

· 学者界面 ·

# 基于大数据技术的社会化媒体传播机制探析

**内容提要** 当前, 社会化媒体的传播机制相对传统媒体是否出现了新的特点? 本文基于大数据技术, 对社会化媒体传播机制进行研究后发现, 这些新特点包括: 从交往机制看, 体现出一般社交行为向“社交网络”扩展延伸的特点; 从联结机制看, 体现出搭建社会化媒体完整网络系统的特点; 从扩散机制看, 展现出信息底层扩散模式的特点。

**关键词** 大数据 社会化媒体 传播机制

文/宋凯 刘叶子

用微博分享所见所闻、用微信与朋友沟通交流占用人们大量时间, 社会化媒体已经成为人们生活的一部分。社会化媒体作为用户彼此间用来分享信息、意见、见解、经验和观点的工具和平台, 缩短了人与人之间的距离, 便利了用户的沟通与交流。社会化媒体的传播机制相对传统媒体是否出现了新特点? 海量数据的出现又为研究社会化媒体传播机制提供了哪些新思路? 社会化媒体平台的传播机制研究可以作为解答这些问题的切入点。

## 一、基于大数据的社会化媒体交往机制

社会化媒体是给予用户极大参与空间的新型在线媒体, 它们建构于互联网技术上, 核心理念是以人为本、开放、互动、共享和关系, 而微博最大的特点是赋予每个个体创造并传播内容的能力, 将人与媒体的关系实现从“信息中介”到“关系网络”的转换, 从根本上改变社会实践的方式<sup>[1]</sup>。社会化媒体的传播机制最重要的特点就是将人的社

会关系引入到信息传播中, 社会关系成为影响信息传播效果的重要因素。

### (一) 内容的自主生产

UGC(用户产生内容)不断演变, 由最初的文本转化为图片及音视频内容, 用户可以随时随地在网络上分享内容, 由此产生了海量的用户数据, 并且由于人人都是传播者, 生产内容的自主性、共享性不断增强, 衍生出了大量关系数据及内容数据。这些数据不同于以往单纯的数字, 它们声色结合、图文并茂, 体现出非结构化的特征。比如, Facebook用户每天共享的东西超过40亿, Twitter每天处理的数据量超过3.4亿; 而每分钟Tumblr博客作者会发布2.7万个新帖子, Instagram用户会共享3600张新照片<sup>[2]</sup>。用户不仅会生产知识内容, 同时, 用户通过社交媒体留下的“行为数据”也是社交媒体平台最为宝贵的资源, 社交平台会不断利用这些“行为数据”攫取用户阅读习惯, 不断给用户推送相关知识, 而用户的互动同时也会生产新的内容, 再由社交媒体

平台推送传播, 形成一条完整的用户内容生产、社会交往与知识传播的链条。用户自主生产的是行为数据和知识内容的结合体, 社会化媒体上的用户以自主生产的内容为传播载体与他人进行互动和交往。

### (二) 消费的自我选择

受众不可能对社会化媒体平台上的所有信息全盘接受, 受众对信息的接受过程是一个能动的选择过程。传播学者经过研究发现, 受众会有选择地接触某一媒介, 其往往只选择那些能加强自己信念的信息, 而拒绝那些与自己固有观点相抵触的信息, 受众通过大众媒介来加强自己的观念而不是改变自己的观念<sup>[3]</sup>。施拉姆曾经提出一个人们为什么选择某种媒介的公式: 选择的或然率=报偿的保证/费力的程度。通过大数据技术我们发现, 在社会化媒体平台上用户接受自己社交圈内的信息, 更愿意通过数字阅读消耗时间, 我们在社交网络上浏览的信息来自同质化网络远大于异质化网络或者自己主动搜索的信息。德弗勒

就曾认为：影响受众选择性注意的因素包括认识结构上的个人差异和社会成员类型等，比如有牢固社会关系的人可能更关注与朋友和家人利益相关的问题。

在基于社会关系的社会化媒体平台上，用户更愿意为社会关系买单这个特点显得尤为显著：用户更加关注自己社交网络中个人传递的信息，比如微信朋友圈的盛行，用户更加愿意接受和选择相信来自朋友的信息，由此我们看到口碑传播在营销中的重要之处。

### （三）分发的自主导向

著名传播学研究专家拉斯韦尔曾指出，传播过程是一个从传播者到受众的单一的线性传播过程。而在社会化媒体平台上的传播过程是一种非线性、圈层、超链接的传播和连接方式。基于社会关系的社会化媒体平台是圈群式的，每个人有着不同的社交圈，比如在微信和QQ等社会平台上我们有着不同的好友群或者QQ群，我们往往会发现每个社交圈的信息具有同质性，不同社交圈之间的信息异质化程度高。从宏观角度来看，社会由于个人社会资本的不同又是分层的，社会化平台同样是层级式的，社会化媒体将具有“差序格局”的社会关系移植到互联网上，正是由于这种社会关系向圈群化和层级化发展，传播过程必然是非线性的，这个传播过程错综复杂，每个圈群内部有着不同的信息传播特点，不同的圈群以个体为纽带相互影响和联系。

## 二、基于大数据的社会化媒体的联结机制

如果将社会化媒体视作一个有机

运作的生命体系统的话，联结机制和交往机制则分别完成了这个系统的基础和搭建工作。交往机制代表了系统的雏形——即从一般社交行为向“社交网络”的扩展延伸，为整个社会化媒体的生态系统提供了素材和血液；联结机制则在此基础上，基于六度空间的理论将各个节点的关系进一步深化，并在大数据技术的支持下让连接机制更加有效地发挥作用，让点——线——面的演变路径逐步实现，搭建起完整的社会化媒体的网络系统。

### （一）六度空间理论及其制约条件

六度分隔是指平均来说，社交网络中任意两个素不相识的人之间，最多只需经过六步即可建立相互联系。也就是说最多通过六个人你就能够认识任何一个陌生人<sup>[4]</sup>。1967年，哈佛大学教授斯坦利·米尔格拉姆通过连锁信件传递实验发现，在发出的160封信中有42封平均经过5.5个中间人到达了目标人物。

六度空间理论虽然展现了我们生活在“小世界”的事实，但是仍然有着三方面的残缺：一是权值问题：我们的世界是一个普遍联系的世界，人

们一生肯定会认识很多人，但实际生活中，人与人的关系是有强弱和权重的。基于血缘形成的亲人、基于地缘形成的邻里、基于爱好形成的朋友，而六度空间理论把这些关系都简单归结为联系，没有强弱之分。二是目的和结果问题：在相关实验中被实验者通过六度空间传递的是信息，但是在社会化媒体时代社会化媒体平台想要传递的是联系方式还是朋友关系是值得深思的问题，到达和建立联系的过程会随着目的的不同而发生改变。三是阻尼问题：阻尼也就是传递的成本和激励，根据六度空间理论，人与人建立联系只有六度，但是每一度的阻尼都是难以超越的。在相关实验中，信件的传递看似成本很低，但是那些人愿意传递其实是看在朋友的面子上，这里面的成本是我们常说的“人情”，也就是关系成本，由于阻尼问题导致了160封邮件只有42封到达了目标人物。

### （二）大数据解放六度空间的制约

上文阐述了六度空间理论的制约条件：权值问题；目的和结果问题；阻尼问题。大数据技术的出现完美解决了权值问题，许多基于大数据技术的社交网络分析软件可以对社会关系进行赋值，对关系强弱进行度量。社会化媒体是一种基于社会关系的传播媒介，平台上信息海量存在，这种社会关系并不是简单的是否存在联系，这种联系还存在着强弱的问题。以社交网络分析软件Ucinet为例，在群体的中心性分析中，点度中心性、接近中心性、特征向量中心性、中介中心性是群体中心性的重要指标。其中点度中心性是代表行动者的



直接社会关系，反映的是个人在网络中的直接影响力；接近中心性基于节点之间的距离，接近中心性越高的行动者越处于整体的中心位置；特征向量中心性是分析与相邻节点是否具有高影响力，特征向量中心性越高说明它距离信息源或高扩散的节点越近；中介中心性刻画的是社会网络中节点对信息流动的影响力，如果一个行动者位于许多网络路径上，他就可以控制两个人之间的信息交换。

这些分析行动者中心性的重要指标在基于大数据技术的社会网络分析软件中可以直接赋值表示出来，社会关系在大数据软件中得以量化，关系有了强弱大小之分。在最初的六度空间实验中，传递的内容是信息，而在社会化媒体平台上我们通过大数据技术可以看到传递的其实是一种社会关系，基于不同的社会关系，目的和结果也会不同，通过大数据可以更加清晰地了解各个行动者在传播的过程中处于何种角色、基于何种目的、造成何种结果。阻尼问题受行动者主观意识控制，无论是传统时代还是大数据时代都属无法解决的问题。

### （三）大数据与社交媒体的小世界理论

通过大数据分析技术穷举搜索 Facebook 关系图上的每条链接，Facebook 数据分析专家小组发现，社交网络中的每个人都可由平均 3.5 个“媒介好友”而联系起来，这一发现刷新了之前长期流行的“六度分隔”理论，即世上任何两人皆可通过某条较短的熟人链条连接起来。这说明社交媒体的出现让世界变得越来越小：以前的六度现在一半就够。3.5 这个数字计算

的是“媒介”的数量，而不是分隔度数，平均 3.5 个媒介对应的是“4.5 度分隔”，六度对应的是“5 个媒介”。

## 三、基于大数据的社交化媒体的扩散机制

传播从本质上来说是信息的流动，信息就是构成传播的基本材料。基于大数据的社交化媒体的扩散机制，则要从系统的实际运行角度切入，探讨整个社交化媒体网络的底层运转机制，探究信息的扩散模式，并结合具体案例进行分析。

### （一）社交化媒体的信息扩散特征

简单随机网络这一概念更符合早期互联网研究者们对网络空间的期许——一个绝对平等的、去中心化的数字空间。在这张网络中，每个节点都遵从着绝对的随机性原则，彼此间没有任何差异，其连接关系也是随机生成的，这就意味着所有节点拥有平等机会获得连接，在网络足够大的前提下，理论上几乎所有节点拥有的连接数，都是基本相同的。理想化的随机网络模型并不现实，事实上，即便是互联网的原教旨主义者，也一样在后续的研究中承认，简单随机网络的模型中需要非随机成分优化改良。在此基础上，复杂网络的概念应运而生。

相比简单随机网络，复杂网络的“复杂性”，来源于它提出的枢纽节点的存在。所谓枢纽，即相对于简单随机网络中的普通节点而言，前者拥有更多的连接数，这里的连接数差异并非随机过程中的数量波动，而是明显的落差对比，这种差距已经大到影响节点之间的地位。在复杂网络中，枢纽节点掌握

了大量的连接数，而众多普通节点仅依靠同枢纽节点连接，便可以完成在网络中的生存，获取所需信息。在节点和连接数的关系上，复杂网络遵从幂律分布，即一条平滑的曲线，两端无限延伸。因为复杂网络中，“特征尺度”意义的消解，这一网络模型也被称为无尺度网络。无尺度网络的扩散机制，显然同简单随机网络存在巨大差异。信息的传递效率与其产生的影响力，取决于枢纽节点们的选择，后者成为整个网络的信息传播中枢。与传统的中心化网络不同的是，这些节点扮演的依然是媒介和中转站的角色，他们负责将信息传输到各自所影响的圈层之中。因为在无尺度的网络模型里，信息的传输路径通常就是通过一个又一个的枢纽节点，进而影响扩散到更大的圈层之中，然后身处圈层中的每一个个体，又在其各自的圈层中扮演起枢纽节点的角色，进而完成又一轮的扩散传递。

反馈到现实中，可以将其简单理解为意见领袖二级传播模式的无限复制——就像各类舆情事件在社交媒体的发酵路径一样，最早的信源曝光后，经历一群“大V”们的转发，才形成了大面积的传播，此后，“大V”们围绕其各自的受众圈再进行一次又一次的扩散，而信息也在这样的过程中完成了传播。这种底层信息的运行机制相比简单随机网络要高效许多，后者需要整个网络的所有节点协同完成信息向全网的扩散，而在复杂网络中，枢纽节点承担起了二传手的作用，让信息可以更为高效地流通传递。

### （二）舆情分析与大数据技术融合当前常用的网络舆情分析方法主

要有网络调查方法、基于统计规则的模式识别方法、基于内容挖掘的主题监测方法等<sup>[5]</sup>。处理网络舆情的技术主要是首先建立一套指标体系,再基于网络爬虫的网络数据抓取、数据的预处理(主要有中文分词、特征提取、建立VSM模型等),最后进行数据分析<sup>[6]</sup>。这种网络舆情技术主要是在于舆情处理的准确性,忽略了网络舆情的处理时间。

在大数据时代的网络舆情分析数据集从传统样本扩展到全体数据,目前的大数据技术可以实现24小时全天候自动获取和采集网络上与关注的舆情事件相关的信息,并对信息进行自动分类,情感分析,及时预警等功能;数据分析技术从描述到预测,以往的网络舆情分析只能描述网络舆情事件发展的状态,何时爆发和何时消退的全过程,通过刻画网民在网络上留下的浏览、评论、转发、点赞等行为详细描述网络舆情事件。在大数据时代的网络舆情分析重点不在描述而是在预测,大数据技术可以通过对信息的分析从整体上把握用户的心理特征、信息传播的规律、网络舆情发展态势。对于政府相关部门和企业来说,最重要的是全面把握公众的意见,识别潜伏期舆论是否存在危机,能够未雨绸缪,防范于未然。大数据技术能为下一步决策提供数据依据,提高政府和企业对危机事件的应对速度。

### (三) 大数据分析下的舆情扩散机制

由于社会化媒体的无门槛性,网民可以在微博、微信、QQ等各种社会化媒体平台上针对任何公共事件发表自己的言论,网民在这一过程中会形成众多的数据,这些数据多是非结构性的,

所以舆情事件的监测和分析需要大数据技术的加入。大数据技术可以提高网络舆情数据的处理效率、网络舆情分析的准确度和精确度和网络舆情处理能力,这对分析社会化媒体舆情产生的原因、发展的过程和预测舆情发展趋势都有着重要的意义。

微博是社交媒体平台的典型代表,同时由于其平台数据的开放性,利用大数据技术手段可以绘制出完整的社会化媒体信息的传播路径、舆情事件各个时期的特点,同时能够体现传播者之间的关联以及其在整个传播链条中的作用。以青岛大虾为例,其微博的传播就是一张典型的“混沌分形图”,美国科学家曼德勃罗特用计算机迭代出来的混沌分形图被称为“上帝的指纹”。舆情事件经过爆发期后,通过“财经网”等几个微博“大V”用户的转发后不断发酵,进入舆情事件的发展期,随着时间的推移,舆情事件不断地衰弱,进入衰弱期。

舆情事件微博的传播一般还具有典型的幂率分布特征,服从“长尾理论”和“马太效应”,这也是现代社会化媒体传播机制中最主要的特点。就微博的信息传播服从“长尾”理论来说,由于参与传播的个体是随机和海量的,且复杂网络的无标度性使网络节点不断增长,微博信息的增长使节点比后来者有更多机会链接,不断转发使链接数量增加,产生无数的节点嵌套结构,网络结构的这种分形特点,使得网络信息的传播结构趋向于“长尾”的无限结构。就微博信息的“马太效应”来说,80%的转发数都来自于占有所有转发用户数20%的微博大V,2/8定律较为明显<sup>[7]</sup>。

社会化媒体信息传播的另外两个特性:自相似性和标度不变性。就“自相似性”来说,信息在传播过程中结构尺度一层一层缩小,同时具有嵌套重复的特点。同一信息在传播的层次和尺度上不同,但是局部和整体的传播结构特征是相似的,可以理解为整体与局部的传播特征对称性。就“标度不变性”来说,在微博信息的传播过程中,每一次转发的延伸组合构成了整个传播链条的“长尾”,每个转发者构成的“小长尾”其形态、不规则性对比整个传播链条是一致的,可以说是社会化媒体传播机制在大数据技术驱动下的一次微观与宏观的映射。

以上三种在大数据技术分析下的传播特性对于认知社会化媒体传播机制具有重要的启示意义,对于社交媒体看似有限的区域和无限的传播个体,通过大数据可视化技术让我们能从一个不断发展、全新的角度去理解信息传播机制的概念。一方面可以利用图形反馈信息传播的发展过程;另一方面,从局部到整体的分形图研究可以完美展示社会化媒体信息传播机制的幂率及自相似性特征。 (作者宋凯是中国传媒大学移动互联网与社会化媒体研究中心副研究员;刘叶子是中国传媒大学互联网信息专业社会化媒体方向硕士研究生)

#### 【注释】

- [1] 乔文娟.浅析微博时代公众话语权的提升[J].新闻世界,2011.
- [2] eNet&Ciweek:社交关系与大数据的结合,催生真正的互联网与金融领导者[N].互联网周刊,2015.
- [3] 段鹏.传播学基础[M].北京:中国传媒大学出版社,2013.
- [4] 艾伯特·巴拉巴西链接——商业、科学与生活的新思维[M].浙江:浙江人民出版社,2013:37.
- [5] 李佳师.大数据:观念影响速度[J].中国电子报,2013(5).
- [6] 李金海.何有世.熊强.基于大数据技术的网络舆情文本挖掘研究[J].情报杂志,2014(10):2.
- [7] 朱海松.读者来稿:杜蕾丝微博研究. <http://chuansong.me/n/137318>,2013-8-3.