

DOI:10.15896/j.xjtuskxb.202203015

殊途同归:社交媒体用户内容生产中信息茧房趋同化 ——基于新浪微博的实证分析

徐翔¹, 敖子棋¹, 史静远¹, 王晓虹²

1. 同济大学 艺术与传媒学院, 上海 200092

2. 复旦大学 新闻学院, 上海 200433



开放科学(资源服务)
标识码(OSID)

【摘要】为了证实用户内容生产的信息“茧房”到底是同质化的还是异质化的,基于典型社交媒体之一的新浪微博样本进行了考察与检验。研究发现:随着用户“茧房度”的加深,越是高“茧房度”的个体,“茧房”之间越是趋同而非趋异;用户“茧房度”越高越会趋同于某种公共“茧房”或社会性的中心信息,越会消磨掉自身的独特性而展现趋同性,发生个体之间的“茧房”趋同化;个体“茧房度”的提高伴随着与“顶部茧房”以及与“相邻茧房”内容相似度的提高。上述“茧房”趋同路径具有一致性,四者之间同步共变。“茧房”趋同化现象虽与直观经验或流行观点有相悖之处,却展现了用户内容生产和社会传播的现实逻辑,可以为网络信息扩散及其结构问题提供参考。

【关键词】社交媒体;用户内容生产;信息茧房;用户趋同;社会信息中心;内容相似度

【中图分类号】G206 **【文献标识码】**A **【文章编号】**1008-245X(2022)03-0133-08

信息茧房(information cocoons)的相关理论最早由桑斯坦^[1]提出,该理论认为用户会因选择性心理使自己桎梏于如蚕茧一般的屏障中,与外界多样化信息隔绝。这一效应被认为是用户在媒介个性化使用上的高度自治化带来的副作用。学者们提到信息“茧房”时,往往都会谈及信息偏食导致的视野局限。长此以往,深陷“茧房”窠巢之中的用户会因偏食形成某种狭隘的价值观和认知结构,导致其发布的帖文在“茧房”内呈现出同质化、极端化倾向。然而,跳出个体“茧房”,甚至群体圈层之外,把目光放远到整个网络社会层面时,用户生产的信息内容的“茧房”到底是异质化还是同质化的?尽管这一问题还存在争论,但一般认为,随着“茧房”程度的加深,要么每个个体内容生产中的“茧房”是趋异化而非趋同的,要么认为用户生产内容时本就容易随机地陷入各种细分性、差异性的“茧房”因而不存在“茧房程度越深就越趋同”的问题。

本文提出并加以实证检验的核心观点是:在社交网

络的内容生产中,随着用户所生产的信息内容的“茧房”程度的加深,各“茧房”之间是彼此趋同化而非趋异化的,用户在朝向“同出一辙”的“茧房”发生趋同,而非朝向“千人千面”的“茧房”产生趋异。这个观点似乎并不符合直观经验,也和一般关于“信息茧房”“过滤气泡”等的观点都存在差异。本文将基于典型社交媒体之一的新浪微博样本,对于“个体所陷入的‘茧房’是越长越像而非各有不同”的观点,从三个关键子假设及其内部关系进行考察与检验。

一、文献回顾与问题分析

在生活中,人们常常会主动接触、阅读、点击和发布与自己兴趣偏好符合的内容,导致自己陷入信息窄化和认知狭隘的风险之中,这一现象在学界受到了广泛关注。目前,国内外学界对于信息“茧房”的研究主要从选择性接触、信息窄化、群体极化、政治偏见、种族主义、媒介素养等角度进行探索,与之密切相关的概念还有“回音室”

【收稿日期】2021-07-28。

【基金项目】国家自然科学基金项目(71804126);上海市级科技重大专项(2021SHZDZX0100)。

【作者简介】徐翔(1983—),男,同济大学艺术与传媒学院副院长,教授。

效应(echo chambers)、“过滤气泡”(filter bubble)等。信息“茧房”强调个体选择,“回音室”效应强调群体封闭,而“过滤气泡”则回应了目前正被学术界热议的说法,即算法技术加剧了信息“茧房”的产生和泛化。在实践中,这三者的研究本质上都可以归结于“网络信息同质化”^[2]。Kossinets等^[3]都认同同圈层内兴趣爱好一致的用户会逐渐趋同,社交网络中的个体之间存在着一种同质性的关系。

用户内容生产的信息“茧房”到底是同质化的还是异质化的呢?现有研究对两方面均有涉及,但支持信息“茧房”的异质化观点较多。Web 2.0时代媒体权力的下放和消解让用户更加主动地参与到传播的各个链路当中,用户的互动特征突出呈现出其在信息消费中的能动性,自媒体用户生产内容(user-generated content, UGC)和网络微名人(micro-celebrity)的出现描绘出更平民化的“人人皆媒体”的传播图景。许多研究者认为和以前相比,人们所阅读、接收、发布和讨论的信息内容大大扩增,又由于不同个体的兴趣爱好并不一致,其感兴趣的议题也并不相同,所以在信息“茧房”等效应的影响下,公众舆论呈现出分众化、圈层化和异质化的传播格局。如Alstyn等^[4]认为同质化的小社群之间的沟通导致圈层内的专业化,导致跨群组交流异常困难,群组间逐渐趋异;汤景泰等^[5]认为分散的社交媒体用户会在兴趣、身份认同、政治立场等因素的驱使下形成多个交流的圈层,“因圈层的排斥性与防御性较强、他者信息流通不畅,可能形成‘回音室’”。姚文康^[6]认为,信息“茧房”导致群体因为缺少异质观点的接触而呈现出异质性。

从表面上看,流动的、消除确定性的网络技术逻辑,确实可能在互联网早期阶段导致用户内容生产的异质化和分散化。然而也有研究认为用户“茧房”之间具有同质化趋向,支持这一观点的研究主要从信息同质化和群体因素两部分进行讨论。前者的相关研究主要认为信息同质化会间接影响用户的价值认同,进而促进“茧房的同质化”^[7]。一是指出信息发布环节的内容同质化。如Pinto等^[8]通过实证检验发现,媒体议程和公共议程相互影响,并且存在媒体框架设置等问题导致公共议程的单一化。Nechushtai等^[9]在近期对气候报道的研究中发现,新闻议程仍然受几家主要媒体控制,来自他们的信息会被各网络平台重点推荐,“即使对有不同内容倾向的用户,过滤气泡仍会分发给他们相似的内容”。Lee^[10]发现社交网络中博客议程和主流媒体议程的趋同化。宋建武等^[11]从垂直互文性和水平互文性两个维度分析了微信新闻公众号所呈现出的同质化表征。传统的业界规矩和惯习形成了小媒体、自媒体在内容筛选上向大媒体看齐的现象,加剧同质化信息在各家传统媒体上的流行,导致用户在接收信息时潜移默化地形成某种趋同

化的认知结构,进而影响到其发布帖文的倾向。二是一定程度上涉及社交媒体中的“茧房”同质化现象。有学者提出社交网站对成员的“显式行为矫正与规训”^[12],即同一社交网络平台具有某种共通的沟通标准,导致用户的内容生产并未呈现任意和自由多样的演进形态,而是趋向于和该平台保持一致。由此可进一步推论,这种社交网络的潜在规约很有可能让用户内容生产的多样性减少,并使得各用户内容趋于某种共同特征和具有相通性的束缚之中。三是认为算法技术导致“茧房”趋同化,推荐系统加剧信息同质化。如Hosanagar等^[13]认为,算法推荐虽然在一定程度上强化了小众兴趣爱好群体的娱乐需求,使用户形成一个个分散的小众群体,但是在整体上用户的消费内容会越来越相似,并且逐步趋近于一个整体,“个性化并没有把网络分裂成一个个孤立的小‘茧房’,它虽是一种帮助用户扩大兴趣的工具,但又使得人们的消费兴趣变得与他人更相似”。延展到信息消费领域这一观点也得到印证。Airolidi等^[14]对UGC平台“推荐观看”音乐视频之间的关联网络的研究表明,网络平台基于用户最频繁的共同浏览量来关联和推荐相似视频,从而塑造出一个彼此关联的视频集群,结果表明这些集群具有很强的内容同质性。

群体因素对“茧房”同质化研究也十分重要,这部分研究大多从用户在群体中的印象管理、风险感知、利益感知等心理层面探讨其发布信息时的同质化倾向。一是出于社交网络中自我形象管理的需要而选择趋同化。如Pedroni等^[15]发现随着好友数量的增加,人们倾向于更多地与所有人都感兴趣的内容交互,以保护自己的形象。Hasen等^[16]¹⁶⁵⁻¹⁷⁹对Facebook用户内容生产的研究表明,人们倾向于选择多数人都感兴趣的内容进行交互,以此凸显个人身份中受大多数好友欣赏的部分,以保护或提升自己在社交平台的形象。二是出于群体内感知到的压力使得用户选择与群体保持一致。Su等^[17]研究发现人们通常通过观察性学习行为更新和表达自己的观点,所以通常在发声时会受到舆论的影响而修改自我的意见。Jun等^[18]测试了在社交网络平台上感知他人的存在是如何影响个人评估信息的方式的。McEwan等^[19]还提出了“自我涉入程度(ego-involvement)”的概念,认为在特定议题的讨论中寻求与自己一致观点的群体的行为,会表现出个体对于某一问题的自我涉入程度,并且参与群体的时间越长,该值越高。三是发现了用户感知到社交网络平台的特点后主动趋同的行为。Mikal等^[20]发现网民受到公开沟通行为标准的显式行为矫正,即在网站中存在着一种公开的沟通行为标准,用户会不约而同地遵守这一标准,相应地修改自己的内容,使之趋同于网站内容。后续的研究还发现,匿名条件下的网络用户在受到与群体显著社会身份一致的行为期望时,更有可能

表现出符合群体规范的行为。总体而言,这些研究对本文具有启发作用,但仍存在巨大的挖掘空间:他们大多未能跳脱出群体内的视角来探讨更广阔的网络社会层面下,群体因素对“茧房”同质化的影响。实际上,网络社会是一个具有主流价值观、社会心态和情感结构的大群体,在这个社会群体的互动中,人们通常会受到上述因素的影响而更新和表达自己的观点,从而体现出在内容发布上的趋同特征。

许多研究都论证了同质化程度越高的信息环境中,越能够形成具有高热度、高流量的社会中心议题。比如Nikolov等^[21]发现社交网络平台呈现出高水平的流行度和同质化偏向。他们通过量化两个概念得出结论:一是同质性偏向,即用户从狭窄的信息源中消费内容的趋势;二是流行性偏向,即热门新闻内容的选择性曝光程度。通过实证分析发现,这两个概念具有一定的相关关系。Qin等^[22]认为社交网络的一大特征是同质驱动性(homophily-driven)。由于社会信息的扩散是盲目随机的,所以越是同质化的信息交流环境,越容易形成传播广、热度高的社会议题。Scheufele等^[23]认为现代社会的信息是通过大量读者评论、Facebook上的“点赞”以及某篇报道被“转发”等方式进行传播的,这导致社交网络呈现出一种趋同化的、群体性偏好的内容取向。

区别于传统媒介,社交媒体平台是各类偶发性信息和意见交换的空间,用户被与熟人网络不同的“弱连接”方式联结起来。基于这样的传播逻辑,试图发现这种社会中心议题的研究主要从用户特征和环境特点的角度开展。Fletcher等^[24]发现即使用户的自我选择权力变大,他们还是会选择(或是“偶遇”)被多数人关注的“主流信息”。Bar-Gill等^[25]总结了对“回音室”敏感的人群的主要特点,一是在搜索过程中很少进行探索,二是在内容选择上注重其流行程度。故陷入“茧房”程度越深的用户,所接收的信息内容也越流行、越多人关注、处于社会舆论场越中心的议题,其发布的信息也越贴合这些议题。所以, Lee^[26]认为尽管信息渠道是多样化的,但人们仍可能在主流媒体和互联网新闻媒体上接收相当稳定的信息。

对于互联网环境特点,相关研究突出了对这种演变机制的讨论,主要从组织传播、网络结构和社会共识等角度展开探索。程士安等^[27]认为信息“茧房”是信息聚合、共识达成的现代社会组织的最小单位,它具备Web 2.0时代下网络组织关系的“纽带”和“锁扣”特征。在生活中,具备该特征的信息“茧房”以惊人的速度吸引、聚合、黏着同类信息需求的人群,通过某一信息的多次共享编码与解码的过程,形成相似的子“茧房”,“最终在特定的大‘茧房’平台上交流并达成共识”。子“茧房”之间的这种黏合性,实际上表现为前文所论述的用户生产内

容的同质性,所谓的大“茧房”也就是社交媒体平台上对于某一热门议题的集中讨论。从这一观点出发,似乎可以认为,社交媒体上的用户都在有序地朝向某一大“茧房”所在的中心节点趋近。从网络结构的角度出发,有研究发现“信息窄化下的社交网络呈现出星形结构”^[28],中心节点的影响越强,则“回音室”效果越强。Webster等^[29]将这个中心节点具体化为最受媒体和大众关注的新闻内容,这为主流新闻作为中心节点的说法提供了参考价值。这些研究都在一定层面上论述了社会舆论信息中心的存在,但还缺少将信息中心与用户的信息窄化结合的讨论。个体信息窄化是如何受到社会公共性、公共信息和公共焦点的共同影响而形成的?这种“窄化”是否是一种“公共性的窄化”而非“个体性的窄化”?这些都是亟待解决的问题。

总体而言,这些相关研究为本文提出用户内容生产中所形成的信息“茧房”趋同化观点提供了参考,也留下了继续挖掘的空间。本文不仅关注用户是否处于“茧房”中、处于何种“茧房”中、因何处于“茧房”中,还关注“茧房”之间的关系。因此,“茧房”是否趋同、如何趋同是本文要探讨的核心问题。

二、研究设计与实施过程

本文的核心问题是探讨随着用户生产的信息内容的“茧房”程度的加深,“茧房”之间是呈现趋同化还是异质化,抑或无明显规律?如果是趋同化,是怎样趋同的?结合理论和文献分析,本文从3个关键假设、2个推论及这五者的整体关系,基于新浪微博的用户数据样本进行实证检验。

(一)假设的内涵及其关系

为了方便描述用户在社交平台上内容生产中的“茧房”程度,引入“茧房度”(degree of cocoon)这一概念来形容个体生产的信息内容自我重复化、相似化的程度。即“茧房度”越高的用户,其生产的内容越是陷入相似性的窠臼,多样性和异质性越少。在此基础上,把具有与某用户相近“茧房度”的个体称为该用户的“相邻茧房”(adjacent cocoons),将全体用户中“茧房度”最高,即自我重复度最高的、处于“顶部”的用户称为“顶部茧房”(top cocoons)。同时需要强调,这里的“茧房度”是针对用户所生产的内容的重复度来衡量的,而不是接收信息、进行选择信息的“茧房度”。此外,结合前文关于公共性的社会信息中心及其对于个体议题的作用力,提出社交网络中的“社会信息中心”(social information center)这一概念,用于描述在社交网络公共领域中具有中心性、高热度、高流量的公共信息,或者说是公共领域的中心信息。

用户所生产的信息内容的“茧房”不是随意地、彼此隔绝地、异质地形成的,而是在复杂多变的传播生态中朝向某个社会信息中心靠拢,这导致“茧房”之间表现出同质化倾向。个体在内容生产时不是陷入任意的“茧房”,而是在趋同于某种社会信息中心的过程中增强自身的“茧房度”;或者说,用户的“茧房度”越高,其所生产的信息内容就越趋同于社会信息中心。据此,本文提出假设1:

假设1:用户个体的“茧房度”越高,该用户所生产的信息内容就越趋同于社会信息中心的内容。

如果假设1成立,即个体的“茧房”随着“茧房度”的提高而呈现出越来越同质化的趋势,并且都趋同于社会信息中心,那么“茧房度”最高的用户也是最接近社会信息中心的用户。微博这类社交媒体中的“顶部茧房”尽管不能等同于社会信息中心,却最接近社会信息中心。换言之,大家在朝向“广场中心”奔跑的过程中,尽管跑在最前面的“排头兵”并非直接的“广场中心”,但由于它是最靠近中心的,这样大家就会显现出朝向“排头兵”奔跑的态势。据此,本文提出假设2:

假设2:用户个体的“茧房度”越高,其与“顶部茧房”的内容相似度越高。

既然个体的“茧房度”越高就越趋同于社会信息中心的内容特征,也越是趋同于“顶部茧房”,那么就意味着,越是高“茧房度”的个体,越是和某种共同“标的”、某种有限的共通“茧房模板”越长越像,也即越是削弱自身的异质性。所以可以推论:高程度的“茧房”彼此更为相似,低程度的“茧房”由于可以保留更多异质性和独特性因而彼此之间的相似度会更低,或者只是趋于某种随机的个体相似度。据此,本文提出假设3:

假设3:用户个体的“茧房度”越高,则彼此之间的内容发布主题越相似。

如果假设1和假设2都成立,把假设2中的“用户茧房度”根据假设1替换为“用户个体与社会信息中心的内容相似度”,则可提出假设4:

假设4:用户个体与社会信息中心的内容相似度越高, 则它与“顶部茧房”的内容越相似。

如果假设1和假设3成立,把假设3中的“用户茧房度”根据假设1替换为“用户个体与社会信息中心的内容相似度”,则可提出假设5:

假设5:用户个体与社会信息中心的内容相似度越高, 则它与“相邻茧房”的内容越相似。

上述5个假设具有一定的内部逻辑结构。其一,假设1的成立是提出假设2的基础,假设1和假设2的成立共同作为提出假设3的基础,在假设1和假设2的基础上可以提出假设4,在假设1和假设3的基础上可以提出假设5。其二,假设1、假设2、假设3为核心论点,而假设4和假设5则是辅助性的推论。前3个假设支撑着关于“茧房趋同化”的根本命题,而后2个假设是前3个假设的内在自然的延伸,而且也使得后文的路径分析形成更为整体化的结构。

(二) 样本选取与数据预处理

根据新浪微博数据中心发布的《2018 微博用户发展报告》,新浪微博月活跃用户 4.62 亿,月阅读量过亿领域达 32 个。可见,微博已经成为目前中国最具影响力的代表性社交媒体平台之一,故本文选取这一平台作为数据抓取的对象。采取成本相对较低的多阶段抽样,抓取工具为八爪鱼软件以及自行采取 selenium 模块编写的网络爬虫。首先,从新浪微博首页 47 个内容版块中,每个版块每天早晚各抓取 1 次,持续 1 个月,获得 65 650 条“种子”帖。种子帖广泛而大致均匀地分布在这 47 个版块中,如图 1 所示。

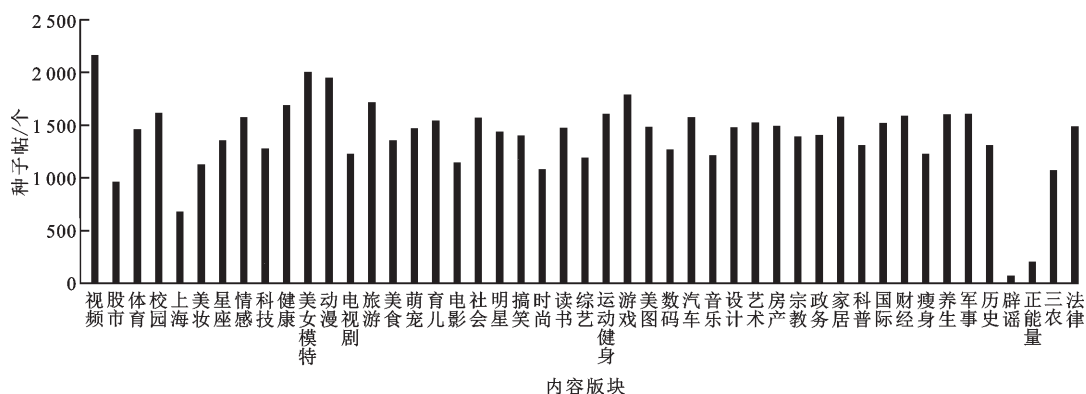


图 1 新浪微博种子帖在 47 个版块中的分布

以种子帖为基础做第一阶段“滚雪球”抽样的扩散, 抓取并采集这些种子帖共 5 053 998 条评论及其评论者。通过“滚雪球”抽样得到的评论者和原种子帖发布者经清洗、去重后, 形成 3 501 153 个用户的初始库。从库中

随机选取 87 739 个有效用户(剔除无效用户之后剩下的种子数),由此开展第二阶段的“滚雪球”抽样,从每人的前 5 页被关注者中随机抽出 2 名。对上述用户横向比较口径进行统一。其一是时间选取在 2018 年 1 月 1 日—

2018年12月31日这一年发布的帖子,其二是每个用户随机选取在上述一年发布的300条帖子,少于该数量的用户不纳入分析。然后,随机选取满足条件的20 000个用户作为最终“粉丝”。以剩下的这些用户作为研究样本,其“粉丝”规模、发微博数、关注他人数量等指标分布具有广泛性与代表性。样本用户的特征分布见表1。

表1 样本用户的特征分布

| 用户特征 | 观测值数 | 均值 | 中值 | 标准差 | 极小值 |
|--------|--------|-------------|-------|---------------|-----|
| “粉丝”规模 | 20 000 | 593 541.061 | 1 193 | 2 426 393.732 | 1 |
| 关注他人数量 | 20 000 | 731.381 | 398 | 1 171.223 | 0 |
| 发微博数 | 20 000 | 19 150.450 | 9 049 | 34 090.372 | 0 |

(三)帖子的向量化转换

把帖子转换为向量后计算两条帖子之间的余弦相似度。对于帖子内容的向量化,本文采用在业界和学界有广泛成熟应用、快速和稳定的Word2vec词向量及其平均池化的方式进行。Word2vec是Mikolov等^[30]在2013年提出的浅层神经网络语言模型,通过无监督训练方式将文本数据转化为低维实数向量,将语义、语法信息经过语言模型训练,投射到若干维的向量空间中,以此完成文本的结构化处理。Word2vec不仅可以在百万数量级的词典和上亿的数据集上进行高效训练,还能有效解决文本高维稀疏性和忽略潜在语义的问题。对于句子采用Word2vec词向量方法经简单等权平均而转换得到的句子向量具有良好的效果。Shen等^[31]的研究将简单词向量模型(SWEM),即对上述的词向量进行等权求平均向量的方法,与循环和卷积网络进行比较发现,在大多数情况下SWEM表现出更高性能。

本文使用Gensim模块编写代码进行Word2vec中CBOW模型的设置与训练。Word2vec使用的语料采用课题组自行抓取的中文语料库,来源包括媒体新闻库、网络论坛帖子、经典名著和文学文本等,训练结果包括5 830 979个词汇的嵌入式表示,效果良好。

对于内容相似度的计算首先需要把微博帖子的短文本转为向量。对样本中的帖子采取当前常用而效果优良的Jieba软件进行分词,然后利用Word2vec将分词后的每一个词转换成一个300维的词向量,对这些词的词向量经过平均池化后得到该帖子、短文本的语句向量。

(四)变量构建与测量

“茧房度”(DC),在实际操作中具体表现为抓取到的每个用户帖文之间的同质化程度。如果一个用户所发布的帖文都是高度相似的,比如都是美食类信息,很少有经济、体育等其他类的信息,那么可以认为用户内容发布的信息“茧房”同质化程度非常高。余弦相似度更好地适应于高维数据和稀疏数据,在文本挖掘中也是常用的经典方法:

$$DC = n^{-1}(n-1)^{-1} \sum_{i=1}^n \sum_{j=1}^n R(t_i, t_j) \quad (1)$$

其中,计算任意两条帖子 t_i 和 t_j 之间余弦相似度的方式表示为 $R(t_i, t_j)$, i, j 表示任意两条帖文, $i \neq j$ 。

与社会信息中心的内容相似度(S_{sc}),指的是个人发布信息的帖文与抓取到的总体数据中热度最高的帖子之间的相似程度。热度的计算方式是把帖子的转发数、点赞数、评论数这3个指标分别经由 $\log_2(x+1)$ 转换,然后进行Min-Max归一化处理各自转为 $[0, 1]$ 区间内的值,最后对归一化之后的3个指标求等权平均值作为该帖子热度。本文将社会信息中心指标的范围规定为总体帖子中热度最高的前10万条之内。由于所选取的是2018年全年的帖子,样本数高达2 011.74万条,因此选择热度最高的前10万条可以充分代表“热帖”所讨论的内容,反映社会公共领域关注的议题,体现出社会信息中心的内容特征,既具有充分的规模,不至于只局限于过度狭小的议题取样范围,又保持了充分的“顶部”热度。其计算公式为

$$S_{sc} = n_1^{-1} n_2^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} R(t_i, t_j) \quad (2)$$

其中, n_1 代表个体的微博帖文数, n_2 代表处于全体样本中热度最高的若干条帖子,这里取最高的10万条帖文。

与“顶部茧房”相似度(S_{tc}),指的是个人发布信息的帖文与抓取到的总体数据中“茧房度”最高的前 k 个用户之间的相似程度,本文取 $k=30$ 。一方面,30个用户在全体20 000个样本中是非常少的一部分,可以充分表现出顶部区别于中部和尾部的特殊性,这个值不能过大,否则就扩大和含糊了顶部的边界;另一方面,当 $k \geq 30$ 的时候,在社会科学意义上一般可以认定为不属于小样本,统计特征和检验具有稳定性。计算公式为

$$S_{tc} = k^{-1} \sum_{i=1}^k [n^{-2} \sum_{j=1}^n \sum_{l=1}^n R(t_i, t_j)] \quad (3)$$

其中, k 取30,即取全体用户中“茧房度”最高的30个用户。

与“相邻茧房”相似度(S_{ac}),指的是个人帖文与其“茧房度”最为相近的 k 个用户的内容相似度。本文计算中,这里的 k 同样取30。

$$S_{ac} = k^{-1} \sum_{i=1}^k [n^{-2} \sum_{j=1}^n \sum_{l=1}^n R(t_i, t_j)] \quad (4)$$

基于文献回顾和变量测量,本文拟通过如图2所示的检验模型来分析“茧房度”、与社会信息中心的内容相似度、与“顶部茧房”相似度以及“相邻茧房”相似度4个变量之间的关系。前文5个假设具体化的可检验形式整理如下:

H1:“茧房度”正相关于与社会信息中心的内容相似度。

H2:“茧房度”正相关于与“顶部茧房”相似度。

H3:“茧房度”正相关于与“相邻茧房”相似度。

H4:与社会信息中心的内容相似度正相关于与“顶部茧房”相似度。

H5:与社会信息中心的内容相似度正相关于与“相邻茧房”相似度。

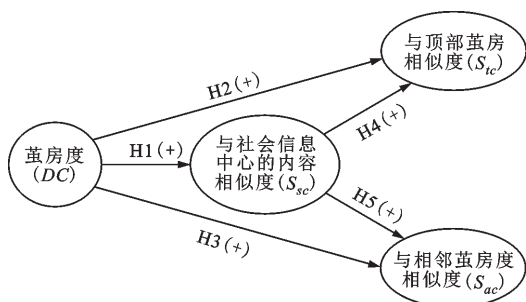


图2 本文假设的检验模型

三、研究结果

(一)相关性分析

首先,对“茧房度”、与社会信息中心的内容相似度、与“顶部茧房”相似度、与“相邻茧房”相似度等4个变量进行相关性分析,探讨其之间是否存在一定程度上的共变性。此处采用皮尔逊相关分析方法,并进行双尾检验,分析结果见表2。从中可见,“茧房度”和与社会信息中心的内容相似度之间呈显著的正相关,和与“顶部茧房”相似度之间呈现显著的正相关,和与“相邻茧房”相似度之间呈现显著的正相关;与社会信息中心的内容相似度和与“顶部茧房”相似度呈现显著的正相关,和与“相邻茧房”相似度呈现显著的正相关。总体而言,这4个变量之间是存在某种共变性的,这也为后续研究在模型中进行路径分析和比较提供了参考。

表2 4个主要变量之间的相关性检验结果

| 变量 | DC | S_{sc} | S_{ic} | S_{ac} |
|----------|---------|----------|----------|----------|
| DC | 1 | | | |
| S_{sc} | 0.689** | 1 | | |
| S_{ic} | 0.643** | 0.528** | 1 | |
| S_{ac} | 0.814** | 0.714** | 0.577** | 1 |

注:1.**表示1%的显著性水平。

2.观测值数为20 000。

(二)路径分析

其次,使用AMOS软件进行路径分析。在进行路径分析时,需要先将数据进行正态化处理再纳入模型中。计算方式采取最大似然法(maximum likelihood, ML)。模型的主要拟合结果见表3。从中可见,模型的 $RMSEA=0.006<0.008$, $CFI=0.998>0.9$, $TLI=0.981>0.9$ 。虽然有部分指标略有超出,但都在可接受范围之内,整体来看模型拟合度优良,可以被接受。

表3 模型拟合结果

| 指数形式 | 检验指标 | 数值 |
|--------|-------|---------|
| 绝对拟合指数 | GFI | 0.998 |
| | AGFI | 0.976 |
| | RMSEA | 0.064 |
| | NFI | 0.998 |
| | RMR | 0.008 |
| 相对拟合指数 | NFI | 0.998 |
| | RFI | 0.988 |
| | IFI | 0.998 |
| | TLI | 0.981 |
| | CFI | 0.998 |
| 信息指数 | AIC | 145.277 |
| | CAIC | 227.981 |

根据之前提出的假设构想和变量思考,将捕捉到的数据进行各项计算处理后纳入模型,得到了表4所示的分析结果。从中可见,假设H1“茧房度”正相关于与社会信息中心的内容相似度获得支持($\beta=0.532, P<0.001$);假设H2“茧房度”正相关于与“顶部茧房”相似度获得支持($\beta=0.689, P<0.001$);假设H3“茧房度”正相关于与“相邻茧房”相似度获得支持($\beta=0.614, P<0.001$);假设H4与社会信息中心的内容相似度正相关于与“顶部茧房”相似度获得支持($\beta=0.161, P<0.001$);假设H5与社会信息中心的内容相似度正相关于与“相邻茧房”相似度获得支持($\beta=0.291, P<0.001$)。总体而言,检验结果与前文对于5个假设及其内部关系的阐述相吻合。

表4 路径分析结果

| 研究假设 | 路径系数 β | 标准误 | t值 | P值 |
|------|--------------|-------|---------|--------|
| H1 | 0.532 | 0.006 | 83.111 | <0.001 |
| H2 | 0.689 | 0.004 | 155.263 | <0.001 |
| H3 | 0.614 | 0.005 | 63.649 | <0.001 |
| H4 | 0.161 | 0.006 | 25.127 | <0.001 |
| H5 | 0.291 | 0.005 | 63.649 | <0.001 |

以上结果显示:首先,假设H1也即本研究的关键假设得到了验证,“茧房度”实际上显著地影响了与社会信息中心的内容相似度;其次,“茧房度”也和与“顶部茧房”相似度、与“相邻茧房”相似度两个变量之间均存在明显关系,足见“茧房度”无论是对于其他三个变量还是对整个模型都具有基础性意义;最后,与社会信息中心的内容相似度(S_{sc})的强化对于与“顶部茧房”相似度(S_{ic})和与“相邻茧房”相似度(S_{ac})都具有关键作用,构成显著且关键的中介路径。

四、结论与讨论

本文基于典型社交媒体之一的新浪微博样本,对于“个体所陷入的‘茧房’是越长越像而非各有不同”的观点,从三个关键子假设及其内部关系进行考察与检验,

主要结论如下。第一,社交空间中各个用户所生产的信息内容的“茧房”会因陷入一种公共性的“回音室”中而呈现出同质化趋势,用户个体的“茧房度”越高则表示其受限于这种公共“茧房”的程度越深,其内容与社会信息中心的内容越相似,并使得各个“茧房”在趋向共同标的的过程中发生个性化的消减、趋同性的增强。第二,个体在“茧房度”增高且与社会信息中心趋近的过程中,会导致与“顶部茧房”的内容相似度的提高。第三,个体在趋近“顶部茧房”和社会信息中心的过程中,彼此之间会变得越来越相似,在实践中具体表现为个体“茧房”与其“相邻茧房”之间内容相似度的提高。第四,由于以上三种趋同路径具有统一性和一致性,所以形成个体的“茧房度”、与社会信息中心的内容相似度、与“顶部茧房”的内容相似度、与“相邻茧房”的内容相似度这4个变量之间存在同步共变的关系。

互联网时代赋予了传播学新的使命和机遇,以连接和网络为核心的传播逻辑正在形塑着如今的媒体生态,用户的自主权和能动性随着技术的发展大大扩增。传统研究视野下,许多学者分别从信息窄化和意见极化的角度关注到信息“茧房”现象,并认为它是作为受众异质化、分散化、圈层化的象征之一。实际上,用户在日常生活中以不同的传播逻辑构筑起自己的信息接收网络,影响其信息获取和信息生产的因素丰富多样。同时,只有更多的用户共同关注并讨论同一个话题,才能够形成网络社会中的公共舆论场。本文从用户生产的信息“茧房”同质化现象切入,发现随着用户所生产的信息内容的“茧房度”增高,各“茧房”之间呈现出彼此趋同化而非趋异化的格局。同时采用实证分析加以验证,在文献回顾的基础上,通过构建信息“茧房”趋同路径模型来阐述它们之间的复杂动态关系。但新的结论也触发了新的需要探讨的问题,这种同步共变的关系会使网络社会产生怎样的变化?这种同质化现象背后又映射出哪些现实生活的问题?此外,本文也局限于一个社交平台,即基于微博数据生态下,这些结论是否在其他社交媒体上也能得到支持,期待后续能有进一步的回答。

参考文献

- [1] 桑斯坦. 信息乌托邦: 众人如何生产知识[M]. 毕竞悦, 译. 北京: 法律出版社, 2008.
- [2] 刘强, 赵茜. 算法中选择的同化与异化: 国外回音室效应研究20年述评与展望[J]. 新闻界, 2021(6): 29-38.
- [3] KOSSINETTS G, WATTS D. Origins of homophily in an evolving social network[J]. Journal of Sociology, 2009, 115(2): 405-450.
- [4] ALSTYNE V, BRYNJOLFSSON E. Global village or cyber-balkans? Modeling and measuring the integration of electronic communities[J]. Management Science, 2006, 46(2): 201-204.
- [5] 汤景泰, 陈秋怡. 意见领袖的跨圈层传播与“回音室效应”: 基于深度学习文本分类及社会网络分析的方法[J]. 现代传播(中国传媒大学学报), 2020(5): 25-33.
- [6] 姚文康. 聚合类新闻客户端的“信息茧房”效应及反思: 以“今日头条”为例[J]. 传媒论坛, 2020(3): 151-153.
- [7] 蒋焱川. 多Agent群体行为中的扩散趋同综述[J]. 模式识别与人工智能, 2009(6): 877-883.
- [8] PINTO S, ALBANESE F, DORSO C, et al. Quantifying time-dependent media agenda and public opinion by topic modeling[J]. Physica A: Statistical Mechanics and its Applications, 2019(524): 614-624.
- [9] NECHUSHTAI E, SETH L. What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations[J]. Computers in Human Behavior, 2019(90): 298-307.
- [10] LEE K. The effect of the internet on homogeneity of the media agenda: a test of the fragmentation thesis[J]. Journalism & Mass Communication Quarterly, 2007, 84(4): 745-760.
- [11] 宋建武, 张琦悦, 王译瑶. 互文视阈下的媒体公众号: 同质化表征及成因[J]. 编辑之友, 2016(8): 47-51.
- [12] MIKAL P, RICE E, KENT G, et al. 100 million strong: a case study of group identification and deindividuation on Imgur. com[J]. New Media & Society, 2016, 18(11): 2485-2506.
- [13] HOSANAGAR K, FLEDER D, LEE D, et al. Will the global village fracture into tribes? Recommender systems and their effects on consumer fragmentation[J]. Management Science, 2014, 60(4): 805-823.
- [14] AIROLDI M, BERALDO D, GANDINI A. Follow the algorithm: an exploratory investigation of music on YouTube[J]. Poetics, 2016(57): 1-13.
- [15] PEDRONI M, FRANCESCA P, SIMONE C. My friends are my audience: mass-mediation of personal content and relations in Facebook[J]. Observatorio, 2014, 8(3): 97-113.
- [16] HANSEN D, SHNEIDERMAN B, SMITH M. Visualizing and interpreting Facebook networks with NodeXL: insights from a connected world [M]. Cambridge: Morgan Kaufmann Publishers, 2011.
- [17] SU J, LIU B, LI Q, et al. Coevolution of opinions and directed adaptive networks in a social group[J]. Journal of Artificial Societies & Social Simulation, 2004, 17(2): 1-4.
- [18] JUN Y, MENG R, JOHAR G. Perceived social presence reduces fact-checking[J]. Proceedings of the National

- Academy of Sciences of the United States of America, 2017, 114(23):17-57.
- [19] MCEWAN B, CARPENTR J, HOPKE E. Mediated skewed diffusion of issues information: a theory[J]. Social Media+Society, 2018, 4(3):1-4.
- [20] MIKAL P, RICE E, KENT G, et al. Common voice: analysis of behavior modification and content convergence in a popular online community[J]. Computers in Human Behavior, 2014(35):506-515.
- [21] NIKOLOV D, LALMAS M, FLAMMINI A, et al. Quantifying biases in online information exposure[J]. Journal of the Association for Information Science & Technology, 2019, 70(3):218-229.
- [22] QIN Z, YOU Z, JIN H, et al. Homophily-driven evolution increases the diffusion accuracy in social networks[J]. IEEE Transactions on Network Science and Engineering, 2020, 7(4):2680-2692.
- [23] SCHEUFELE A, NISBET C. Commentary: online news and the demise of political disagreement[J]. Annals of the International Communication Association, 2013, 36(1):45-53.
- [24] FLETCHER R, NIELSEN K. Are news audiences increasingly fragmented? A cross-national comparative analysis of cross-platform news audience fragmentation and duplication[J]. Journal of Communication, 2017, 67(4):476-498.
- [25] BAR-GILL S, GANDAL N. Online exploration, content choice & echo chambers: an experiment[R]. CEPR Discussion Paper, 2017:1-32.
- [26] LEE K. Effect of the Internet on homogeneity of the media agenda: a test of the fragmentation thesis[J]. Journalism & Mass Communication Quarterly, 2007, 84(4):745-760.
- [27] 程士安, 沈恩绍. 数字化时代组织传播理论的解释与重构:以科技进步与传播规律的演进为视角[J]. 新闻大学, 2009(2):119-124.
- [28] MIN Y, JIANG T, JIN C, et al. Endogenetic structure of filter bubble in social networks[J]. Royal Society Open Science, 2019, 6(11):1-11.
- [29] WEBSTER G, KSIAZEK B. The dynamics of audience fragmentation: public attention in an age of digital media [J]. Journal of Communication, 2012, 62(1):39-56.
- [30] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[C]//Proceedings of the 26th International Conference on Neural Information Processing Systems. New York:Curran Associates Inc., 2013:3111-3119.
- [31] SHEN D, WANG G, WANG W, et al. Baseline needs more love: on simple word-embedding-based models and associated pooling mechanisms[C]//56th Annual Meeting of the Association for Computational Linguistics. Melbourne:Association for Computational Linguistics, 2018:440-450.

(编辑:张园,高原)

Different Approaches Lead to the Same Goal: The Homogenization of Information Cocoons of User-Generated Content on Social Media

—An Empirical Research Based on Sina Weibo

XU Xiang¹, AO Ziqi¹, SHI Jingyuan¹, WANG Xiaohong²

1. College of Arts and Media, Tongji University, Shanghai 200092, China

2. School of Journalism, Fudan University, Shanghai 200433, China

Abstract In order to confirm whether the information cocoon room of user content production is homogeneous or heterogeneous, we investigated and tested the sample of Sina Weibo, one of the typical social media in China. This study demonstrates that with the deepening of degree of cocoon, the more the individual cocoon room converges. The higher the cocoon room degree of users, the more they will converge to some kind of public cocoon room or social central information, the more they will wear off their uniqueness and show convergence, and the cocoon room between individuals will tend to be assimilated. The improvement of degree of cocoon is accompanied by the improvement of content similarity with top cocoons and adjacent cocoons. The convergence path of the kinds of cocoon room above is consistent, and the four change synchronously. Although the phenomenon of cocoon room assimilation is contrary to intuitive experience or popular views, it shows the realistic logic of user content production and social communication, and it can provide further reference for network information diffusion and its structure.

Keywords social media; user-generated content; information cocoon; users' homogenization; social information center; content similarity