兴趣社交网站中意见领袖识别的研究*

——以“点点网”为例

王娟 曹树金 姜灵敏 唐宝珍

[摘 要] 以国内最大的轻博客网站——点点网为研究对象,根据点点网用户间兴趣关系进行社区划分,通过统计兴趣社区的拓扑特性,发现其都具有小世界和无标度特性,说明有少数用户在信息传播中起着至关重要的作用。之后利用节点中心性测量方法进行意见领袖的识别,分析 4 个已有的中心性指标——连接度、中介度、接近度和核数在点点网意见领袖识别中的不足,构建新的意见领袖识别指标,并通过实验证实该指标具有更高的准确性。

[关键词] 意见领袖 兴趣图谱 在线社会网络 轻博客 社会网络分析
[分类号] G350
DOI:10.7536/j.issn.0252-3116.2013.14.017

1 引 言

早在 1940 年,美国哥伦比亚大学的传播学者拉扎斯菲尔德[1]即在研究“两级流动传播理论”时发现并提出“意见领袖”这一概念。他认为意见领袖是指在人际传播网络中经常为他人提供信息,同时对他人施加影响的活跃分子。随着 Web 2.0 技术的快速发展,在线论坛、博客、微博等在线社会网络逐渐成为网络中的“民意集散地”,意见领袖较好的意见引导和示范作用在这里得到了进一步的放大和扩展。科技新闻聚合网站 Techmeme 的编辑马亨德拉·帕素雷在知名科技博客 TechCrunch 上称网络正在从简单的社交共享进入“内容相关性时代”,配合社交图谱的兴趣图谱变得越来越重要[2]。G. Tan 2013 年 1 月的调查报告也表明,2007 成立的轻博客网站 Tumblr 已经超越 Facebook,成为美国年轻人访问最多的社交网站[3],其注册用户达到 1.084 亿,日均 PV(Pageview,访问量)超过 2.6 亿,UV(Unique Vistor,独立访客)超过 1 000 万。Tumblr 定位于兴趣爱好的社交平台,人们通过共同爱好形成各种兴趣小组,小组内用户针对某个或某几个主题展开交流,作为领域专家的意见领袖在引导小组兴趣主题的发展上起着主导作用。

目前,国内学者对意见领袖的研究主要集中在对营销领域[4]、网络舆情[5-7]中意见领袖的识别,针对兴趣社交网站中意见领袖识别的研究尚未见到。因此,研究兴趣社交网站中意见领袖识别方法以有效识别兴趣社区中的意见领袖不仅有助于满足网络中其他用户某种程度的信息需求,维持其他用户的网络粘性,而且可将真正的意见领袖凸显出来,使意见领袖得到更广泛的关注,知识和经验的分享变得更加容易,由此营造出一个健康向上的网络环境。

本文之所以以国内最大的轻博客平台——点点网作为研究对象,是因为相比其他类似网站,被称为“Tumblr 中文版”的点点网是最纯粹的轻博客,其网络结构特征具有很强的代表性。研究内容包括两部分:基于兴趣关系的点点网拓扑结构的研究和兴趣社区中意见领袖识别的研究。

2 相关研究综述

2.1 OSN(online social netwok,在线社会网络)拓扑特性分析

社会网络是由关系连接而成的社会实体网络。网络中的“节点”是个人,“连接”则是按某种方式定义的

*本文系 2012 年度国家社会科学基金重大项目“基于特定领域的网络资源知识组织与导航机制研究”(项目编号:12&ZD222)和 2012 年度教育部人文社会科学研究青年基金项目“面向轻博客热点话题情感倾向性分析的研究”(项目编号:12YJC870023)研究成果之一。
[作者简介] 王娟,广东外语外贸大学思科信息学院讲师,中山大学资讯管理学院博士研究生,E-mail:misisipiwj@126.com;曹树金,中山大学资讯管理学院教授;姜灵敏,广东外语外贸大学思科信息学院教授;唐宝珍,广东外语外贸大学思科信息学院讲师。
收稿日期:2013-06-24 本文起止页码:97-104,22 本文责任编辑:易飞

人与人之间的关系。网络中信息传播的速度受社会网络结构的影响,关系紧密的网络的信息传播会比较容易,而关系疏远的网络通常会出现信息不畅的问题。网络中节点也因为所处位置不同,彼此间掌握和控制资源的能力存在较大差异。作为现实社会网络在互联网中的映射与扩展,在线社会网络重建了社会连接与纽带,其网络的结构特征和发展也引起了越来越多学者的关注。

以 1998 年康奈尔大学的 D. J. Watts 和 S. H. Strogatz 建立的小世界模型[8]、1999 年圣母大学的 A. L. Barabási 和 R. Albert 建立的无标度模型[9]为标志,研究人员广泛使用复杂网络理论对社会网络进行分析,理解社会网络性质和功能。A. Java 等[10]研究了微博网络 Twitter 的网络拓扑和地理分布特性,并借助相应的社区结构去自动识别用户发贴的兴趣和动机。A. Mislove[11]针对 4 家流行的、以分享内容为主的交友网站——Flick、YouTube、LiveJournal 和 Orkut 开展了大规模的网络测量,结果验证了 OSN 的幂率特性、小世界效应以及无标度特性。Fu Feng[12]研究了大学在线交友网——校内网,验证了校内网中用户好友关系网络的小世界和无标度特性,并表现为同配模式。C. Wilson[13]依据获取到的 22 个区域的用户信息对 Facebook 的拓扑特性进行了研究,发现 420 万节点度的中值为 144,均值为 179.53,这与 R. I. M. Dunbar[14]提出的“150 法则”不谋而合,即一个人最多能维持 150 个好友关系。

从现有研究文献可以看出,近几年对各种在线社会网络拓扑特性的研究取得了很大的进展。但是,针对兴趣社交网站拓扑结构的研究尚未出现,有必要对此开研究。

2.2　意见领袖的特征与识别方法

E. M. Rogers[15]总结了 4 种有效测量意见领袖的方法:关键人物访谈法、观察法、自我报告法和社会网络分析法。B. Lyons 等[16]采用传统意见领袖的定义,将研究对象分为意见领袖和跟随者两种,并采用 T. L. Childers[17]量表测量出其中的意见领袖。丁汉青[6]采用观察法识别 SNS 网络空间中的意见领袖,第一个层面考察发帖人在小组和网络层面的位置,即中心性;第二个层面考察发帖人在话题层面的发言频率、质量和效果,即活跃性、吸聚力和传染力。王君泽[7]采用观察法将关注用户数量、粉丝数量、是否被验证身份和发布的微博数量作为微博意见领袖识别模型的 4 个维度。

目前,国内很大一部分学者都沿用关键人物访谈法、观察法或者依靠主观判断确定网络中的意见领袖。这些方法虽然便于操作,但往往带有很强的主观性,在筛选意见领袖时,容易受个人主观意志的影响而产生偏差,特别是网络的匿名性和随意性也为这些研究方法的信度和效度打了折扣。而社会网络分析法从关系取向和位置取向进行定量分析,关系取向考察宏观效应,关注节点之间的社会性粘着关系;位置取向考察微观效应,关注节点位置的影响,本文认为综合宏观和微观两个层面可以更加准确地识别兴趣社区中的“意见领袖”。

3　点点网的数据采集

网站的数据采集方法主要有基于 API 和基于网络爬虫的数据采集方法。通过调用网站提供的 API 接口可以实现网站数据的便捷抓取与解析,但也要注意:一是 API 内容开放不全面,例如点点网 API 2011 年 12 月才对外开放,API 的种类也很少,目前不到 30 个;二是 API 服务商对用户的 API 接口调用频率与查询的返回结果的最大数量有限制,点点网就规定查询的返回结果不超过 20 个;三是使用 API 接口需要解决用户认证问题,如果待获取用户条目太多,则会占用大量系统开销来等待用户授权许可。因此,本文采用基于网络爬虫的数据采集技术,在开源软件 Heritrix 的基础上,自行开发数据采集器来获取点点网的数据。

点点网自身统计数据显示,截至 2013 年 3 月,点点网注册用户数已经达到 1 919 万,帖子数达到 3 547 万,数据采集量十分庞大且处于动态变化之中,要获取整个网络的拓扑数据十分困难,因此本文采用滚雪球采样法,依据“兴趣标签”,随机选择两个标签下面的“杰出轻博客”中的某篇轻博文(http://cdpjohnyu.diandian.com/post/2013-03-11/40049888596 和 http://10min.diandian.com/post/2012-12-09/40046688258)作为种子,利用点点网用户之间的兴趣关系进行广度优先搜索。搜索对象为含有“post/”和含有“n/common/comment”的 URL,这些 URL 代表着两类页面:帖子页面和热度页面。

从图 1 可以发现,点点网的每篇轻博文下面都有“热度”,标注喜欢、转载和推荐该文的用户列表。查看源码,发现“热度”是一个内嵌网页,该网页采用 AJAX 技术,页面源码中内容比较少,更多的内容实际上是采用 Javascript 驱动的异步请求/响应机制加载出来的。如果直接用 Heritirx 原有的抓取方法,则抓取不到真正的用户列表。因此,必须对 Heritirx 的 Extracotr

图 1 点点网页面及相关源代码片段呈现

类进行扩展，扩展后的新类 DiandianExtractor 的工作流程采用 Selenium-WebDriver + PhantomJS + Jsoup 技术架构，如图 2 所示。通过 Selenium WebDriver API 驱动浏览器内核 PhantomJS，模拟浏览器获取 AJAX 内容，得到和页面呈现一致的页面内容，再通过 Jsoup 解析页面内容，至此，AJAX 页面采集问题得到真正解决。

图 2 扩展类 DiandianExtrator 的工作流程

采集器共抓取了近 600 万页面，总容量接近 60G。经过实时抽取后共有 1 898 356 条记录存到 MySQL 数据库。其中，数据表结构包括 id、username（用户名）、inname（链入用户名）、type（链入用户是哪种类型用户：喜欢、转载还是推荐）、link（该记录从哪个链接得来）。经过去重，即从数据表中删除 username 和 inname 都相同的记录，最终得到用户数为 128 786，用户间的兴趣关系数为 825 057。

4 点点网拓扑结构测量

4.1 兴趣关系网络模型

在点点网中，一个用户“喜欢”或是“转载”、“推荐”另一个用户的轻博文是因为他对此文感兴趣，而不是因为他跟另一个用户有某种社会联系。正是因为这种基于兴趣的互动关系，更多素不相识的点点网用户通过关注“兴趣标签”建立起社交关系，网络逐渐由松散的个体聚合为或大或小的有边界群体，进而形成一个庞大的兴趣社区。

基于获取的实验数据集，本文将点点网抽象为一个基于兴趣关系的有向网络 G(V,E)：V 代表节点集合，每个点点网的注册用户即为一个节点；E 代表边集合，若用户 i 发布轻博文，用户 j 对该文感兴趣，则存在一条从节点 j 指向节点 i 的有向边，如图 3 所示：

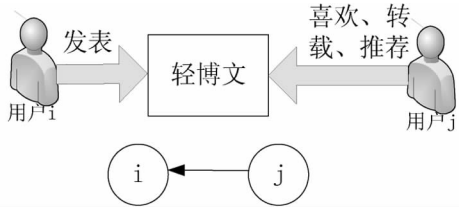

图 3 网络抽象示意

整个网络的节点数为 128 786，连接数为825 057，其中双向连接数为 4 328，占连接总数的 0.5%。网络密度为 4.97×10^{-5}，表明网络很稀疏，由许多孤立的节点组成。因此，为了更好地进行统计分析，本文进行了噪音过滤，在 Pajek 中打开“点点. NET”文件，通过

Reduction(简化)操作,删除所有度低于 2 的节点,再通过 Strong Components(强连接部分),找到网络的极大连通子图,该连通子图共包括 4 673 个节点和84 975 条边。利用这个网络图,就可以找到不同类型的子图,如兴趣社区;也可以找到子图的中心性节点,即最有影响力的用户——意见领袖。

4.2　点点网的兴趣社区

社区[18]的概念最初是在社会学领域提出的,通常是指一组具有某种共同属性或起某种相似作用的节点集合。从社会网络、信息网络再到生物网络,网络的社区结构随处可见。在点点网中也存在着许多隐性子社区,这些隐性的子社区往往是由一些兴趣相同或相似的用户组成的小组,一个小组内部用户间的互动频率明显高于不同小组用户间的互动频率。

由于社区发现算法不是本文研究重点,本文采用了 Pajek 自带的社区发现算法——Louvain Method[19]来进行点点网的兴趣社区划分,该算法被证明是目前在保证一定精确度的前提下运行速度最快的算法之一。根据 Louvain Method,点点网可划分为 5 个社区(见表 1),其中最小社区包含的节点数为 3,最大社区包含的节点数为 4 010,模块度 Q = 0.515 672,具有显著的社区结构。

表 1　社区分析结果

社区	规模	内连接度之和	外连接度之和	兴趣标签
1	4 010	141 206	7 908	视觉、摄影、色彩、风景
2	158	2 096	1 447	美女写真
3	484	10 478	6 541	游戏、动漫、手绘、ACG
4	18	160	107	仁爱基金会、公益、慈善
5	3	4	3	UI 设计

为了进一步验证 Louvain Method 划分的有效性,本文对每个社区的内连接度和外连接度之和做了统计,可以看出:每个社区的内连接度之和都远大于外连接度之和,符合 F. Radicchi 等人提出的社区弱定义“一个社区结构被定义为一组节点的集合,它的内部节点度之和大于指向外部的节点度之和[20]”。

4.3　点点网的拓扑特性分析

4.3.1　网络测量指标　多数在线社会网络和现实网络一样,都具有复杂网络的一些共性:小世界和无标度特性。兴趣社交网站是否同样具有这些性质?本文利用 Pajek 对点点网的网络拓扑特性进行统计分析,测量指标如下:

- 平均路径长度。网络中节点 i 到节点 j 的最短路径上的边数定义为该节点对的最短距离,所有节点对之间的最短距离的平均值称为网络的平均路径长度。
- 聚集系数。用于描述一个节点邻居之间相互连接的紧密程度。节点 i 的聚集系数定义为该节点的所有邻居节点之间实际存在的连接边数与可能的最大连接边数的比值。整个网络所有节点聚集系数的平均值称为网络的聚集系数。
- 度分布。网络中节点的度分布情况可以用累计分布函数(Cumulative Distribution Function, CDF)或互补累计分布函数(Complementary CDF, CCDF)来表示。前者描述网络中度不小于 x 的节点数占网络节点总数的比值;后者描述网络中度大于 x 的节点数占网络节点总数的比值。近几年大量研究表明,许多 OSN 的度分布都表现为幂率(Power – law)分布,即:

$$P[X \leq x] = F(x) \approx cx^{-\gamma}(1 \leqslant \gamma \leqslant 3) \text{ 或 } P[X > x] = 1 - F(x) \approx cx^{-\alpha}(0 \leqslant \leqslant 2) \quad (1)$$

- 连接度相关性。节点之间的连接通常会表现出某种倾向性。如果高度节点倾向于连接其他高度节点,该网络是正相关的;反之,该网络是负相关的。本文采用 Newman 相关系数 r[13]来度量网络节点的连接倾向性,若 r > 0,则称网络节点连接是正相关的;若 r < 0,则是负相关的。

4.3.2　结果分析　对于一些对节点度的大小有要求的测量指标来说,社区 4 和社区 5 中的节点数太少,因此,本文对这 2 个社区的拓扑特性将不做讨论。表 2 是前 3 个社区的拓扑特性统计结果,可以看出,3 个社区中节点的平均路径长度均小于或约等于 4,意味着在一个社区中平均只需要 4 个人就可以为任意两个用户建立联系。3 个社区的聚集系数 C 在 0.118 和0.231 之间,远大于同等规模的随机网络聚集系数 C_{rand}。较小的平均路径长度与较大的聚集系数说明 3 个社区都具有典型的小世界特征,即信息在网络中流通顺畅,网络中心节点的影响力能够快捷地辐射至整个社区。

表 2　点点网的拓扑特性统计

社区	平均路径长度	平均度	聚集系数 C	C_{rand}	连接度相关性 r
1	4.02	35.21	0.118	0.0087	–0.1446
2	3.70	13.27	0.231	0.084	–0.0463
3	3.29	21.65	0.227	0.045	–0.1642

图 4 给出的是 3 个社区中节点入度、出度分布的 CCDF 图,可以看出 3 个社区中普遍存在着出入度大的节点占比例较小,而出入度小的节点占比例较大的现

象。通过对点集进行鲁棒最小二乘拟合，拟合曲线表达式为：

$$P[X>x]\approx cx^{-\alpha},\ \text{wherein}\ \begin{cases}\alpha_1=1.185, k=k_{in}\\ \alpha_1=1.529, k=k_{out}\\ \alpha_2=1.213, k=k_{in}\\ \alpha_2=1.308, k=k_{out}\\ \alpha_3=1.019, k=k_{in}\\ \alpha_3=1.188, k=k_{out}\end{cases} \quad (2)$$

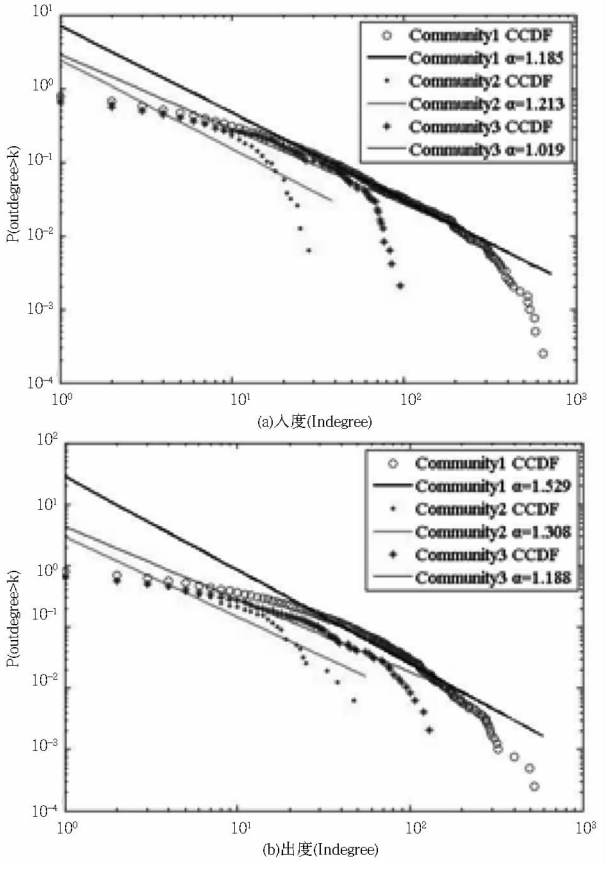

图 4 网络的出入度双对数坐标分布

公式 2 中，3 个社区的出入度分布的幂率指数α均满足大多数社交网络的度分布函数的幂率指数范围 1 <α<2[21]，这种无标度特性揭示出网络存在中心节点、网络资源占有不平衡的现象，即社区中存在少数用户，他们或是具有较高的“吸引力”，通过专业、高质量的发文，引起许多用户的兴趣；或是具有较高的“活跃度”，通过频繁与其他用户的互动（喜欢、转载和推荐其他用户的轻博文），推动兴趣主题的发展。

进一步研究发现，点点网 3 个社区的 Newman 相关系数 r 均小于 0，呈现连接度负相关性，与现实社会网络不同。在现实生活中，人与人之间的关系表现出正相关，知名人士倾向于跟同类型（同等地位、同等知识背景等）的人交往。而在点点网的 3 个社区中，度大的节点多与度小的节点相连，即在信息发布和互动过程中，影响总是倾向于从中心用户流到普通用户那里去，从而产生了信息传播中的“意见领袖”。

5 点点网意见领袖的识别

5.1 备选指标分析

R. Vander Merwe 等[22]综合应用关键人物访谈法、自我报告法和社会网络分析法对意见领袖测量进行分析，发现在社会网络中具有较高“中心度”的个体具备了意见领袖的特征。本文在已有文献研究的基础上，分别选取连接度[23]、中介度[24]、接近度[25]和核数[26]4 个备选指标对点点网节点的中心程度进行考察。为方便比较，仅使用这 4 个指标对社区 1 中的所有节点进行统计，并按照节点中心性值的降序进行排列（前 10 位），如表 3 所示：

表 3 节点的中心性测量结果

排序	连接度				中介度		接近度	
	节点	点入度值	点出度值	连接度值	节点	中介度值	节点	接近度值
1	596	746	10	756	1 113	0.055 1	596	0.516 9
2	1 113	603	138	741	259	0.052 8	1 113	0.516 4
3	461	678	54	732	348	0.035 4	461	0.515 8
4	4 603	634	1	635	539	0.035 3	12	0.512 3
5	1 389	620	10	630	212	0.034 1	926	0.511 0
6	926	560	65	625	208	0.030 7	4 603	0.510 5
7	4 202	613	4	617	789	0.030 2	348	0.506 6
8	12	9	607	615	461	0.029 2	1 389	0.503 4
9	348	440	163	603	632	0.029 0	62	0.502 9
10	2 026	545	6	551	94	0.028 1	4 202	0.499 0

- 连接度。是一个最简单、最具有直观性的指标，考察与某节点直接相连的其他节点的个数。连接度越大，该节点与其他节点的互动越频繁，其中心性越高。

按照连接度从大到小排列，位列前 10 位的节点分别是 596、1 113、461、4 603、1 389、926、4 202、12、348 和 2 026，说明这 10 个节点与很多其他节点相连。其中，节点 596 与其他节点联系最为紧密，应该处于网络中心。但观察发现，10 个节点存在点入度与点出度严重不均衡的现象。节点 596、4 603、1 389、4 202 和 2 026 的点入度值远大于点出度值，说明这 5 个用户发文比较专业，具有较高的“吸引力”，但很少与其他用户互动，不利于相关兴趣主题的发展。节点 12 的点出度值远大于点入度，说明该用户互动比较多，但他主要是通过不断地“喜欢、转载、推荐”其他用户的轻博文提高自身的“活跃度”，自己的轻博文却很少受到关注，不能形成较大的意见影响力。而节点 1 113、461、926 和

348 则不同,他们的点入度值和点出度值都较高,说明这 4 个用户既吸引了很多用户的注意,又积极推动主题信息的流动,属于高“吸引力”、高“活跃度”的中心用户,应该获得意见的高位。

• 中介度。考察一个节点落在其他任意两个节点最短路径中的程度,中介度越大,表明该节点越处于信息枢纽的位置,对信息的控制能力越强。

从表 3 可以看出,中介度排在前 10 位的节点是 1 113、259、348、539、212、208、789、461、632 和 94,他们在网络中起到“桥”作用,控制着他人的交往。其中,节点 1 113 和 259 的中介度远大于第三位,说明这两个节点在网络的交流中有着举足轻重的作用,如果少了这两个节点,网络中很多用户就无法实现信息的交流。而节点 4 124 的中介度为 0,说明该用户不能控制任何信息的流动。

本文利用 Pajek 对节点进行连通性测试。选择菜单 Network → Create New Network → Remove → Selected Vertices,分别删除节点 1 113 和节点 4 124,发现删除节点 1 113 对网络连通性的破坏大,原来的 1 个连通子图转为 19 个连通子图;而删除节点 4 124 并没有破坏网络的连通性,仍然保持为 1 个连通子图。根据安世虎[27]提出的“破坏性等价于重要性”原则,节点 1 113 要比节点 4 124 重要,这也再次验证了中介度大的节点要比中介度小的节点重要。但是整个网络的中介中心势为 0.054,比较低,说明网络中绝大部分节点的“桥”作用不明显。

• 接近度。主要是从时间和成本效率来衡量。A. Bavelas 指出,在网络中最中心的节点上产生的消息将以最短的时间传遍整个网络[28]。接近度越高,表明该节点比其他节点能更快地到达网络中的所有节点,这意味着更少的信息中转、更少的时间、更低的成本。

可以发现,接近度测量结果的排名与连接度基本一致。网络中节点的接近度值均在 0.2 – 0.6 之间,连接较紧密,信息可以快速在节点之间进行传播。接近度排名前 10 位的节点分别是 596、1 113、461、12、926、4 603、348、1 389、62 和 4 202,不论是发布信息还是获取信息,较之其他节点,这十个节点不容易受到他人控制,独立性强。整个网络的接近中心势较低,入中心势和出中心势分别为 0.182 和 0.167,说明网络中的大部分节点在发布信息时,会受到少数节点的控制;在获取信息时,更会受到少数节点的控制。

• 核数。网络中的 k – core 是指反复去掉度小于或等于 k 的节点及其连接的边之后所剩余的子网。若一个节点存在于 k – core,而在(k + 1) – core 中被移去,那么该节点的核数为 k[29]。核数描述了网络拓扑的层次特征,核值大的节点聚集在网络的核心,核值小的节点聚集在网络边缘。

通过 k-core 分析,点点网的最大 k 值为 47,包括 178 个节点,连接度、中介度、接近度排在前 10 位的节点均在其中。因此,可以认为这 178 个节点在整个网络中占据了重要位置,并主导了信息的交流。

5.2　构建识别指标

由 4 个备选指标对节点中心程度的测量结果可知:

• 网络中存在连接度很大的节点,但这些节点并不都是点入度与点出度均衡的节点。存在少数高“吸引力”、低“活跃度”节点,如节点 596 的点入度值远大于点出度值;亦存在高“活跃度”、低“吸引力”节点,如节点 12 的点出度值远大于点入度。前者不能积极推动兴趣主题的发展,后者不能形成较大的意见影响力。

• 网络中存在中介度值和连接度值都较小但处于整体网络核心的节点,如节点 1 227,核数为 47,但其中介度值只有 0.000 6,连接度值也只有 84,说明该用户虽然处于网络的核心位置,但其“吸引力”不高,作为“桥”的能力也不强。同时,网络中也存在不处于整体网络中心但中介度值和连接度值都较大的节点,如节点 188,中介度值为 0.014 6,连接度值为 145,都比节点 1 227 大,说明前者的“吸引力”和“桥”的能力都强于后者,但前者的核数为 41,不在网络的核心位置。

因此,本文认为单独使用 4 个指标均不能准确描述节点的中心程度,需要综合连接度、中介度、接近度和核数 4 个指标才能更加有效地识别意见领袖。基于点点网的有向性,本文又将连接度分为内连接度和外连接度两个指标。表 4 是 5 个指标的相关系数矩阵:

表 4　节点中心性指标的相关系数矩阵

指标	内连接度	外连接度	中介度	接近度	核数
内连接度	1.000	.132	.614	.514	.490
外连接度	.132	1.000	.528	.616	.591
中介度	.614	.528	1.000	.474	.413
接近度	.514	.616	.474	1.000	.908
核数	.490	.591	.413	.908	1.000

从表 4 可以看出,接近度与核数两个指标存在着极其显著的关系,内连接度与中介度、接近度指标存在着显著关系,外连接度与中介度、接近度、核数指标存在着显著关系,说明这些指标存在信息重叠。因此,本文采用 SPSS 提供的主成分分析方法,得到 5 个中心性

指标的权重,形成一个新的综合指标 F,表达式为:

$$F = 0.382Z_{内连接度} + 0.410Z_{外连接度} + 0.419Z_{中介度} + 0.512\ Z_{接近度} + 0.498Z_{核数} \quad (3)$$

为了衡量该综合指标在识别意见领袖时的准确度,本文采用"覆盖率(coverage ratio)[30]"进行评价。覆盖率指前 n% 意见领袖团体所覆盖的用户与总用户的比例,用来描述意见领袖团体影响范围。从图 5 可以清晰地看出,曲线斜率在 5% 就开始趋于平缓,说明采用综合指标 F 得到的前 5% 的用户基本上对网络中 80% 的用户产生影响,这个指标能够有效地让兴趣社区中的意见领袖凸现出来。

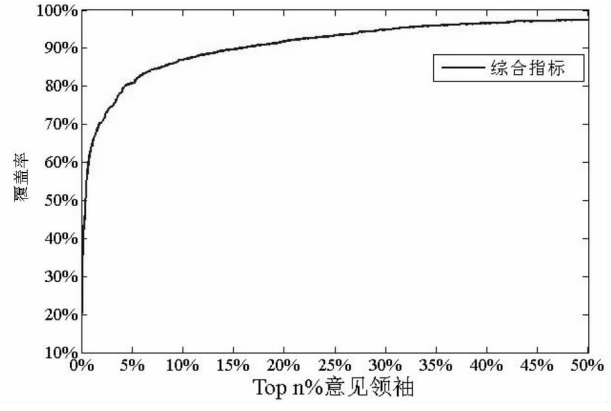

图 5 覆盖率示意

5.3 实证分析

本文依据公式(3)分别计算点点网的 5 个兴趣社区中节点的 F 值,取排名第一的用户作为该兴趣社区的意见领袖,如表 5 所示:

表 5 兴趣社区的意见领袖

社区	兴趣标签	节点	用户名	F 值	内连接度	外连接度	连接度	中介度	接近度	核数
1	风景摄影	1 113	oui0404	18.341	603	138	741	0.055 124	0.516 414	47
2	美女写真	924	328120879	2.136	6	71	77	0.002 780	0.417 441	38
3	动漫	137	momo - ada	7.646	213	110	323	0.017 322	0.463 814	46
4	仁爱慈善	651	lingjiu	2.348	31	48	79	0.007 213	0.401 340	30
5	UI 设计	3 431	vivid - yang	-2	1	1	2	0.000 021	0.296 353	2

查看点点网页面数据可知:用户 oui0404、328120879、momo - ada 和 vivid - yang 发文频率比较高,所发轻博文均是图文并茂、高质量的内容,专业性强且有个性。用户 lingjiu 作为仁爱志愿者一员,也是许多仁爱慈善活动的目击者,有第一手资料并能及时将资料发布出去,让活动信息得到更加迅速和广泛的传播。

据此,与基于 4 个单指标的意见领袖识别方法相比,本文构建的基于点点网的意见领袖识别综合指标具有更高的准确性,识别出的意见领袖均是所在兴趣社区中具有强大影响力的活跃人物,在信息发布和传播过程中起着重要的引导和示范作用。

6 结 论

随着 Web 2.0 技术的快速发展,网络正在从简单的社交共享进入"内容相关性时代"。人们通过共同爱好形成各种兴趣小组,小组内用户互动频率明显高于不同小组用户间的互动频率,而作为领域专家的意见领袖在引导小组兴趣主题的发展上起着主导作用。目前,国内学者对意见领袖的研究主要集中在对营销领域、网络舆情领域中意见领袖的识别,针对兴趣社交网站中意见领袖识别的研究尚未见到报道,有必要予以专门研究。

本文选取国内最大的轻博客平台——点点网作为研究对象,主要是因为轻博客的定位就是一个兴趣爱好的社交平台,而点点网号称国内最纯粹的轻博客,其网络结构特征具有很强的代表性。对点点网意见领袖识别的研究包括两个部分:

其一,对基于兴趣关系的点点网拓扑结构展开测量。根据采集下来的点点网样本数据,构造一个基于"发文←喜欢、转载、推荐"互动的兴趣关系网络。再通过 Pajek 提供的社区划分算法 Louvain Method 对点点网进行兴趣社区划分,统计每个社区的拓扑特性,如平均路径长度、聚集系数、出入度分布及连接度相关性等,发现均存在小世界和无标度特性,说明社区中信息流通顺畅,少数用户在信息传播中起着至关重要的作用,其影响力能够快捷地辐射至整个社区,这为进一步识别网络中的意见领袖奠定了基础。

其二,研究兴趣社区中意见领袖的识别方法。通过对描述节点中心性的 4 个指标——连接度、中介度、接近度和核数进行实证分析,发现:

- 直接使用连接度指标,会筛选出高"吸引力"、低"活跃度"的用户,不利于相关兴趣主题的发展;或是会筛选出高"活跃度"、低"吸引力"的用户,不能形成较大的意见影响力。而直接使用接近度进行识别的效果与连接度基本一致。
- 通过连通性测试,发现使用中介度筛选出的"桥"节点对整个网络的连通性破坏大,但整个网络的中介中心势比较低,中介度在识别意见领袖中的作用不显著。
- 直接使用核数指标,会筛选出"吸引力"不高、"桥"能力也不强的用户。

因此，本文采用社会网络分析法，在点点网数据基础上，通过主成分分析方法构建一个新的综合指标 F，并通过实验证实该指标在意见领袖识别上的准确性，具有重要的实践意义，不仅有助于满足其他用户某种程度的信息需求，维持其他用户的网络粘性；同时也将真正的意见领袖凸显出来，使意见领袖得到更广泛的关注，知识和经验的分享变得更加容易，由此营造出一个健康向上的网络环境。

下一步的研究工作主要包括两个部分：一是鉴于兴趣社交网站中意见领袖识别指标的构建仅基于点点网，其在其他兴趣社交网站上的适用性如何，仍然需要考证；二是需要结合用户行为特征，如用户发文数量和频率、用户被关注数等，进一步完善本文所构建的意见领袖识别指标。

参考文献：

[1] Lazarsfeld F, Berelson B, Gaudet H, et al. The people's choice [M]. New York: Columbia University Press, 1948.

[2] 高兴. 最新历史版本：兴趣社交网络的崛起[EB/OL]. [2013-02-22]. http://www.techcn.com.cn/index.php? edition-view-182991-1.html.

[3] 果子. 影子大亨 Tumblr 的成功之道 [EB/OL]. [2013-02-21]. http://www.36kr.com/p/201458.html? ref=weixin0222m.

[4] 罗晓光，溪璐路. 基于社会网络分析方法的顾客口碑意见领袖研究[J]. 管理评论，2012，24(1)：75-81.

[5] 刘志明，刘鲁. 微博网络舆情中的意见领袖识别及分析[J]. 系统工程，2011，29(6)：8-16.

[6] 丁汉青，王亚萍. SNS 网络空间中“意见领袖”特征之分析———以豆瓣网为例[J]. 新闻与传播研究，2010(3)：83-91.

[7] 王君泽，王雅蕾，禹航，等. 微博客意见领袖识别模型研究[J]. 新闻与传播研究，2011(6)：81-88.

[8] Watts D J, Strogatz S H. Collective dynamics of small world networks[J]. Nature, 1998, 393(6684): 440-442.

[9] Barabási A L, Albert R. Emergence of scaling in random networks [J]. Science, 1999, 286(5439): 509-512.

[10] Java A, Song Xiaodan, Finin T, et al. Why we Twitter: Understanding micro-blogging usage and communities [C]// Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis. New York: ACM Press, 2007: 56-65.

[11] Mislove A, Marcon M, Gummadi K P. Measurement and analysis of online social networks [C]//Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement. New York: ACM Press, 2007: 29-42.

[12] Fu Feng, Liu Lianghuan, Wang Long. Empirical analysis of online social networks in the age of Web 2.0[J]. Physica A, 2008, 387(2): 675-684.

[13] Wilson C, Boe B, Sala A, et al. User interactions in social networks and their implications[C]//Proceedings of the 4th ACM European Conference on Computer Systems. New York: ACM Press, 2009: 205-218.

[14] Dunbar R I M. Coevolution of neocortical size, group size and language in humans[J]. Behavioral and Brain Sciences, 1993, 16(4): 681-735.

[15] Rogers E M. Diffusion of innovations(5th Ed)[M]. New York: Free Press, 2003.

[16] Lyons B, Henderson K. Opinion leadership in a computer-mediated environment[J]. Journal of Consumer Behavior, 2005, 4(5): 319-329.

[17] Childers T L. Measurement of opinion leadership[J]. Journal of Marketing Research, 1986, 23(2): 184-188.

[18] Wasserman S, Faust K. Social network analysis: Methods and applications[M]. Cambridge: Cambridge University Press, 1994.

[19] Blondel V D, Guillaume J L, Lambiottel R, et al. Fast unfolding of communities in large networks [J]. Journal of Statistical Mechanics: Theory and Experiment, 2008(10): 10008-10019.

[20] Radicchi F, Castellano C, Cecconi F, et al. Defining and identifying communities in networks [J]. Proceedings of the National Academy of the Sciences of the United States of America, 2004, 101(9): 2658-2663.

[21] Albert R, Barabasi A L. Statistical mechanics of complex networks [J]. Reviews of Modern Physics, 2002, 74(1): 47-97.

[22] van der Merwe R, van Heerden G. Finding and utilizing opinion leaders: Social networks and the power of relationships[J]. South African Journal of Business Management, 2009, 40(3): 65-75.

[23] Callaway D S, Newman M E J, Strogatz S H, et al. Network robustness and fragility: Percolation on random graphs[J]. Physics Review Letters, 2000, 85(25): 5468-5471.

[24] Freeman L. C. A set of measures of centrality based upon betweenness[J]. Sociometry, 1977, 40(1): 35-41.

[25] 陈静，孙林夫. 复杂网络中节点重要度评估[J]. 西南交通大学学报，2009，44(3)：426-429.

[26] Kitsak M, Gallos L K, Havlin S, et al. Identifying influential spreaders in complex networks [J]. Nature Physics, 2010, 6(11): 888-893.

[27] 安世虎，都艺兵，曲吉林. 节点集重要性测度[J]. 中国管理科学，2006，14(1)：106-111.

[28] Bavelas A. A mathematical model for group structures[J]. Human Organization, 1948, 7(3): 16-30.

[29] Tomasz L. Size and connectivity of the k-core of a random graph [J]. Discrete Mathematics, 1991, 91(1): 61-68.

[30] Lancaster F W, Fayen E G. Information retrieval on-line los angeles[M]. Los Angeles: Melville Publishing Co, 1973.

（下转第 22 页）

参考文献：

[1] 李梅军. 高校图书馆面向社会服务研究[J]. 图书馆工作与研究,2008(5): 87-91.

[2] 张建国,田秋菊. 我国高校图书馆向社会开放的冷思考[J]. 大学图书情报学刊, 2009(1):19-22.

[3] 武继山. 应当避免对高校图书馆向社会开放的误读[J]. 大学图书馆学报,2009(3):16-19.

[4] 马娴. 高校图书馆社会化服务新探[J]. 高校图书馆工作,2011(3):70-72.

[5] 王玉林、曾咏梅,崔然,等. 我国高校图书馆面向社会开放现状调查[J]. 图书与情报, 2011(6):26-32.

[6] 金声. 高校图书馆社会化服务问题辨析[J]. 教育教学论坛, 2012(旬刊):196-199.

[7] 韩宇. 美国若干所著名大学图书馆的读者权利管理[J]. 大学图书馆学报, 2008(2):22-26.

[8] 帕提曼. 高校图书馆向社会开放:被误读的取向[J]. 图书情报工作, 2010,54(19):133-136.

[9] 吉聪. 高校图书馆校外读者权限管理若干问题探讨[J]. 图书馆论坛,2010(4):173-175.

[10] 李美琴. 浅论高校图书馆社会化服务中需要处理的几个关系[J]. 科技情报开发与经济,2012(6):20-22.

[11] 崔红雁. 刍议现阶段高校图书馆对社会开放的原则[J]. 大学图书情报学刊, 2009(4):38-41.

[12] 百度百科. 信息资源共享[EB/OL]. [2013-03-15]. http://baike.baidu.com/view/965696.htm

[13] 张雪梅. 地方高校图书馆对社会开放的探讨[J]. 龙岩学院学报, 2010(2):133-136.

[14] 谢叶. 图书馆经济效益与社会效益初探[J]. 情报探索,2007(4): 15-17.

The Relationship Involved with University Libraries Open to Public

Wu Jin　Liu Sisi

Shenyang Normal University Library, Shenyang 110034

[**Abstract**] With regard to university libraries open to public, the paper analyzes and clarifies the relationship between intramural resources and social service , social readers and intramural readers, paid service and free service, information resource sharing and intellectual property protection, general service and specialized service, social benefit and economic benefit as well, putting forward the effective measures to solve these problems, in order to promote the practice of social service of university libraries.

[**Keywords**] university library　university library open to society　the university　relationship

(上接第 104 页)

Research on the Recognition of Opinion Leaders in the Interest Community: A Case Study of diandian.com

Wang Juan[1,2]　Cao Shujin[2]　Jiang Lingmin[1]　Tang Baozhen[1]

[1]Cisco School of Informatics, Guangdong University of Foreign Studies, Guangzhou Guangdong 510420

[2]School of Information Management, Sun Yat-Sen University, Guangzhou Guangdong 510275

[**Abstract**] This paper takes diandian.com as the study object. It partitions users of diandian.com into five interest communities according to their similar interest relation, and analyzes their topology properties to find that each of them has a small world and scale free characteristics. The result shows that a few users play a crucial role in the process of information dissemination. Then, it recognizes opinion leaders in the interest community with the measurement method of complex network centrality. Then this paper analyzes four centrality indicators which include node degree, betweenness degree, closeness degree and k-core, and finds some problems in the process of recognizing opinion leaders by them. Finally, it proposes a new recognition indicator, and confirms its higher accuracy through an experiment.

[**Keywords**] opinion leader　interest graph　online social network　light blogging　social network analysis