

新浪微博的大规模信息传播规律研究*

易成岐¹, 鲍媛媛^{2,3}, 薛一波^{2,3+}, 姜京池¹

1. 哈尔滨理工大学 计算机科学与技术学院 哈尔滨 150080
2. 清华大学 信息技术研究院 北京 100084
3. 清华大学 信息科学技术国家实验室(筹) 北京 100084

Research on Mechanism of Large-Scale Information Dissemination Based on Sina Weibo*

YI Chengqi¹, BAO Yuanyuan^{2,3}, XUE Yibo^{2,3+}, JIANG Jingchi¹

1. School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China
 2. Research Institute of Information Technology, Tsinghua University, Beijing 100084, China
 3. Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China
- + Corresponding author: E-mail: yiboxue@tsinghua.edu.cn

YI Chengqi, BAO Yuanyuan, XUE Yibo, et al. Research on mechanism of large-scale information dissemination based on Sina Weibo. Journal of Frontiers of Computer Science and Technology, 2013, 7(6) :551-561.

Abstract: Web2.0, as a representative social media, is bringing a profound impact and change on the mode of information dissemination, people's new way of socializing and popular lifestyle. Social network has been a hot research area for academic and industry. This paper relies on the large-scale data of the famous Chinese social network websites (Sina Weibo), addressing the problems of information dissemination on social network. This paper reveals the mechanism of information dissemination on social network, and proposes a few models of information dissemination to try to analyze the rule of information dissemination. These research results can be widely used in the prediction of information dissemination, the perception of dissemination situation, the personalized information recommendation, precision advertising and so on. It also can be used to find out the illegal users and information to ensure the security of network and information.

Key words: social network; information dissemination; dissemination mechanism

* The National Science and Technology Pillar Program of China under Grant No. 2012BAH42B04 (国家科技支撑计划).

Received 2012-11, Accepted 2013-01.

CNKI网络优先出版 2013-03-05, <http://www.cnki.net/kcms/detail/11.5602.TP.20130305.1702.001.html>.

摘要:以 Web2.0 为代表的社会媒体正在深刻影响和改变信息的传播方式、人们的社交方式、大众的生活理念, 社会网络也成为学术界和工业界的研究热点。以中国著名社交网站“新浪微博”的大规模数据为依托, 研究了社会网络的信息传播问题, 揭示了社会网络的信息传播机制, 提出了几种信息传播模型, 尝试剖析社会网络的信息传播规律。这些研究成果既可以应用于信息传播预测、传播态势感知、个性化信息推荐、广告精准投放等多个领域, 也可以用于发现不良用户和信息, 保障网络和信息安全。

关键词: 社会网络; 信息传播; 传播规律

文献标志码: A **中图分类号:** TP393

1 引言

当前, 随着社交网站的迅速发展, 以人与人关系为基础的社会网络正在日益受到广大网民的喜爱。Facebook 创始人 Zuckerberg 在 2012 年 10 月 4 日宣布, 该社交网站活跃用户突破 10 亿大关。这一数字是指平均每月至少登入一次的用户, 相当于全球每七个人当中就有一人拥有 Facebook 账户, 并经常登录。分析机构 SemioCast 发布的数据显示^[1], 截至 2012 年 7 月 1 日, Twitter 的用户数已经突破 5 亿, 达到 5.17 亿, 成为仅次于 Facebook 的全球第二大社交网站。到目前为止, Facebook 和 Twitter 是世界上最成功的社会交友网络, 也是互联网上最不可思议的神话。

在中国, 社会网络作为新媒体时代的一种新兴事物, 也已日益深入到人们的生活中, 深刻影响和改变着人们的人际沟通和信息传递的方式。新浪微博作为一个新兴的中国社交网站, 其日益攀升的用户规模和迅猛发展的影响力在中国掀起了一场“微革命”。2012 年 11 月 17 日, 新浪公布了第三季度的财务报告, 新浪 CEO 曹国伟称, 新浪微博的注册账户已突破 4 亿。

在社会网络这项革命性的创新背后, 也隐藏着一些制约其发展的因素和问题, 例如社会网络信息传播的不确定性、盈利模式、广告营销策略、信息安全问题等, 其模糊不清的发展前景也让人担忧。基于目前现状, 本文针对独具特色的新浪微博进行研究, 分析新浪微博的传播特性, 建立其信息传播模型, 探究其信息传播规律。

通过对社会网络信息传播规律的研究, 不仅可以对信息传播预测提供帮助, 也有助于对消息传播态势进行宏观感知; 另外, 对于个性化信息推荐、广

告的精准投放以及网络和信息安全等方面都大有裨益。

本文组织结构如下: 第 2 章主要介绍信息传播领域的相关工作, 第 3 章对新浪微博的信息传播机制进行分析, 第 4 章通过传播机制的分析, 提出 7 种传播模型, 第 5 章总结全文, 并对未来研究工作进行展望。

2 相关研究

社会网络中的信息传播过程与复杂网络中的传染病模型传播过程有着诸多相似的地方, 因此对社会网络信息传播过程的研究大多都是采用不同传染机制的传染病模型。目前比较经典、研究最为彻底的传播模型有 SIS(susceptible infected susceptible)模型^[2]和 SIR(susceptible infected recovered)模型^[3], 两者的优点是将医学领域中的传染病模型引入到社会网络信息传播中, 但由于信息传播影响因素更多, 并不能完全对其进行刻画。

Watts 和 Strogatz 在提出 SW(small world)网络模型的开创性工作^[4]中指出 SW 效应会加快传染病的传播过程。随后, Pastor-Satorras 和 Vespignani^[5]利用平均场理论, 研究了小世界网络中的传播行为, 突破性地将小世界网络上的传播行为等价于某种逾渗问题。

也有不同的学者对无标度网络中的传播行为进行了研究。其中 Pastor-Satorras 和 Vespignani^[5]研究了无标度网络的传播临界值问题, 得到了无标度网络在网络规模趋于无穷大时, 缺乏传播阈值的结论, 具有突破性意义, 引发了无标度网络传播行为的研究热潮。文献[6]在此基础上, 研究了有限规模无标度网络的传播临界值, 得到了临界值的解析结果。

Boguna 和 Pastor-Satorras^[7]研究了关联网络的传播临界值特性,利用条件概率理论,得知关联网络的传播率的临界值由度与条件概率值形成的矩阵特征值决定。

此外,文献[8]提出了一种用于预测社交网络中消息传播的即时状态的实用算法,并且在 Twitter 数据集上验证了方法的有效性。文献[9-10]分别通过基于图形的分析方法和线性阈值方法研究了 Twitter 中 URL 的传播情况。文献[11-12]基于数据挖掘中的“生存分析”理论研究了 Twitter 中主题的扩散情况和标签内容的传播情况。文献[13-15]则研究了社会网络的传播模型、影响力及规则。目前大多数信息传播研究都比较微观,并没有从宏观的传播态势角度分析其传播模型。

本文与上述相关研究不同,更侧重于以新浪微博中较大传播量的微博为研究对象,主要研究社会网络结构设计对信息传播的影响;根据真实传播数据研究社会网络的信息传播问题,探究社会网络信息传播规律;并且在大规模数据分析基础上,提出了几种社会网络的信息传播模型。

3 新浪微博的传播机制分析

3.1 社会网络结构设计分析

社会网络是以用户节点为基础的真实人际关系建立起来的虚拟网络,用户节点不仅源源不断地生产、使用着信息,而且也是社交网络中信息传播的重要枢纽。信息传播影响的强弱及传播范围的大小,都与用户节点之间的连通性是分不开的。好友可以及时、迅速地收到用户传播的信息,因此一个用户节点有多少好友,是考察一个用户节点信息影响程度的重要因素。在社会网络中有越多的好友,对信息传播的广度、深度、效果等就有越大的影响力。

如果将社会网络进一步细分,则可以分为很多种,例如手机社会网络、即时通信社会网络、邮件社会网络、微博类社会网络等,每种社会网络中的信息传播速度和影响力也不尽相同。

下面将从社会网络的连通性、交互性、隐私性和转发性等四个层次,分析不同类型社会网络的结构

设计差异。

(1)连通性

连通性是点集拓扑学中的基本概念,在本文略有不同,对其定义如下:

定义1(连通性) 连通性是指用户与用户之间建立好友关系的认证限制。

邮件通信社会网络与手机社会网络类似,无论是语音模式还是短信模式,都不需要与对方成为好友,即可进行连通,满足单向同意模式;在微博类社会网络中,用户A与用户B之间建立好友关系,单方只需执行“关注”命令即可,不需要等待另一方同意或认证,也满足单向同意模式,而即时通信网络与上述三者均不同,双方建立好友关系前,必须得到双方同意后,才能建立好友关系,属于双向同意模式,大大限制了其信息传播的连通性。

(2)交互性

定义2(交互性) 交互性是指用户与用户之间的信息交互限制。

在微博类社会网络中,用户A与用户B之间进行信息交互,不需要对方同意,类似“留言”功能;而即时通信网络与邮件通信网络同样满足单向同意模式,但手机通信网络与上述三者则不同,对于其最主要的语音通话功能而言,发送方必须征得接收方同意才能实现语音通话功能,因此手机通信网络在交互性方面,限制了信息的传播速度。

(3)隐私性

定义3(隐私性) 隐私性是指用户与用户之间传递的信息是否可以被其他用户查看。

在微博类社会网络中,用户的个人信息、交互信息、转发信息等全部公开,可以直接被他人查看;而在手机通信网络和邮件通信网络中,用户的所有信息都是私密的;在即时通信网络中,个人信息一般是开放的,但交互信息是私密的。从隐私性可以看出,微博类社会网络开放程度最高。

(4)转发性

定义4(转发性) 转发性是指用户发布的信息是否容易被转发。

在微博类社会网络和邮件通信网络中,两者都

内嵌转发功能,非常便于信息的传播;而手机通信网络和即时通信网络则不具备转发功能,大大阻碍了信息的传播。

通过对比分析发现,微博类社会网络(Facebook、Twitter、新浪微博等)之所以近些年得到非常迅猛的发展,其信息传播能呈爆发式增长,这与其社会网络的结构设计密切相关。具体对比如表1所示。

Table 1 The comparison of social network structure design

表1 社会网络结构设计对比

性能	手机 社会网络	即时通信 社会网络	邮件 社会网络	微博类 社会网络
连通性	单向同意	双向同意	单向同意	单向同意
交互性	双向同意	单向同意	单向同意	单向同意
隐私性	私密	私密	私密	公开
转发性	非内嵌转 发功能	非内嵌转 发功能	内嵌转 发功能	内嵌转 发功能

综上所述,微博类社会网络具有连通性(单向同意)、交互性(单向同意)、隐私性(公开)和转发性(内嵌转发功能)等非常优越的特点,因此具有建立关系方便、交互程度深、开放程度高、信息传播快等优点,对信息的快速传播具有得天独厚的优势,也是其迅猛发展的原因所在。

3.2 新浪微博信息传播机制

区别于传统社会网络(手机通信网络、即时通信网络、邮件通信网络)中以信息内容为主体的传播方式,微博类社会网络人际结构具有扁平化的特点,信息依赖于人与人之间的好友关系进行传播,人与人的结构在信息传播过程中起到非常关键的作用。

新浪微博中的用户通常表现出较强的个体意识,通过用户与用户之间的关系影响用户与信息的关系。另外,新浪微博提供的转发机制使人们在获悉信息时,能够对信息的来源一目了然,因此可以更关注来源的可靠性,增加了信息转发的便利性和趣味性。

新浪微博中的每个用户通常都会拥有一定数量的“粉丝”;“粉丝”不通过目标用户的验证即可建立好友关系。用户发布的信息会以广播的形式,通过

类似滚动黑板报的方式将信息通知到每个“粉丝”,“粉丝”收到信息时决定是否转发。新浪微博可以通过转发功能将喜欢的信息转发到自己的微博,在转发过程中原始微博的内容不会被篡改,且转发记录会以“//@用户名”的方式保留转发路径,此处“用户名”是指转发微博的父节点,同时会以“@用户名”的方式保留原始微博节点。

新浪微博信息转发示意图如图1所示。其中,“北美新浪”代表原始微博信息节点;“X曼妙游离X”代表信息转发节点;“李开复”也是信息转发节点。由“//@X曼妙游离X”可知,“李开复”节点转发了“X曼妙游离X”节点的信息,其信息转发链为:“北美新浪”→“X曼妙游离X”→“李开复”。

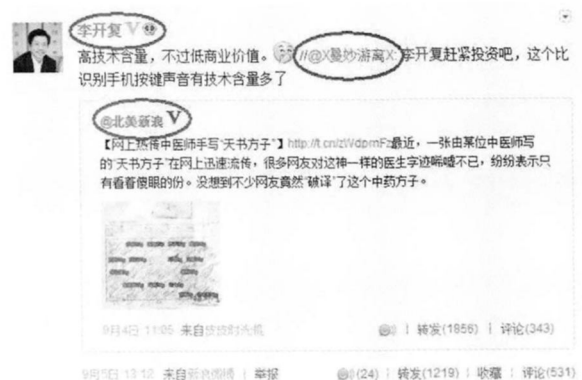


Fig.1 Repost information based on Sina Weibo

图1 新浪微博信息转发示意图

信息M在新浪微博上的传播过程如图2所示。

由此可知,新浪微博在信息传播机制方面设计简洁明了,从易用性角度看,可以加快信息的传播速度。

4 新浪微博的传播模型分析

4.1 大规模数据集获取

实验数据通过新浪微博提供的开放平台(<http://open.weibo.com/>)中的信息转发 API(application programming interface)接口“statuses/repost_timeline”进行获取,为了平衡选取数据的非主观性,以及大规模数据分析的实验定位,在数据集采集过程中,对每条信息的转发节点的个数设定了阈值范围 $\alpha > 5\,000$,

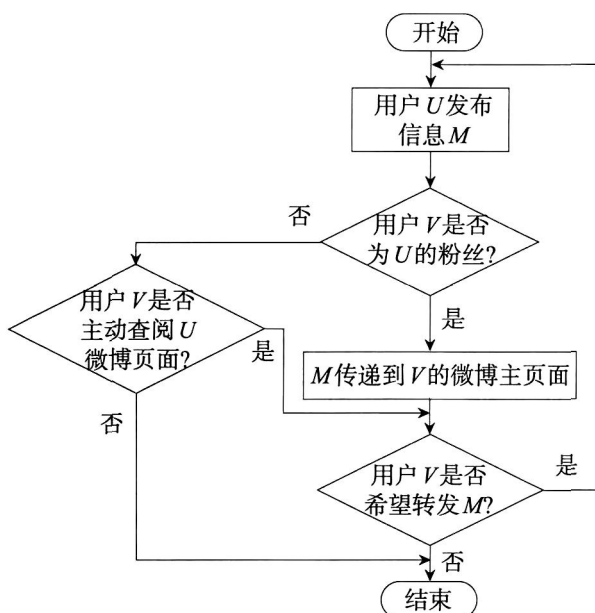


Fig.2 Information dissemination process on Sina Weibo

图2 新浪微博信息传播过程

即只获取转发量在5 000条以上的微博信息。在这一数据获取前提下,本文随机抽取新浪微博中从2012年4月到2012年9月期间“热门微博”板块中500条热门微博信息及其相关的转发微博的详细信息,其中选取的微博的转发量均在5 055到370 269之间。

本文研究了此类微博的转发时间、转发层次关系以及微博传播的中间过程及最终状态,根据转发微博与原微博的转发层次、转发时间等方面的关系,形成微博信息传播树。对于原微博而言,其周围的一层节点代表对于原微博的一次转发,即周围节点与源节点的位置关系代表转发微博与原微博的层次关系。

一棵较复杂的信息传播树示意图如图3所示。其中,给定信息传播图 $G=\langle V, E \rangle$, V 代表用户节点, E 代表信息转发边。图3展示的是一条微博信息 M 在 V_0 和 V_6 之间的传递情况。 E_1 代表 V_1 转发了 V_0 的微博, E_2 代表 V_1 自转发了微博 M ,即 V_1 转发了两次微博,第一次是从 V_0 转发,第二次是对自己已经转发过的微博再次转发。 V_3 节点比较特殊,因为由图3可知, V_3 从 V_1 V_2 两个节点接听到了微博信息,并且

转发了两次,分别传递给 V_4 和 V_5 ,最后 V_6 从 V_4 和 V_5 节点接收到两次微博信息 M 。

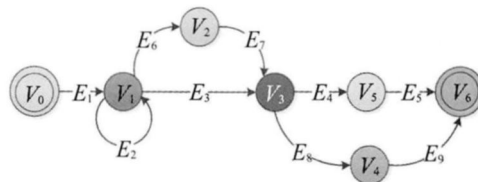


Fig.3 Information dissemination tree

图3 信息传播树示意图

由于本文侧重于大数据传播可视化基础上的微博传播模型分析,为了更好地实现微博消息传播的可视化效果,选取Gephi这一开源社交网络分析工具,应用Yifan Hu算法按照上述微博消息传播树构建的思路,对500棵传播树进行了可视化,得到了传播树的最终状态。经过统计分析,发现大多数传播树具有类似的传播模型,可以总结为:波纹式模型、蒲公英式模型、菌落式模型、烟花式模型、蜂巢式模型、双子星式模型、随机引爆式模型。

在500棵传播树中,7种模型的数量分别为:波纹式模型126棵(25.2%),蒲公英式模型138棵(27.6%),菌落式模型60棵(12.0%),烟花式模型58棵(11.6%),蜂巢式模型28棵(5.6%),双子星式模型24棵(4.8%),随机引爆式模型30棵(6.0%),以及其他36棵(7.2%),如图4所示。

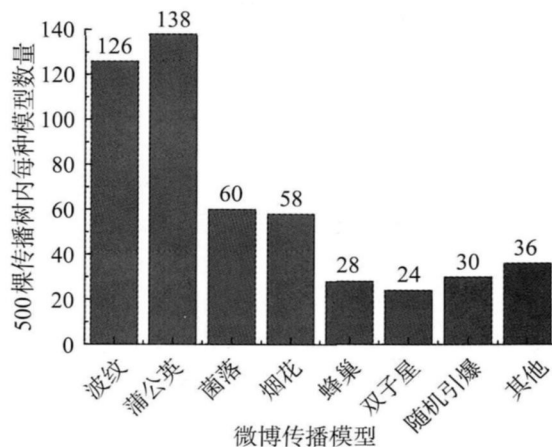


Fig.4 The quantitative comparison of various models

图4 各种传播模型数量比较

下面对7种主要传播模型进行解析。由于其他模型相比较波纹模型而言,传播更为复杂,不确定性更多,本文更着重对波纹模型进行模拟,其他模型的模拟会在后续工作中进一步研究。

4.2 传播模型

4.2.1 波纹式模型

图5展示的是波纹式模型,该案例的基本信息如表2所示。

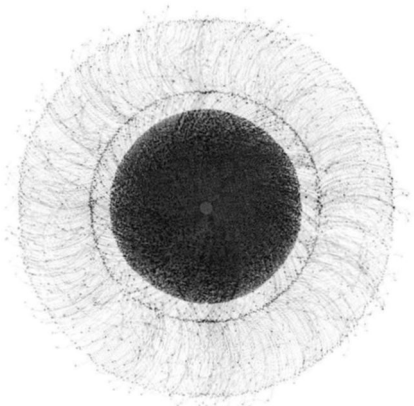


Fig.5 Ripple model

图5 波纹式模型

Table 2 A case of ripple model

表2 波纹式模型案例

名称	描述
用户名	音悦台
用户ID	1698229264
微博ID	3433159230421868
微博内容	“原谅是容易的。再次信任 就没那么容易了。”
转发次数	15 278
微博地址	http://e.weibo.com/1698229264/ydUCD1LKk

波纹式传播模型是最典型的传播模型,在所有传播模型中占有非常高的比例,因为该模型充分体现了社会网络信息传播过程中一对多的广播特点。

当用户发布一条微博信息后,该用户的粉丝会同时收到该条信息,并且一部分粉丝会先后对此微博进行转发,这部分转发仅限于第一层转发。但该模型的微博拥有较少的二次或者多次转发,因此该微博的传播主要集中在一个以原创微博为中心的圈内,类似水波纹从中心点向外扩散。

为了更加真实地模拟波纹式模型,本文只考虑传播过程中的扩散和衰减因素,而不考虑折射、反射等因素。

(1)扩散:当用户发布信息后,即产生入水点,水波上的每一点并不是以入水点为中心向外扩散的,而是水波上的任何一点都是以自己为圆心向四周扩散的。之所以会形成一个环状的水波,是因为水波内部因为扩散对称而相互抵消。

(2)衰减:传播介质(人物网络关系)是有阻尼的,否则信息就会永不停止地扩散。

假设 N_0 点波纹的振幅除了受 N_0 点自身振幅的影响外,同时受来自它周围 n 个点(粉丝)的影响 $S=\{N_1, N_2, \dots, N_n\}$,而且这 n 个点对 N_0 点的影响机会是均等的,那么可以假设这个公式为:

$$N_0' = \alpha \sum_{i=1}^n N_i + \beta N_0 \quad (N_i \in S) \quad (1)$$

其中, α β 为波纹待定系数; N_0' 为目标节点下一时刻的振幅; N_i 为周围节点当前时刻的振幅。

假设水的阻尼系数为 γ ,在这种理想条件下,波纹的总势能将保持不变,那么可以得到下面这个公式:

$$N_0' + N_1' + \dots + N_n' = N_0 + N_1 + \dots + N_n + \gamma \quad (2)$$

将每个点代入式(1),可以得到:

$$(n\alpha + \beta)N_0 + (n\alpha + \beta)N_1 + \dots + (n\alpha + \beta)N_n = N_0 + N_1 + \dots + N_n + \gamma \quad (3)$$

因为不能忽略阻尼,否则信息将无限制传播,所以 $\gamma \neq 0$,即 $(n\alpha + \beta) \neq 1$ 。可以通过调整波纹待定系数 α β 和阻尼系数 γ 模拟波纹式传播。

波纹式传播模型常见于名人、专业机构以及某些领域内的专家微博,而且很多热门的微博,往往由很多个小的波纹式模型组成。

4.2.2 蒲公英式模型

图6展示的是蒲公英式模型,该案例的基本信息如表3所示。

蒲公英式传播模型也是一个典型的传播模型,该模型由波纹式模型发展而来。在波纹式传播模型形成后,会有少数、稀疏的用户在二次转发的基础上进行再次转发,其中包含一些破碎的小型爆发点。

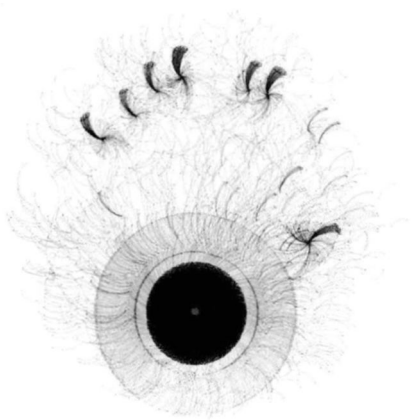


Fig.6 Dandelion model
图6 蒲公英式模型

Table 3 A case of dandelion model
表3 蒲公英式模型案例

名称	描述
用户名	每日微小说
用户 ID	2147536551
微博 ID	3436460755909917
微博内容	“在一号线上看到条导盲犬,主人是个女孩,一路都睡的很香甜,到站时主人轻轻动了动狗狗马上站起来,一边打哈欠一边等主人,下了地铁狗狗一路领着她在人群里穿梭,避开别人的推搡和拥挤,看了有点感动,真的很有爱! [心]”
转发次数	18 632
微博地址	http://weibo.com/2147536551/yfivFoNrf

蒲公英式模型与波纹式模型相比,用户的关注程度加深,转发层次更深,但仍然不具备大规模爆发的能力。

4.2.3 菌落式模型

图7展示的是菌落式模型,该案例的基本信息如表4所示。

菌落式传播模型特点非常明显,从原创微博信息开始传播,并且形成有限个超大规模的波纹式模型。

从图7可以看出,在该传播模型中,较高影响力节点的中心性非常强,除了4个超大波纹式模型外,其他转发树枝非常稀少。这种模型很大程度是由几个高影响力节点传播而形成,一些企业会参照该模型通过旗下多个子品牌设计营销策略。

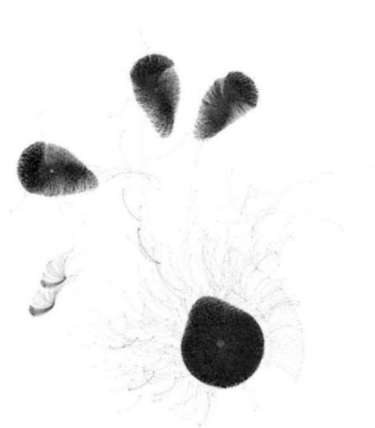


Fig.7 Colony model
图7 菌落式模型

Table 4 A case of colony model
表4 菌落式模型案例

名称	描述
用户名	路过心上的话儿
用户 ID	2370022990
微博 ID	3436475146635229
微博内容	“两个人在一起,更多的不是改变对方,而是接受对方,这就是包容。如果光想着改变对方,那不是生活,那是战争。”
转发次数	16 993
微博地址	http://weibo.com/2370022990/yfiSSrQ7P

4.2.4 烟花式模型

图8展示的是烟花式模型,该案例的基本信息如表5所示。

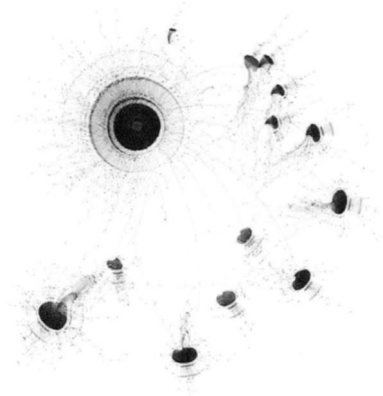


Fig.8 Fireworks model
图8 烟花式模型

Table 5 A case of fireworks model

表5 烟花式模型案例

名称	描述
用户名	精彩电影
用户ID	1843443790
微博ID	3436810799792502
微博内容	“《吸血鬼日记》的执行制片人朱莉.佩勒克透露: elena在本季剧终会在 stefan 和 damon 中作出选择。。“希望本剧的粉丝以及埃琳娜的粉丝都会理解并且尊重她的选择,这将是一个喜忧参半的选择。”第三季剧终将在5月10日播出,你期待她跟谁在一起??? http://t.cn/zOOIafF ”
转发次数	12 843
微博地址	http://weibo.com/1843443790/yfrCfF5tA

烟花式传播模型也是十分常见的传播模型,这些微博本身具备非常强大的传播力,同时在二次或者多次转发中,又有很多关键节点协助转发,最后达到上万或者上十万的转发规模。

这种方式传播的微博信息往往是用户关注度非常高的热门话题,能够引起很多用户的共鸣,并且得到了很多影响力较高的关键节点的推荐,这样的微博具有强大的影响力。但是该模型的主从规模分辨能力很强,往往原始节点影响规模最大,其他节点影响规模相对较小。烟花式模型传播能力很强,当形成一定规模后,很难控制。

4.2.5 蜂巢式模型

图9展示的是蜂巢式模型,该案例的基本信息如表6所示。

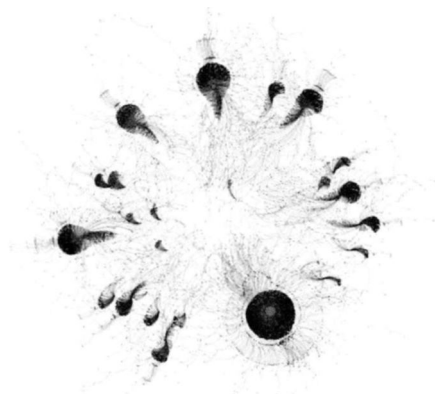


Fig.9 Beehive model

图9 蜂巢式模型

Table 6 A case of beehive model

表6 蜂巢式模型案例

名称	描述
用户名	全球潮宝宝
用户ID	2054793217
微博ID	3436813114947191
微博内容	“OMG! 贵妃醉酒? @全球潮宝宝”
转发次数	14 503
微博地址	http://weibo.com/2054793217/yfrFZkKZp

蜂巢式传播模型与烟花式传播模型从拓扑模型看,似乎有些类似,但两者最大的区别是,前者主从规模分辨能力强,原始节点在信息传播过程中的贡献最大,因此对原始节点要求也很高;而后者主从规模分辨能力弱,转发后的爆发点与原始爆发点影响规模相近,转发节点的贡献值也高,因此具备更强大的影响力和持续的爆发能力。

由于原创微博节点外形成了若干个局部的波纹式模型,其对微博传播的贡献和规模接近或者超过了原创本身。蜂巢式模型中,随转发的深入,会出现几个关键节点,对转发进行不断推升。

4.2.6 双子星式模型

图10展示的是双子星式模型,该案例的基本信息如表7所示。

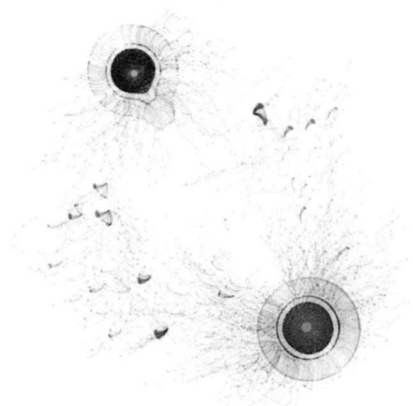


Fig.10 Binary star model

图10 双子星式模型

双子星式模型很特殊,形成该模型的前提是在整个传播过程中存在两个高影响力,并且影响力相近的关键节点。同时,这两个节点会进行激烈交互,

Table 7 A case of binary star model

表7 双子星式模型案例

名称	描述
用户名	袁弘
用户ID	1223179064
微博ID	3436888147053016
微博内容	“虽然第一个答案让我有点紧迫感,但最后一个回答还是蛮让人欣慰的。@胡歌 @林更新”
转发次数	12 212
微博地址	http://weibo.com/1223179064/yftD0tAOk

两者的粉丝中有绝大部分会伴随各自节点参与转发,也会有一定数量的粉丝活跃于两个节点之间,保证两点的连通性。

4.2.7 随机引爆式模型

图 11 展示的是随机引爆式模型,该案例的基本信息如表 8 所示。

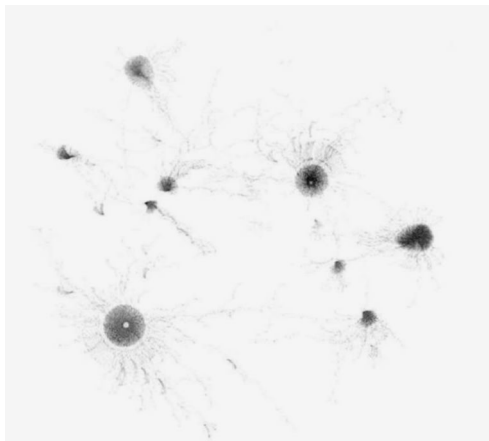


Fig.11 Random detonate model

图11 随机引爆式模型

随机引爆式模型的主要特征是不确定性因素加强,传播拓扑集合了波纹式模型、蒲公英式模型、轻量级烟花式模型等关键特征。微博信息传播引爆点较杂乱,随机性增强。

4.3 对传播模型的总结分析

对上述 7 种传播模型进行分析,总结如下:

(1)对于名人微博、企业微博和草根明星微博而言,由于其拥有庞大的粉丝群体,容易形成波纹式传播、蒲公英式传播、烟花式传播、蜂巢式传播。

(2)对于普通用户微博而言,在传播过程中,如

Table 8 A case of random detonate model

表8 随机引爆式模型案例

名称	描述
用户名	雷军
用户ID	1749127163
微博ID	3461549869799578
微博内容	“周老板,刚看了您的神机配置,有三点请教: 1.神机怎么只有双核 1.2G,而不是双核 1.5G? 2.相机为啥只有 500 万像素? 哪个厂的? 3.你天天鼓吹的神屏是哪里产的? 网上传言是国产山寨屏,请周老板辟一下谣。@周鸿祎”
转发次数	5 348
微博地址	http://weibo.com/1749127163/ypPbYF7jI

果没有高影响力用户参与,很难形成传播规模大的信息传播树。

(3)目前一些企业多参考菌落式模型进行营销,通过企业旗下多个子账号参与信息传播。

(4)烟花式模型和蜂巢式模型用户共鸣度最高,该类微博具有极强的影响力,而且可控性较差。

(5)信息形成双子星式模型的条件较苛刻,但该类微博的两个节点间互动性较强。

(6)随机引爆式模型不确定性最强,传播引爆点杂乱,规律性差。

5 结束语

本文以中国著名社交网站“新浪微博”为研究对象,主要研究了社会网络的信息传播问题。首先分析了不同社会网络的结构设计对信息传播的影响;接着对新浪微博的信息传播机制进行了研究和分析;最后通过对大规模数据的分析,抽象出了 7 种信息传播模型,剖析了不同传播模型的形成原因与主要差异,为后续更加深入的研究奠定了基础。

下一步工作会研究没有抽象的传播模型,以及形成不同传播模型的关键特征,并且将不同传播模型与传播预测方法进行关联、分析。在此基础上,不仅可以实现信息的传播预测和态势感知,还可以进行个性化的信息推荐、广告的精准投放、好友推荐等。另外,从网络和信息安全角度,可以及时发现不良用户和不良信息。

致谢 特别感谢周正隆、孙浩硕同学为本文部分实验所做的工作。

References:

- [1] AOL Inc. Analyst: Twitter passed 500M users in June 2012, 140M of them in US; Jakarta 'Biggest Tweeting' city[EB/OL]. (2012-07-30)[2012-10]. <http://techcrunch.com/2012/07/30/analyst-twitter-passed-500m-users-in-june-2012-140m-of-them-in-us-jakarta-biggest-tweeting-city/>.
- [2] Li Cong, van de Bovenkamp R, van Mieghem P. Susceptible-infected-susceptible model: a comparison of N-intertwined and heterogeneous mean-field approximations[J]. *Physical Review E*, 2012, 86: 026116.
- [3] Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks[J]. *Physical Review Letters*, 2001, 86(14): 3200-3203.
- [4] Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks[J]. *Nature*, 1988, 393: 440-442.
- [5] Pastor-Satorras R, Vespignani A. Epidemic dynamics and endemic states in complex networks[J]. *Physical Review E*, 2001, 63: 066117.
- [6] Zhanette D H. Critical behavior of propagation on small-world networks[J]. *Physical Review E*, 2001, 64: 050901.
- [7] Boguna M, Pastor-Satorras R. Epidemic spreading in correlated complex networks[J]. *Physical Review E*, 2002, 66: 047104.
- [8] Guille A, Hacid H. A predictive model for the temporal dynamics of information diffusion in online social networks[C]// *Proceedings of the 21st International Conference Companion on World Wide Web (WWW '12 Companion)*, Lyon, France, Apr 16-20, 2012. New York, NY, USA: ACM, 2012: 1145-1152.
- [9] Bakshy E, Hofman J M, Mason W A, et al. Everyone's an influencer: quantifying influence on Twitter[C]// *Proceedings of the 4th ACM International Conference on Web Search and Data Mining (WSDM '11)*, Hong Kong, China, Feb 9-12, 2011. New York, NY, USA: ACM, 2011: 65-74.
- [10] Galuba W, Aberer K, Chakraborty D, et al. Outtweeting the twitterers- predicting information cascades in microblogs[C]// *Proceedings of the 3rd International Conference on Online Social Networks (WOSN '10)*, Boston, USA, June 22-25, 2010. Berkeley, USA: USENIX Association, 2010: 3-11.
- [11] Yang J, Counts S. Predicting the speed, scale and range of information diffusion in Twitter[C]// *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM '10)*, Washington, DC, USA, May 23-26, 2010. [S.l.]: Association for the Advancement of Artificial Intelligence, 2010: 355-358.
- [12] Yang J, Leskovec J. Modeling information diffusion in implicit networks[C]// *Proceedings of the 2010 IEEE International Conference on Data Mining (ICDM '10)*, Sydney, Australia, Dec 14-17, 2010. Washington, DC, USA: IEEE Computer Society, 2010: 599-608.
- [13] Bakshy E, Rosenn I. The role of social networks in information diffusion[C]// *Proceedings of the 21st International Conference on World Wide Web (WWW '12)*, Lyon, France, Apr 16-20, 2012. New York, NY, USA: ACM, 2012: 519-528.
- [14] Starbird K, Palen L. Will the revolution be retweeted: information diffusion and the 2011 Egyptian uprising[C]// *Proceedings of the 2012 ACM Conference on Computer Supported Cooperative Work (CSCW '12)*, Seattle, USA, Feb 11-15, 2012. New York, NY, USA: ACM, 2012: 7-16.
- [15] Matsubara Y, Sakurai Y, Prakash B A. Rise and fall patterns of information diffusion: model and implications[C]// *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '12)*, Beijing, China, Aug 12-16, 2012. New York, NY, USA: ACM, 2012: 6-14.



YI Chengqi was born in 1988. He is a Ph.D. candidate at School of Computer Science and Technology, Harbin University of Science and Technology. His research interests include social network, data mining and cloud computing, etc.

易成岐(1988—),男,黑龙江嫩江人,哈尔滨理工大学计算机科学与技术学院博士研究生,主要研究领域为社交网络、数据挖掘、云计算等。



BAO Yuanyuan was born in 1984. She is a postdoctor at Research Institute of Information Technology, Tsinghua University. Her research interests include social network and behavior dynamics, etc.

鲍媛媛(1984—),女,山东东营人,清华大学信息技术研究院博士后,主要研究领域为社交网络、行为动力学等。



XUE Yibo was born in 1967. He received his M.S. degree in computer science and engineering at Harbin Institute of Technology in 1992, and Ph.D. degree in computer architecture at Institute of Computing Technology, Chinese Academy of Sciences in 1995. Now he is a professor at Research Institute of Information Technology, Tsinghua University, the senior member of CCF and the member of IEEE/ACM. His research interests include computer network, information security, parallel processing and distributed system. He has published more than 100 papers in journals and conferences.

薛一波(1967—),男,1992年于哈尔滨工业大学计算机科学与工程专业获得硕士学位,1995年于中国科学院计算技术研究所计算机结构专业获得博士学位,现为清华大学信息技术研究院研究员,CCF高级会员,IEEE/ACM会员,主要研究领域为计算机网络、信息安全、并行处理、分布式系统。发表论文100余篇。



JIANG Jingchi was born in 1990. He is a master candidate at School of Computer Science and Technology, Harbin University of Science and Technology. His research interests include Web crawler and social network, etc.

姜京池(1990—),男,黑龙江佳木斯人,哈尔滨理工大学计算机科学与技术学院硕士研究生,主要研究领域为网络爬虫、社交网络等。