龙脉图书

图書館學 情報學 檔案學理論與實践系列叢書

名家视点 Mmaster viewport 第4辑

微博与信息传播

《图书情报工作》杂志社 编

WEIBO YU XINXI CHUANBO

海洋出版社

微博与信息传播

《图书情报工作》杂志社　编

海洋出版社
2013 年 · 北京

图书在版编目（CIP）数据

微博与信息传播/图书情报工作杂志社编著．—北京：海洋出版社，2013.10
（名家视点．第4辑）
ISBN 978-7-5027-8656-4

Ⅰ．①微… Ⅱ．①图… Ⅲ．①互连网络-传播媒介-文集
Ⅳ．①G206.2-53

中国版本图书馆CIP数据核字（2013）第215121号

责任编辑：杨海萍
责任印制：赵麟苏
海洋出版社 出版发行
http://www.oceanpress.com.cn
北京市海淀区大慧寺路8号 邮编：100081
北京旺都印务有限公司印刷 新华书店北京发行所经销
2013年10月第1版 2013年10月第1次印刷
开本：787 mm×1092 mm 1/16 印张：20.25
字数：477千字 定价：45.00元
发行部：62132549 邮购部：68038093 总编室：62114335

《名家视点丛书》编委会

序

由《图书情报工作》杂志社编辑、海洋出版社出版的《名家视点：图书馆学情报学档案学理论与实践系列丛书》第4辑即将付梓问世。作为我担任《图书情报工作》杂志社社长、主编后经手策划的第一套丛书，我很高兴看到，经过相当长时间的讨论、选题、编辑、加工、出版等一系列环节，第4辑共5本书，就要正式出版。我有种释然的感觉，又觉得有必要多说几句话。

近些年，我们所处的信息环境和文献情报领域发生了非常重大的变化。大英图书馆2008—2011年的战略规划指出：我们所处的环境在过去的二十年里发生的变化超过了过去两百年的变化（初景利、吴冬曼．国际图书馆发展趋势调研报告（一）：环境分析与主要战略．国家图书馆学刊，2010年第1期）；美国一位学者Scott Nicholson也曾提出：图书馆界在过去五年的变化超过了前面一百年的变化，而未来五年的变化将使过去五年的变化微不足道（张晓林．颠覆数字图书馆的大趋势．中国图书馆学报，2011第5期）。

我们需要敏感地认识到这种变化，并积极地应对变化，直面变化所带来的挑战。变化是永恒的（change is constant），但变化也是机会。没有一个学科、一个领域不受快速发展的信息技术所影响，不受快速变化的信息环境所影响。文献情报工作在这种大变革的环境下很可能受到的冲击最大，但也可能是孕育的机会最多的领域。关键是，我们能不能抓住变化的机会，寻求新的业务生长点和自我创新发展的路径。

图书馆学、情报学、档案学的研究者、从业人员、教师、学生和管理者，必须从自身业务上的例行事务中跳出来，睁大眼睛看世界，跟踪和了解国际国内学界业界正在思考的问题，正在发生的变化，正在设计的未来路线。近年来文献情报及相关领域发生的变化可以从《图书情报工作》每年发表的众多文章中感受到这种律动，也可从我们精选的部分文章编辑出版的这套丛书可见一斑。无论是作为图书馆服务的热点的学科服务、知识服务，还是与文献情报有密切关系信息环境和信息化的微博、电子政务、电子商务，都在经历着变革与创新，而正是这种变革与创新不断地推动着文献情报工作及相关领域工作的不断深化和不断向前发展。

我们编辑的这套丛书共5本，分别为《知识服务的现在与未来》、《学科

服务进展与创新》、《微博与信息传播》、《电子政务研究与实践进展》、《电子商务研究与实践进展》，基本都是从《图书情报工作》2009 年到 2013 年初所正式发表的文章精选出来的。5 个主题所研究的问题各有侧重，但都注重理论与实践的结合，体现了作者对相关问题的理论思考和实践探索，反映了当前业界学界对这些问题的研究水平和业务进展。相信会对广大读者有一定的帮助，或具有一定的启示作用。他山之石，可以攻玉。我们也都需要通过学习、交流和借鉴，相互沟通，取长补短，共同成长，共同提高。《图书情报工作》是严谨的学术期刊。作为半月刊，每年发文在 700 篇左右（来稿有 7000 篇左右），同时我们还创办了纯网络的电子期刊《知识管理论坛》（原名《图书情报工作网刊》）。这么多的文章全部阅读完，还是有些困难的。为此，我们选择了 5 个专题，从大量的发表的文章中筛选出一些质量好、有特色的文章，编辑了一个专辑 5 本书。读者可以选择其中感兴趣的主题阅读相关的文章，并追踪阅读和利用该领域更多的研究成果与实践进展。

这是自 2009 年《图书情报工作》杂志社与海洋出版社首次合作出版第 1 辑（4 本）、2011 年出版第 2 辑（5 本）、2012 年出版第 3 辑（4 本）之后的再度合作。我们希望通过对《图书情报工作》所发表的文章的精华部分以书的形式出版，形成对这些研究成果的再利用，更充分地发挥这些研究成果的价值和影响力，为读者提供增值的服务，使这些论文的学术思想、理论创新、实践经验、专业成就得到最大限度地利用。

感谢本丛书的多位作者为丛书所提供的重要的科研成果与实践创新案例，这些成果尽管只是《图书情报工作》发表的，但也一定程度上代表了国内这些领域最新的研究成果和取得的学术成就，为读者了解、学习、借鉴和应用这些成果提供了有价值的参考源，并在此基础上进行深入的研究与探索，不断深化所研究的问题，不断创造出更多更好的成果。

丛书的出版，是《图书情报工作》杂志社、海洋出版社和广大作者共同努力的结果，是我们三方共同奉献给业内广大读者的一份礼物。感谢本专辑的作者，感谢海洋出版社。但愿本专辑的出版，能对图书馆学情报学档案学的相关理论研究与实践创新有所裨益、有所推动，体现出其应有的社会价值，为人们的学习、研究、实践提供必要的支持，为发展壮大我们的学科，为图书情报服务的持续创新，做出应有的贡献。

初景利

《图书情报工作》杂志社社长、主编

2013 年 7 月 3 日于中关村

目　　次

基　础　篇

技　术　篇

用　户　篇

实　务　篇

综　合　篇

基　础　篇

微博客信息传播结构、路径及其影响因素分析*

袁　毅

（华东师范大学商学院　上海 200241）

摘　要　以新浪微博客为研究平台，采集事件传播路径中的用户属性数据及行为数据，利用社会网络分析软件绘制信息传播网络图，并对传播网络的结构、路径及其影响因素进行分析，最后，发现传播网络的形态与用户的影响力、节点的合理布局及外部干扰因素有关。

关键词　微博客　信息传播　社会网络分析

分类号　G206　TP393

微博客是一种被定义为书写文字不超过140个字符，记录用户当前活动、意见和状态，通过关注、粉丝、评论、转发等功能实现信息传播和共享的博客变体。微博客较其他互联工具最大的优势在于能够更为快速地传播信息，其快速传播信息的原因除了发送方式的多样化以外，一个非常重要的因素是它具有转发功能，并且转发贴能通过“粉丝”的“粉丝”迅速传播扩散。当某一原创贴发布后，它通常被原创作者的粉丝发现并可能转发，转发贴又可能被粉丝的粉丝发现并可能转发，以此层层传播扩散，形成了一个典型的级联传播网络。为了考察微博客的传播规律，本文以新浪微博客为研究平台，跟踪了事件在研究周期内被传播的路径，绘制了节点和路径形成的传播网络结构图，最后对影响信息传播网络结构及路径形成的因素进行了分析。

1　数据选择与处理

数据选择与处理通过以下步骤完成：

* 本文系2008年国家社会科学基金项目“网络社区信息运动模式研究”（项目编号：08BTQ029）研究成果之一。

1.1 微博客平台的选择

由于新浪微博客是目前我国用户数最多，也是最有影响的微博客，因此，本研究确定在新浪微博中选择研究样本。

1.2 传播事件的选择

为了使研究结论更具有一般性，本研究避免选择被过分热捧或有商业炒作嫌疑的事件，而是随机选择一条农民工生存状态的话题（以下称原创贴），原创作者为薛晓棠，地址为：http：//t. sina. com. cn/1554014995/k4CeUoLKt，话题以文字配发图片的形式出现，图片显示一个满脸皱纹、衣着破旧的农民工半蹲着吃干粮的图片，文字对农民工艰苦的生活状态表示了同情。

1.3 确定转发原创贴的用户集，并采集每个用户信息

转发用户及其信息按以下步骤采集：①以原创贴为起点，采集原创贴页面的评论信息，包括：评论用户名，评论用户名的 ID，评论中引用的用户名，评论内容，评论时间；②搜索评论中所引用的用户的 ID；如“@老沉：盟国//@USMBA：有机会要去感受一下这种友谊”这条评论中取老沉和 USMBA 的 ID；③获取上述两类用户（一类是评论用户，一类是评论用户引用的用户）的关注和粉丝（由于系统限制，粉丝至多只能取到前 1 000 位）。④搜索上述用户的微博客（截止到原创内容发表的时间），观察是否转发了原创内容，如果转发，记录该用户，该用户即为本研究要找的转发用户，再取其关注和粉丝，循环上面过程，直到无新的转发者。

在搜索转发原创贴用户的同时，采集该用户 7 项信息：用户 ID、关注数、粉丝数、发表的微博文数、是否认证用户、被转发数、传播级别。其中，被转发数是指用户发布或转发该原创贴后，被其他用户转发的次数，传播级别是本研究中为了考察信息传播的路径而设定的参数，在某条传播路径上，根据用户转发时间的先后划分为不同的传播等级，转发原创节点的节点 A 设为 1 级，转发节点 A 的节点 B 设为 2 级，以此类推。

原创贴发布时间 2010 年 3 月 8 日 23 点 55 分，采集时间从 3 月 25 日 13 点整到 3 月 29 日 13 点整，即数据存在时间域为 3 月 8 日 23 点 55 分 - 3 月 29 日 13 时。共获取转发数为 2 530 次，其中 2 340 个用户转发了 1 次，61 个用户转发了 2 次，11 个用户转发了 3 次，转发 4 次和 9 次的各 1 个用户，转发 5 次和 6 次的各 2 个用户，也就是说，2 351 个用户参与转发，这些用户即为转发原创贴的用户集，用户集中每个用户的 7 项信息被采集入库，以备分析所用。

2 微博客信息传播网络结构与路径

首先定义传播图 G = {V, E, W} 来描述微博客信息传播网络，其中，V 是微博客的节点集合，将微博客用户及其微博客页面统一看作微博客网络的节点，本研究的节点即传播该农民工贴的所有微博客用户及其微博站点。节点 a_1 转发了原创节点 1 的原创贴，则生成一条由 1 指向 a_1 的有向链，a_1 称为一级传播节点，当 a_2 又转发了 a_1 所转发的贴，则又生成一条由 a_1 指向 a_2 的有向链，a_2 称为二级传播节点，以前类推，于是形成了从 $1\rightarrow a_1\rightarrow a_2\rightarrow\cdots$ 的传播链，节点与节点的转发路径数称为距离，例如，节点 a_2 到原创节点 1 的距离为 2，节点到原创点的距离称为链长，当一条传播链传播到 5 级传播节点后，再没有被转发，则称该传播链链长为 5；E 为由原创节点出发，所有转发原创贴的用户及微博站点形成的所有路径集合；在所有转发节点中，由于不同节点的粉丝数、关注数、性别、地域、是否是认证用户均不相同，用户的某些特征可能会造成节点影响力不同，将节点的影响力集合记为 W。也就是说，整个传播网络由节点、路径以及节点的影响力所构成。此外，为了叙述方便，根据一个节点被转发次数的多寡，该节点被称为强势节点、次强势节点及弱势节点。

利用社会网络工具 Netdraw 绘制信息传播网络图[1]。Netdraw 的文本文件程序为（部分）：

```
* node data
uid          level   forward   focus   fans   blogs   vip
216523996    3       16        531     2796   5367    1
227842593    7       2         108     64     240     0
1001560384   6       1         266     117    175     0
…            …       …         …       …      …       …
* tie data
From         to           distance
1554014995   1074250082   1
1554014995   1088047653   1
…            …            …
1074250082   1226270601   2
1074250082   1448166187   2
…            …            …
216523996    1169245034   3
216523996    1363780062   3
…            …            …
```

文本文件中的节点“* node data”不仅给出了节点 ID，同时给出了各节点的属性信息，即包括传播级别（level）、被转发次数（forward）、关注数

(focus)、粉丝数(fans)、发表的微博文数(blogs)以及是否是认证用户(vip)。这样做的优点在于,当使用 Netdraw,绘制原创贴传播的网络图时,鼠标放在任何一个节点上,点击右键,即可看到每个节点的属性信息,便于进行数据观察和分析,更重要的是,可以任意选择节点的某一项属性显示传播网络,使图形更易于观察,如选择节点粉丝数这一属性显示网络时,粉丝数量大的用户的节点尺寸大于粉丝数量小的用户的节点尺寸,一般根据研究的目的及观察的需要,确定选择何种属性显示网络。

文本文件中的"* tie data"中,第一列与第二列都是节点的 ID,根据第二列节点到原点(原点 ID 为 1554014995)的链长(distance)分为若干组,如第一组的链长为 1,说明第一组中第二列节点都是到原点链长为 1 的节点,即在原创贴发布后最早转发的一级传播节点;第二组的链长为 2,以此类推。

调用 Netdraw,绘制原创贴传播的网络图(见图 1),为了图形的清晰,图 1 未显示每个节点的 ID,但在重要的节点标记了数字(如原创节点命名为节点 1)。此外,在软件中设置了节点属性,根据节点被其他用户转发的次数多寡显示节点的大小(如节点 3 被转发了 77 次,图 1 中显示节点 3 的面积最大),以便从图中可以清楚地观测到被转发次数多的用户节点,即所谓强势节点;图 1 中省略了对原创节点仅进行了一级传播和二级传播便无进一步传播的节点,由于原创节点附近有大量这类节点,省略这部分节点,可以更为清楚地观测信息传播的主要路径,但对于大于或等于三级传播的路径,图中保留从 1 级到最终的路径,以保证信息传播路径的完整显示。

图 1 也清楚地显示了微博客信息传播是一个从原创贴为中心层层扩散的级联传播结构,在数据采集周期内,从原创节点出发,最长传播路径链长为 12,即最后一次传播为 12 级传播,以下将对传播过程中的几种典型类型进行分析:

- 第 1 种类型:偶发型。仅发生一级传播,没有其他用户再次传播上一次的传播。它包括两种情况:第一种是对原创节点的一级传播,即 A 转发了原创节点后再没有其他用户转发 A 的转发,从而中断了传播链。在 2 351 个参与转发的用户中,有 1 770 个用户一次传播原创贴后再未被其他用户转发,说明微博客的偶发性传播所占的比例是非常大的;第二种情况是在二级(包括二级)以上传播链中,B 又转发了上一个用户的转发,但没有其他用户再次传播 B(如节点 8),图 1 中可以看到大量这类只有一个分支的情况,这种偶发型的用户通常是在网络中没有影响力的用户,因为没有影响力,信息到此难以继续被传播。
- 第 2 种类型:偶遇机会型。在传播过程中,偶遇强势节点,扩大了传

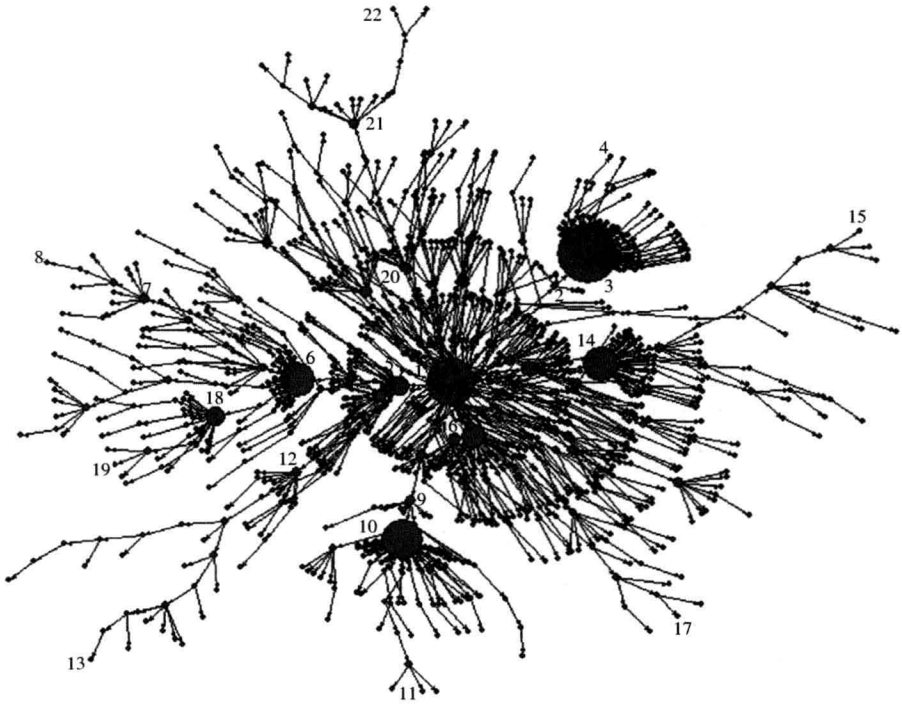

图 1　微博客信息传播结构与路径

播的面积及传播的链长。如原创节点 1 经过节点 2、3 到节点 4 的分支，如果根据新浪微博的用户 ID 排序，该分支按以下用户逐级传播：1554014995（节点 1）－1404971935－1662345577－1212040547（节点 2）－1253531973（节点 3）－1699941995－1698108537－1681717110（节点 4）。该分支在传播过程中，到了第 4 级时，遇到节点 3，该用户有很强的传染力，引发了大量转发，扩大了转发的宽度，将传播延长到 7 级，如果未遇到节点 3，传播可能在 5 级时就中断。同样分支 1－9－10－11 也属此类型，该分支最长距离为 9；分支 1－14－15 情况类似，但由于节点 14 在后面又遇到较为强势的节点，导致传播链长达 11。

● 第 3 种类型：强势节点相互呼应型。从节点 5 发出以下分支：分支 1－5－6－7－8，分支 1－5－6－18－19，分支 1－5－12－13。该子网主要由上述三个分支网络组成，是整个传播网络中传播面积最大且传播链最长（链长为 12）的一个子网。该传播网络强势的原因在于：①在数据采集时间域，节点 5（ID 为 1641428154）的粉丝数高达 24 655 个（可通过 http：//t. sina. com. cn/

1641428154 访问），并且不同于第二种类型的是，在第二种类型中，虽然节点3（ID 为 1253531973）和节点 10（ID 为 1680313495）的粉丝分别高达 108 627 和 42 734，但是由于该两个节点分别处于第 4 级和第 5 级传播，传播级别高于节点 5 的传播级别（节点 5 的传播级别为 1 级），从转发发生的时间看，节点 5 比节点 3 和节点 10 分别早 6 小时 41 分和 10 小时 34 分（节点 5 是 3 月 25 日零点 31 分，而节点 3 和节点 10 的转发时间分别是 3 月 25 日 7 点 12 分和 3 月 25 日 11 点 05 分）。在常规情况下，节点转发时间相差 6 – 10 小时并不会明显地影响传播效率，但对传播量较为密集的时间段（如突发事件往往在短时间快速传播），则会有较大的影响，本文原创贴传播虽然时间周期界定在 3 月 8 日 23 点 55 分 – 3 月 29 日 13 点，但大量的转发均密集地发生在 3 月 25 日这天（原因将在第 4 节分析），因此，在 25 日这一天时间周期里传播提早 6 – 10 小时，就占据了传播上的优势，尽管节点 5 的粉丝量低于节点 3 和节点 10，但由于在传播级别及传播时间上占据优势，从而使其传播具有更大的扩散面积及更长的传播链。②强势节点相互承接配合。继节点 5 后，又有 6，18，12 等强势节点在不同的时刻逐步扩散。从图 1 中可以看到，还有其他次强势节点也在网络中不断承接扩散，形成一个强势、次强势节点逐层承接、均匀地分布较大的传播网络，其中最大链长达到 12。分支 1 – 16 – 17（见图 1 第 4 象限）的情况，属第 2 型与第 3 型中间状态，由于节点 16 属 2 级传播节点，边上又有两个次强势节点配合，使早期传播的面积扩散较快，但整个子网络后面缺乏强势节点承接，分布不够均匀，因此传播网络从面积上弱于第 3 型。

- 第 4 种类型：以节点 1 – 20 – 21 – 22 为主支及临近 4 个支干形成的子网，该子网没有明显的强势传播节点，但是在图 1 中可以看到传播早期有一些次强势节点较为均匀地分布着，中期和后期又有几个次强势节点承接，导致整个网络的面积较大，且传播链长也达到 11。

从上述四种类型可以看到，当传播中强势节点在传播的早期传播，并且网络中强势节点均匀分布时，传播的面积及链长较为理想。这点在网络传播及营销方面有重要意义，合理设置、增加或改变强势节点及其节点合理布局，会改善网络传播结构、速度、面积及链长。

3　网络结构及路径形成的影响因素

传播网络的形成与多种因素有关，除了偶发因素外，发现以下因素对网络的结构及路径的形成产生较大的影响：

3.1　高影响力用户数量影响网络规模及结构

在事件传播网络中，转发是信息传播的关键，当网络中有较多的高被转

发的用户时，才易形成较大规模的网络，那么用户被转发与哪些因素有关就成了问题的关键？为此，从转发用户集中提取各用户的ID、所处的传播级别、被转发的数量，转发时间，关注数、粉丝数、发表的微博数量、是否是认证用户，再将用户被转发数分别与其他字段进行相关性分析，结果发现，用户被转发的数量与该用户的粉丝数存在高度相关，相关系数达0.704 0，与该用户是否是认证用户存在低度相关，而与所发表的微博客条数等没有相关性，如表1、表2所示：

表1　样本用户的主要属性（部分）

用户ID	传播级	被转发数	转发时间	关注数	粉丝数	微博数	是否认证
1554014995	1	63	201003082355	195	363	329	0
1670201604	2	1	201003251006	169	150	256	0
1677638975	3	1	201003251012	369	535	1096	0
1357428174	4	1	201003251018	753	219	35	0
1357426835	5	1	201003251033	631	136	39	0
1357426104	6	3	201003251034	590	111	29	0
1357425703	7	2	201003251036	736	109	32	0

注：用户ID－新浪微博客给予的用户标识；传播级－从原创节点开始发散的一条路径中，各节点按传播的先后命名为传播级，原点本应是零传播级，但为了相关系数计算的方便，将原点设置为1级，其他节点相应均增加一级；是否认证－记0为非认证用户，1为认证用户。

表2　用户被转发数与其他字段的相关系数

相关系数	传播级	转发时间	关注数	粉丝数	博客数	是否认证
被转发数	－0.1334	－0.2750	0.1944	0.7040	0.1563	0.3318

上述研究说明，用户被转发的数量只与用户的粉丝量高度相关，即拥有粉丝数多的用户所发布或转发的贴子才易于被他人发现并继续转发，拥有较多粉丝的用户被称为有影响力的用户，当网络中拥有较多的有影响力用户时，才能吸引更多的关注，从而引起信息的持续传播，形成更大的网络规模，这些有影响力的用户则成了网络的信息中枢。从图1中也可以看到重要的传播链中都有高影响力的用户。

3.2　传播时间、传播级别及高影响力用户的合理分布决定传播的路径

从信息传播的周期看，一般来说，信息传播分为早期、中期和晚期三个阶段，信息发布的早期较易受到关注或引发兴趣，即使是有影响力的用户，