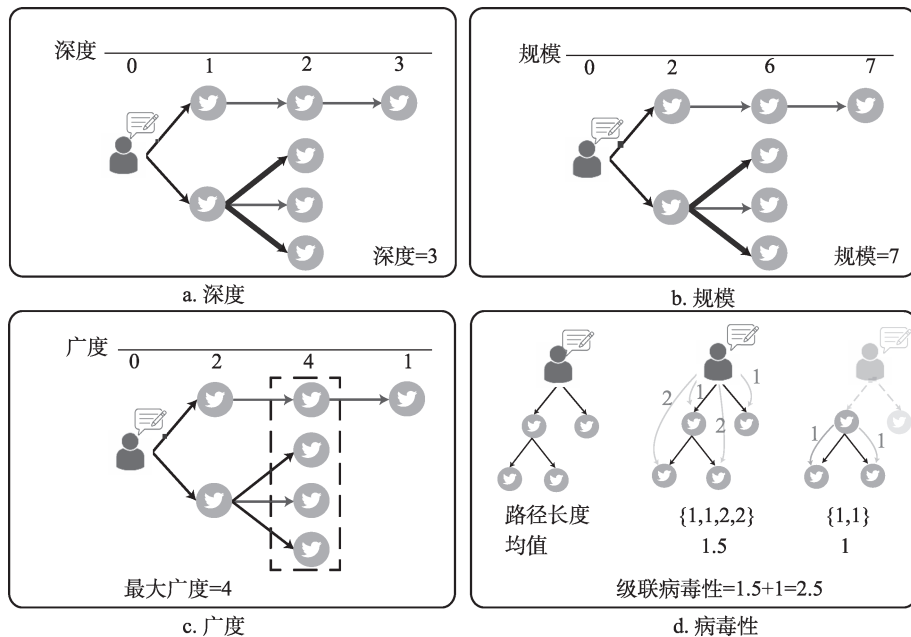


图3 静态网络结构指标计算过程

### (2) 基于深度的动态网络结构指标

**规模随深度变化:** 指不同深度下的累计受众规模, 即信息达到某一深度时累计触达的用户数量。图 4a 展示了当前示例推文级联规模随深度变化的曲线。

**广度随深度变化:** 指不同深度下的受众规模, 即信息在特定传播深度下触达的用户数量。图 4b 展示了当前示例推文级联广度随深度变化的曲线。

### (3) 基于时间的动态网络结构指标

基于时间的动态网络结构指标用于描述规模、

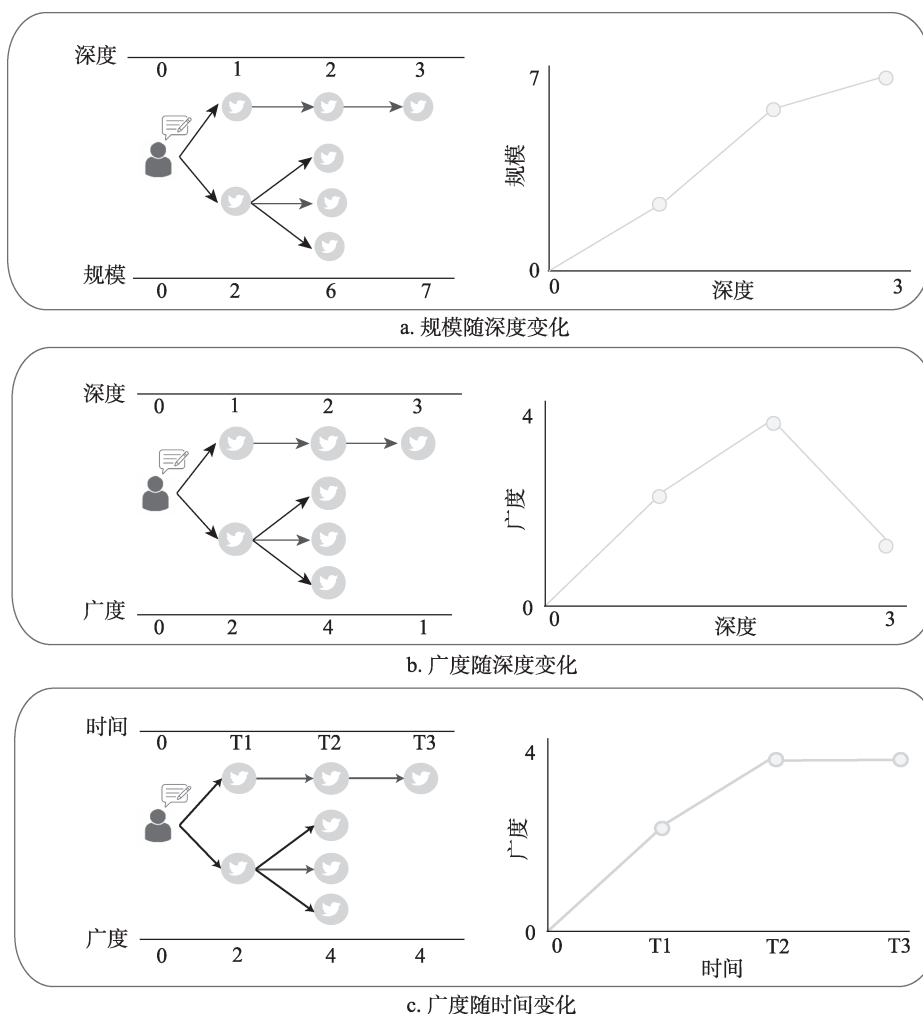


图4 基于深度的动态特征

广度、深度及病毒性随传播时间的变化趋势。图4c展示了当前示例推文级联广度随时间变化的曲线。在T1时刻，推文深度达到1且触及2个用户，此时广度为2；在T2时刻，推文深度达到2且触及4个用户，此时广度为4；在T3时刻，推文深度达到3，但仅触及1个用户，此时广度不变。

### 3 不同文本属性特征的级联传播效果分析

#### 3.1 推文文本内容

##### (1) 标题帖和非标题帖的转发情况

首先对标题帖和非标题帖的转发情况进行基本分析。图5a和图5b分别展示了3种不同类型推文的转发情况。通过比较发现，局部引用标题的推文不仅转发概率较高(0.39)，而且在实际传播中能获得大量转发。相比之下，与标题完全不同的推文传

播效果略逊于局部引用标题的推文，而直接引用标题的推文转发数最低(转发均值：7.18 vs. 3.13 vs. 2.20)。直接引用标题的推文转发较少的原因可能与信息重复性有关。在Twitter平台上，指向科学论文的超链接通常会以缩略图的形式展示，如图5c和图5d所示，该缩略图显示了论文的第一张图片或期刊图标以及论文标题。因此，当推文内容与缩略图中的论文标题一致时，信息显得冗余，从而降低了用户的转发意愿。已有研究表明，反复接触相同信息内容会减少用户的转发意愿<sup>[41]</sup>。

##### (2) 级联传播效果分析

本文通过设定规模、广度、深度和病毒性4个指标，分析标题帖与非标题帖在级联演化趋势上的差异。传播时间定义为第一条推文与后续推文之间的时间间隔，以研究这些传播指标随时间演变的趋势。从时间维度来看，非标题帖在传播初期迅速获得了大量的关注和分享，并在随后的传播过程中持

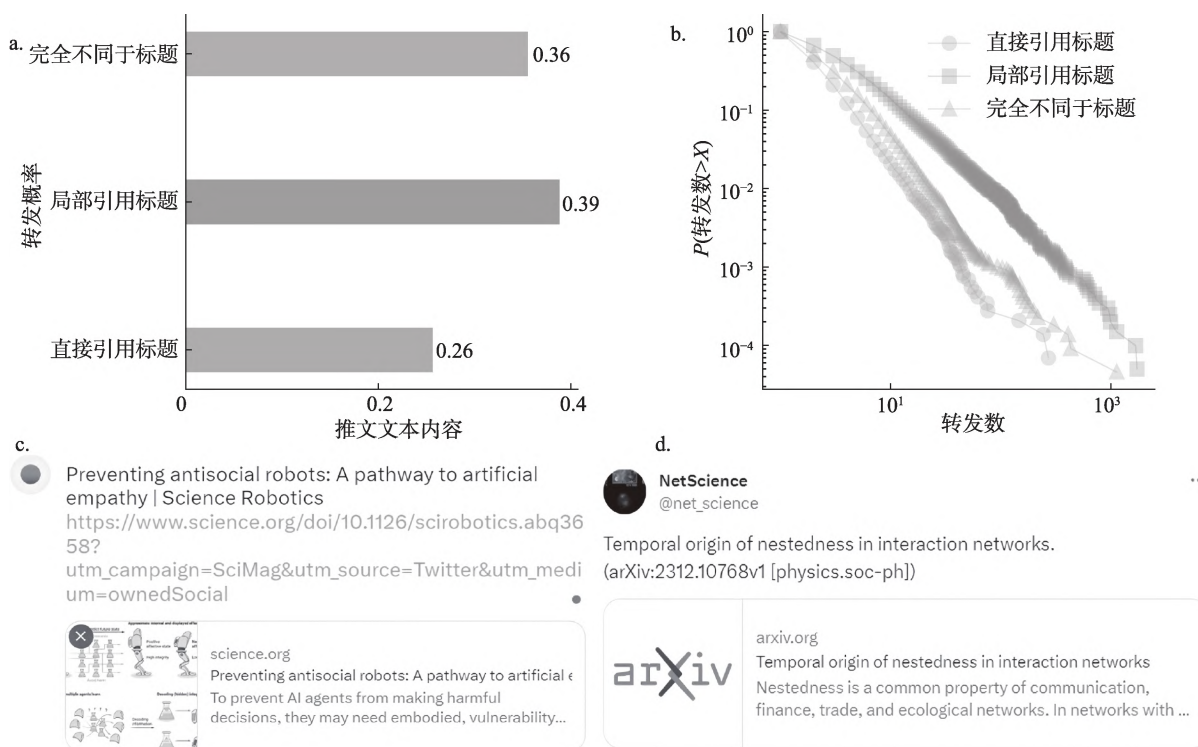


图5 不同推文文本内容下的推文转发情况

续扩大其传播优势。在规模、广度、深度和病毒性4个指标上,非标题帖始终高于标题帖,如图6a~图6d所示。从传播深度的角度来看,非标题帖在各深度下持续保持较高的广度,并且随着传播深度的增加,其广度呈现逐步上升的趋势,始终高于标题帖推文,如图6f所示。同样地,非标题帖推文在每一深度下均能够触及大量的用户,展现出较强的持续传播能力,如图6e所示。综合来看,非标题帖凭借初期传播的迅速突破,为后续传播奠定了坚实基础,推动了传播范围和影响力的持续扩展。这一现象体现了级联传播中的马太效应,呈现“强者愈强”的动态趋势。

### 3.2 多媒体元素

#### (1) 推文的转发情况

根据原创推文中是否包含多媒体信息(视频、图片和动图),将推文分成多媒体式和纯文本式两种。通过比较两种推文的转发情况,发现包含图片、视频以及动图的多媒体推文更容易吸引用户关注,并会影响用户转发行为,如图7所示。具体来说,多媒体推文的转发概率为0.41,而纯文本推文的转发概率仅为0.37。多媒体推文所获得转发数均值是纯文本的2倍之多(均值:8.17 vs. 3.74; K-S

检验:  $D=0.24$ ,  $p<0.001$ )。从多媒体信息的类型来看,带有视频的推文要比动图、图片、纯文本式更容易吸引用户的关注,并获得大量的转发。

#### (2) 级联传播效果分析

从时间维度来看,多媒体推文在传播初期就取得了显著的传播效果,并在后续传播过程中得到加强。具体来说,首条推文为多媒体的级联,在规模、广度、深度和病毒性4个指标上均表现出显著的优势,如图8a~图8d所示。随着传播时间的推移,这种优势不仅得以维持,而且与纯文本推文之间的差距进一步扩大。从传播深度的视角来看,多媒体推文的广度随传播深度呈现先增长后下降的趋势,但在传播链条的各个层次中,其广度始终显著高于纯文本推文,如图8f所示。这表明多媒体推文即使在传播链条的更深层次中,其受众覆盖能力依然保持优势。规模随传播深度的变化表明,多媒体推文在每一传播深度中都维持了较大的用户覆盖,并且在更深层次的传播阶段表现出更强的扩散能力,如图8e所示。总的来说,传播初期的成功为多媒体推文的后续传播奠定了坚实的基础,促进了传播效果的进一步提升和增强,呈现一种典型的级联马太效应,即“强者愈强”的动态特征。

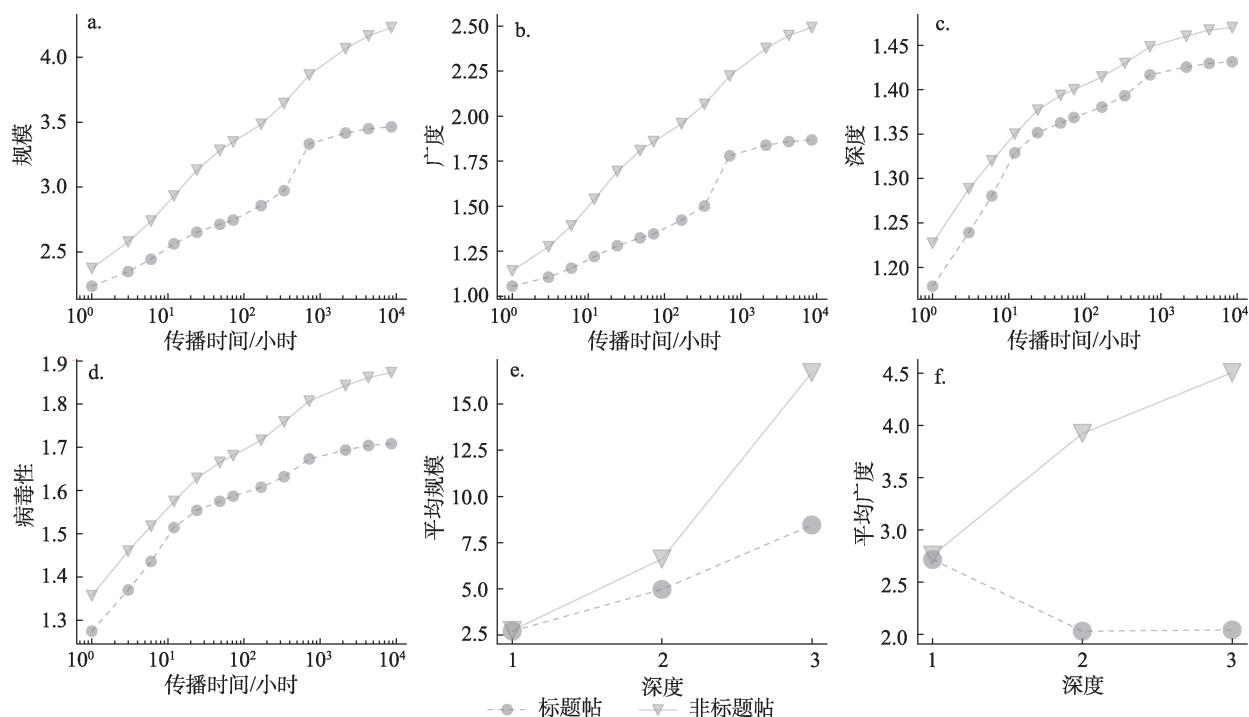


图6 首条推文为标题帖和非标题帖的级联结构演化趋势

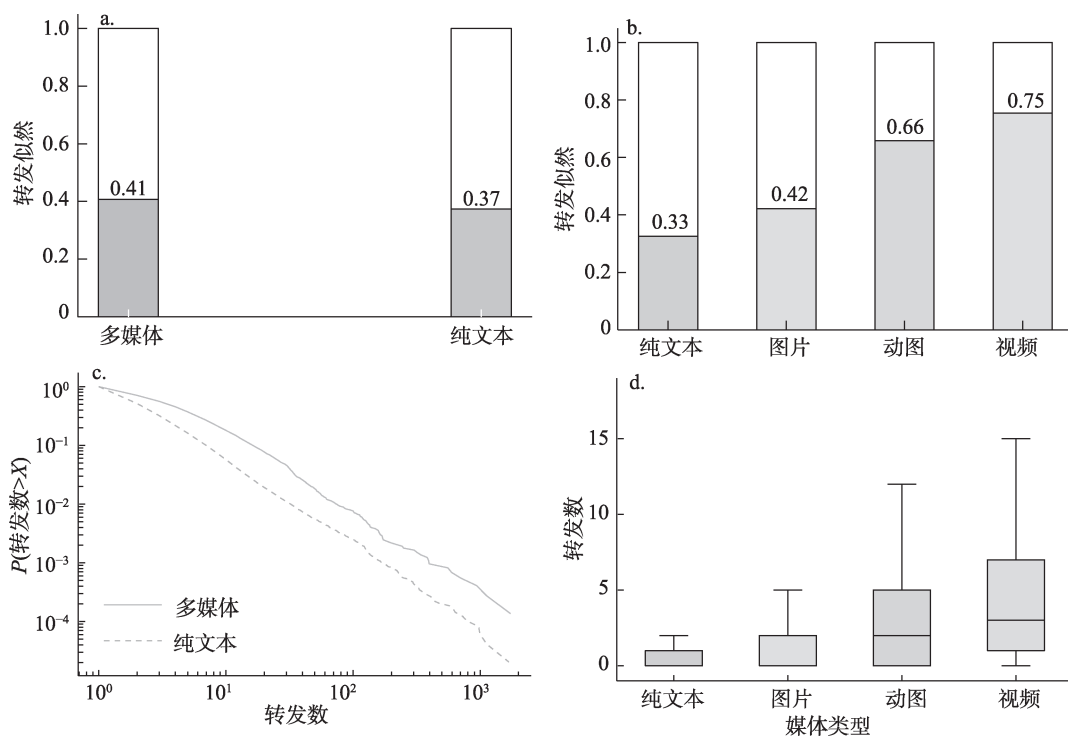


图7 不同多媒体形式下的推文转发情况

### 3.3 表情符号

#### (1) 表情符号的使用频率

表情符号被视为一种社交互动的手段。它们在有限的字符空间内为文本增添情感和视觉效果。统计发现，用户在编辑学术推文时更倾向于在文字内

容中添加一些手势和箭头符号（如👉👈👉👈）来突出论文链接的位置，以更加直观地引导关注者注意链接所在的位置，进而提高科学论文的阅读和传播，如图9所示。此外，一些与医学相关的表情符号也被频繁使用（如💊🩺🩹），以更深入地传达论

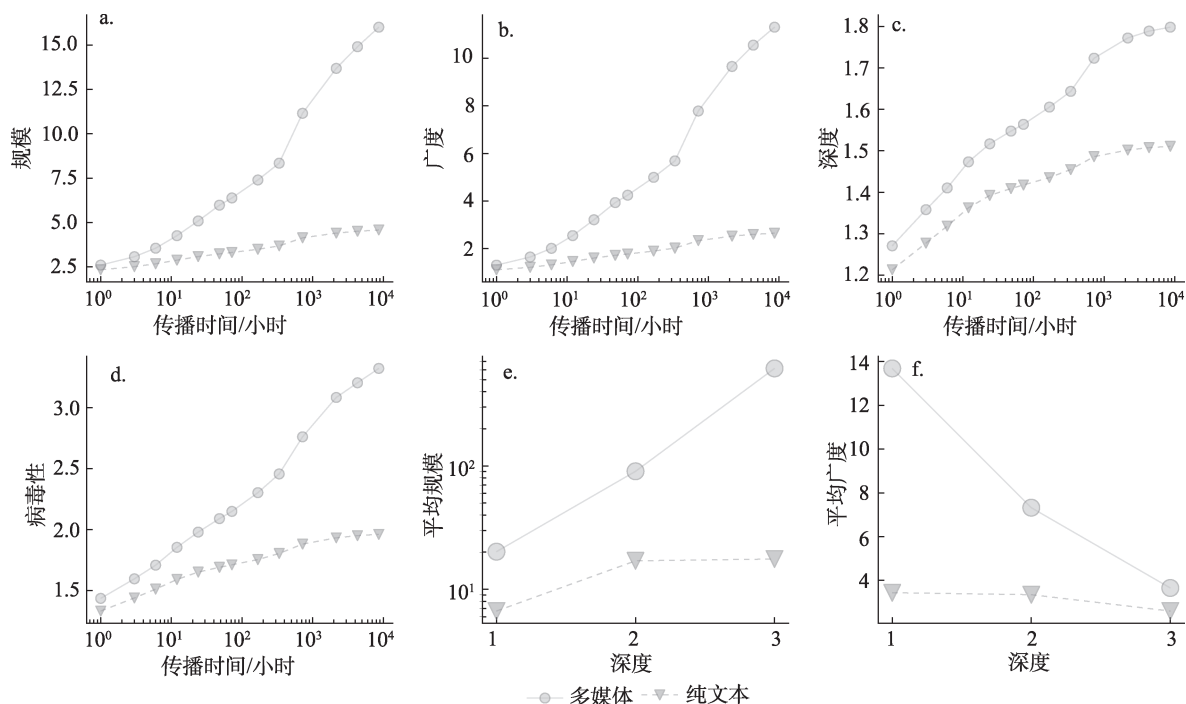


图8 首条推文为多媒体式和纯文本式推文的级联结构演化趋势

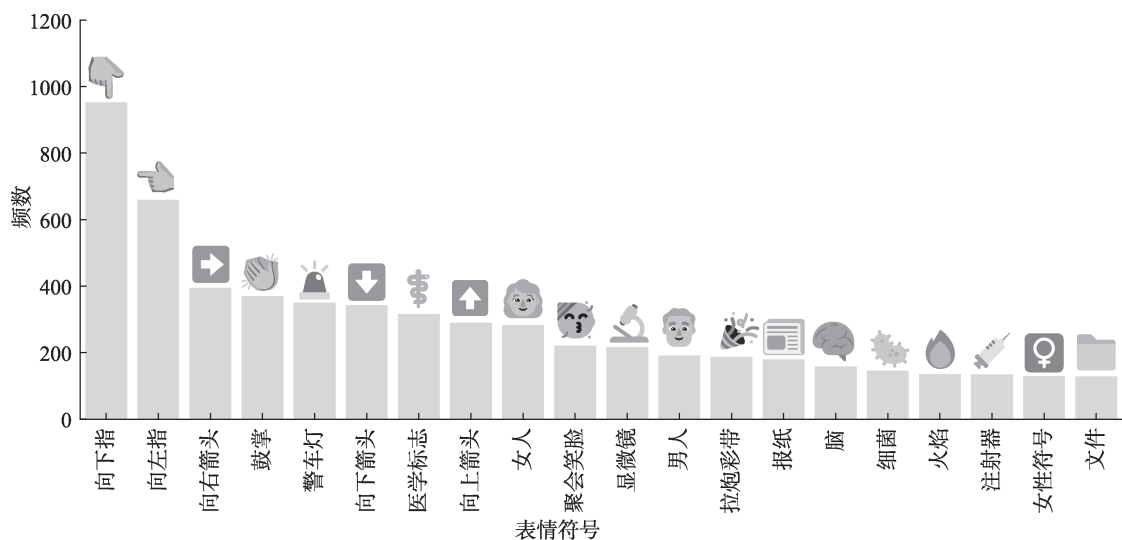


图9 科学论文在社交媒体传播中的表情符号使用情况

文研究内容的氛围和主题，特别是对于涉及医学或生命科学的研究。这有助于读者更形象地理解研究的领域和关注点。

## (2) 级联传播效果分析

从时间维度来看，表情帖在传播初期迅速吸引了大量用户的关注，并在后续传播过程中持续扩大其覆盖范围。表情帖在规模、广度、深度和病毒性4个级联传播的关键指标上均显现出明显的优势，如图10a~图10d所示。随着传播的推进，这些优势不

仅得以保持，还使表情帖与纯文本推文之间的传播效果差距进一步拉大。从传播深度的角度分析，表情帖的广度随传播深度逐步增加，并在深层传播阶段保持高水平，其覆盖范围在传播链条的各个层次均显著高于纯文本推文，如图10f所示。这说明表情帖不仅在传播初期具有优势，甚至在传播链条的深层次中仍能有效触及更多受众。同时，规模随传播深度的变化表明，表情帖在各个传播深度中均能够吸引大量用户的参与，其扩散效应随着传播的深

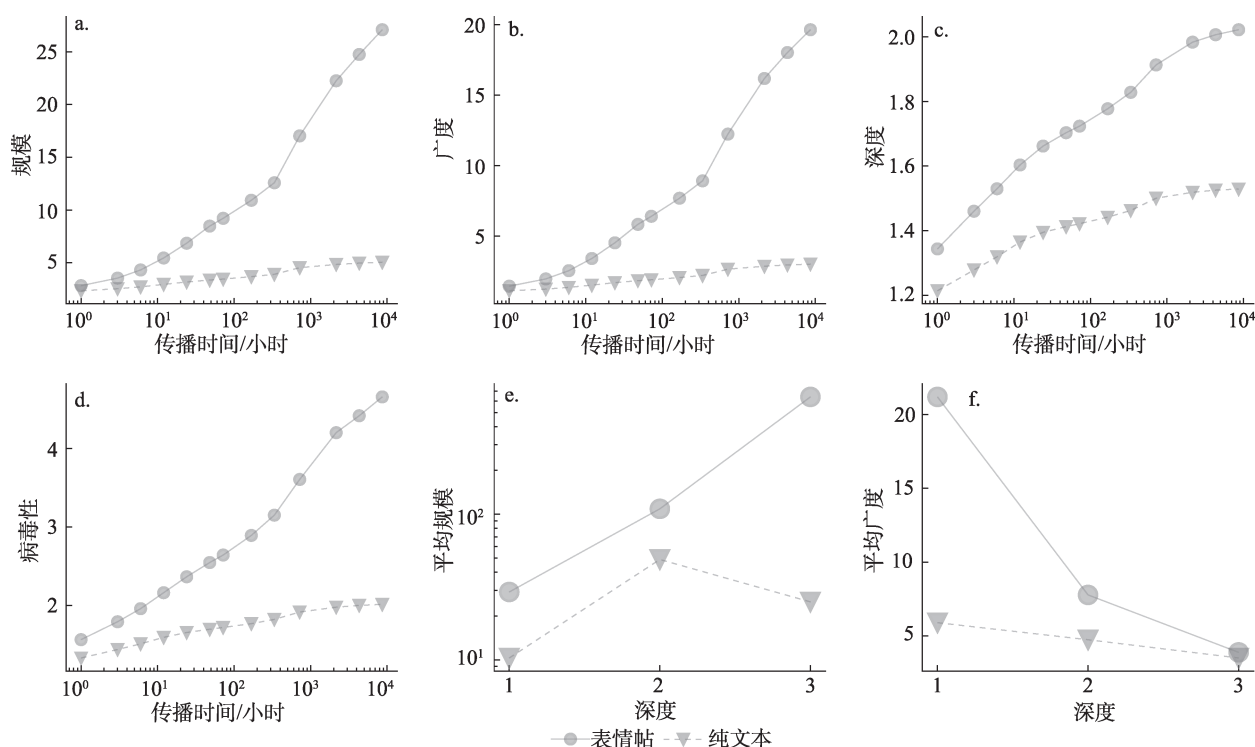


图 10 首条推文为表情帖和非表情帖的级联结构演化趋势

入持续显现，在深层传播阶段展现出更大规模的扩散能力，如图 10e 所示。这一结果表明，表情帖在传播链的初期便展现出显著的传播优势，并在后续传播阶段进一步扩大了其覆盖范围和影响力，充分体现了“强者愈强”的级联传播特性。

## 4 结论与讨论

本文从科学推文的文本内容、多媒体信息和表情符号 3 种文本属性特征入手，通过级联传播的视角，比较分析具有不同文本属性特征的科学推文的级联演化趋势。研究发现，将论文亮点内容作为科学推文的文本内容，在科学推文中加入图片、视频、表情符号等视觉元素能够显著提升科学论文的传播范围。这种提升不仅在传播初期就有所显现，而且在后续的传播过程中得到进一步增强，形成了“强者愈强”的传播趋势，即级联马太效应。

级联传播是信息传播中的一种常见现象，广泛存在于多种传播场景，如网络博客<sup>[11]</sup>、电子邮件<sup>[12]</sup>、产品推荐<sup>[13]</sup>以及社交媒体平台<sup>[14-16]</sup>。在这个过程当中，初始用户的传播行为会对后续用户产生影响，形成链式反应，最终实现大规模扩散的过程。已有研究通过网络视角，分析了信息的级联结构特征，并对级联传播的动态规律进行了建模与仿真<sup>[42-45]</sup>。相关研究表明，即便两个级联的传播规模

相似，其在时间演化、转发深度分布、级联规模的可预测性以及传播用户的群体特征上可能存在显著差异<sup>[42]</sup>。这意味着级联传播可以通过多种截然不同的方式实现大规模扩散。与前人研究不同，本文结合传播时间和级联深度这两个维度，探讨了初期传播效果与后续传播效果之间的潜在联系，揭示了科学论文在社交媒体传播中的“级联马太效应”：初期扩散优势在后续传播中持续放大，呈现“强者愈强”的动态演化过程。

与本文结论相似的一项工作是 Cheng 等<sup>[46]</sup>对 Facebook 数据的分析，他们分析了信息流行度的增减趋势，发现初期传播中的爆发性传播峰值越高，后续传播更容易经历多轮次的爆发。该研究结论与本文观察到的“级联马太效应”具有相似之处，均表明初期传播的成功为后续传播的持续扩展提供了动力，为本文的结论提供了支持。此外，纪晓玉等<sup>[47]</sup>通过对社交媒体平台的信息传播研究发现，初期传播的成功会促进信息在后续传播过程中的持续扩展，尤其是在社交网络中，传播的早期成功能够显著增加信息被转发和扩散的概率。这些研究结果与本文的“级联马太效应”相呼应，都强调了传播初期的影响对后续传播过程的重要推动作用，为本文提供了有力的理论支撑。

本文的主要贡献体现在 3 个方面。首先，本文

基于计算传播学的视角,结合社交媒体中的级联传播现象,深入分析了科学推文的文本属性特征(如多媒体信息、表情符号等)对传播效果的影响,为揭示科学论文在社交媒体上的传播机制提供了新的理论框架。其次,本文通过大规模社交媒体数据的实证分析,发现不同文本属性特征对级联传播的范围和深度具有显著影响,揭示了“级联马太效应”,这一发现为科学传播策略的优化提供了有力的实证依据。进一步地,本文为科学传播提供了策略性指导,特别是在科学推文内容设计方面,强调如何有效利用多媒体信息和情绪化表达等元素提升传播效果,可为科研人员和科学传播机构提供具体的社交媒体传播策略优化建议。

然而,本文也存在一定的局限性。首先,尽管本文使用了大规模的社交媒体数据进行分析,却并未聚焦于特定的学科领域,未来研究可以扩展至不同学科,验证研究结论的广泛适用性。其次,推文的传播效果不仅受到文本属性的影响,还可能受到用户特征、社交关系网络等多维度因素的共同作用,但这些因素未在本文进行深入探讨。未来应进一步考察这些多维度特征对科学推文传播的影响,以全面揭示科学推文传播的规律,为科学传播的研究提供更为丰富的视角和理论支持。

### 参 考 文 献

- [1] Bik H M, Goldstein M C. An introduction to social media for scientists[J]. PLoS Biology, 2013, 11(4): e1001535.
- [2] Osterrieder A. The value and use of social media as communication tool in the plant sciences[J]. Plant Methods, 2013, 9(1): Article No.26.
- [3] 贾鹤鹏, 范敬群, 彭光芒. 从公众参与科学视角看微博对科学传播的挑战[J]. 科普研究, 2014, 9(2): 10-17, 32.
- [4] van Eperen L, Marincola F M. How scientists use social media to communicate their research[J]. Journal of Translational Medicine, 2011, 9: Article No.199.
- [5] Márquez M C, Porras A M. Science communication in multiple languages is critical to its effectiveness[J]. Frontiers in Communication, 2020, 5: 31.
- [6] Collins K, Shiffman D, Rock J. How are scientists using social media in the workplace?[J]. PLoS One, 2016, 11(10): e0162680.
- [7] Dudo A. Scientists, the media, and the public communication of science[J]. Sociology Compass, 2015, 9(9): 761-775.
- [8] López-Goñi I, Sánchez-Angulo M. Social networks as a tool for science communication and public engagement: focus on Twitter[J]. FEMS Microbiology Letters, 2018, 365(2): fnx246.
- [9] Cheng J, Adamic L, Dow P A, et al. Can cascades be predicted? [C]// Proceedings of the 23rd International Conference on World Wide Web. New York: ACM Press, 2014: 925-936.
- [10] Gao L Q, Zhou B, Jia Y, et al. Information cascading in social networks[M]// MDATA: A New Knowledge Representation Model. Cham: Springer, 2021: 234-254.
- [11] Adar E, Adamic L A. Tracking information epidemics in blog-space[C]// Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence. Washington, DC: IEEE Computer Society, 2005: 207-214.
- [12] Golub B, Jackson M O. Using selection bias to explain the observed structure of internet diffusions[J]. Proceedings of the National Academy of Sciences of the United States of America, 2010, 107(24): 10833-10836.
- [13] Leskovec J, Adamic L A, Huberman B A. The dynamics of viral marketing[J]. ACM Transactions on the Web, 2007, 1(1): 5-es.
- [14] Cao R M, Liu X F, Fang Z C, et al. How do scientific papers from different journal tiers gain attention on social media?[J]. Information Processing & Management, 2023, 60(1): 103152.
- [15] Cao R M, Wang X W, Xu X K, et al. Multiple bursts of highly retweeted articles on social media[J]. Scientometrics, 2021, 126(6): 5165-5179.
- [16] Ye Y E, Na J C, Oh P. Are automated accounts driving scholarly communication on Twitter? A case study of dissemination of COVID-19 publications[J]. Scientometrics, 2022, 127(5): 2151-2172.
- [17] Can E F, Oktay H, Manmatha R. Predicting retweet count using visual cues[C]// Proceedings of the 22nd ACM International Conference on Information & Knowledge Management. New York: ACM Press, 2013: 1481-1484.
- [18] Gayberi M, Oguducu S G. Popularity prediction of posts in social networks based on user, post and image features[C]// Proceedings of the 11th International Conference on Management of Digital EcoSystems. New York: ACM Press, 2020: 9-15.
- [19] Hong L J, Dan O, Davison B D. Predicting popular messages in Twitter[C]// Proceedings of the 20th International Conference Companion on World Wide Web. New York: ACM Press, 2011: 57-58.
- [20] Jenders M, Kasneci G, Naumann F. Analyzing and predicting viral tweets[C]// Proceedings of the 22nd International Conference on World Wide Web. New York: ACM Press, 2013: 657-664.
- [21] Ma Z Y, Sun A X, Cong G. On predicting the popularity of newly emerging hashtags in Twitter[J]. Journal of the American Society for Information Science and Technology, 2013, 64(7): 1399-1410.
- [22] Suh B, Hong L C, Piroli P, et al. Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network [C]// Proceedings of the 2010 IEEE Second International Conference on Social Computing. Washington, DC: IEEE Computer Society, 2010: 177-184.
- [23] 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.
- [24] Zhang J M, Wang Y, Shi M L, et al. Factors driving the popularity and virality of COVID-19 vaccine discourse on Twitter: text mining and data visualization study[J]. JMIR Public Health and Surveillance, 2021, 7(12): e32814.
- [25] Pancer E, Poole M. The popularity and virality of political social media: hashtags, mentions, and links predict likes and retweets of 2016 U. S. presidential nominees' tweets[J]. Social Influence, 2016, 11(4): 259-270.
- [26] Naveed N, Gottron T, Kunegis J, et al. Bad news travel fast: a content-based analysis of interestingness on Twitter[C]// Proceedings of the 3rd International Web Science Conference. New York: ACM Press, 2011: 1-7.
- [27] Dang-Xuan L, Stieglitz S, Wladarsch J, et al. An investigation of influentials and the role of sentiment in political communication on Twitter during election periods[J]. Information, Communication & Society, 2013, 16(5): 795-825.
- [28] Fan R, Zhao J C, Chen Y, et al. Anger is more influential than joy: sentiment correlation in Weibo[J]. PLoS One, 2014, 9(10): e110184.
- [29] Möller A M, Kühne R, Baumgartner S E, et al. Exploring user responses to entertainment and political videos: an automated content analysis of YouTube[J]. Social Science Computer Review, 2019, 37(4): 510-528.
- [30] Stieglitz S, Dang-Xuan L. Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior[J]. Journal of Management Information Systems, 2013, 29(4): 217-248.
- [31] Kim J, Yoo J. Role of sentiment in message propagation: reply vs. retweet behavior in political communication[C]// Proceedings of the 2012 International Conference on Social Informatics. Washington, DC: IEEE Computer Society, 2012: 131-136.
- [32] Wu S M, Tan C H, Kleinberg J, et al. Does bad news go away faster? [C]// Proceedings of the 5th International AAAI Conference on Web and Social Media. Palo Alto: AAAI Press, 2011: 646-649.
- [33] Vosoughi S, Roy D, Aral S. The spread of true and false news online[J]. Science, 2018, 359(6380): 1146-1151.
- [34] Romero D M, Meeder B, Kleinberg J. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on Twitter[C]// Proceedings of the 20th International Conference on World Wide Web. New York: ACM Press, 2011: 695-704.
- [35] Zhao Z L. Propagation structure feature of entertainment news in the Weibo online social network[J]. Europhysics Letters, 2021, 135(1): 16002.
- [36] 钟智锦, 周金金, 徐铭达, 等. 娱乐信息与公共信息的扩散竞争: 网络结构和传播主体视角[J]. 新闻与传播研究, 2023, 30(3): 88-107, 128.
- [37] van Honk J, Costas R. Integrating context in Twitter metrics: preliminary investigation on the possibilities of hashtags as an alt-metric resource[C]// Proceedings of the 2016 Altmetrics Workshop Bucharest. London: Digital Science and Research Solutions, 2016. DOI: 10.5446/46291.
- [38] Ye Y E, Na J C. To get cited or get tweeted: a study of psychological academic articles[J]. Online Information Review, 2018, 42(7): 1065-1081.
- [39] Maisner R S, Goydos C, Ong M, et al. Promoting publications through plastic surgery journal instagram accounts: is it worth it? [J]. Annals of Plastic Surgery, 2023, 90(6S): S693-S698.
- [40] Zhang Y F, Wang L, Zhu J J H, et al. Viral vs. broadcast: characterizing the virality and growth of cascades[J]. Europhysics Letters, 2020, 131(2): 28002.
- [41] Cao R M, Geng Y, Xu X K, et al. How does duplicate tweeting boost social media exposure to scholarly articles? [J]. Journal of Informetrics, 2022, 16(1): 101249.
- [42] Dow P A, Adamic L, Friggeri A. The anatomy of large Facebook cascades[C]// Proceedings of the 7th International AAAI Conference on Web and Social Media. Palo Alto: AAAI Press, 2013: 145-154.
- [43] Kumar R, Mahdian M, McGlohon M. Dynamics of conversations [C]// Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2010: 553-562.
- [44] Liben-Nowell D, Kleinberg J. Tracing information flow on a global scale using internet chain-letter data[J]. Proceedings of the National Academy of Sciences of the United States of America, 2008, 105(12): 4633-4638.
- [45] Gruhl D, Guha R, Liben-Nowell D, et al. Information diffusion through blogspace[C]// Proceedings of the 13th International Conference on World Wide Web. New York: ACM Press, 2004: 491-501.
- [46] Cheng J, Adamic L A, Kleinberg J M, et al. Do cascades recur? [C]// Proceedings of the 25th International Conference on World Wide Web. Republic and Canton of Geneva: International World Wide Web Conferences Steering Committee, 2016: 671-681.
- [47] 纪晓玉, 曹仁猛, 张玮, 等. 风险扩散视角下网络突发事件的信息传播机制研究[J]. 新闻大学, 2024(10): 74-89, 124.

(责任编辑 魏瑞斌)