

微博转发网络中意见领袖的识别与分析

熊 涛 何 跃

(四川大学商学院 成都 610064)

【摘要】以微博中的转发关系构建邻接矩阵,通过改进后的 HITS 算法识别微博意见领袖,并构建基于转发关系的意见领袖网,验证算法有效性并分析意见领袖在网络中的作用。研究表明:改进后的 HITS 算法能够有效地识别意见领袖;意见领袖的中心值与其粉丝数高度正相关。通过对意见领袖网的分析发现:意见领袖在网络的关键节点中占有重要地位,意见领袖的作用并没有因为微博中信息源的增多而削弱。

【关键词】微博 意见领袖 转发关系 HITS 算法

【分类号】G202

The Identification and Analysis of Micro – blogging Opinion Leaders in the Network of Retweet Relationship

Xiong Tao He Yue

(Business School , Sichuan University , Chengdu 610064 , China)

【Abstract】 This paper builds the adjacent matrix based on the retweet relationship in micro – blogging to find the opinion leaders through improved HITS algorithm. Then a net of opinion leaders is built based on the relationship of retweet to prove the efficiency of the improved algorithm and to analyze the function of opinion leaders. The research shows that the improved HITS algorithm can find the opinion leaders effectively. The hub of an opinion leader is highly positive correlated with the amount of his fans. According to the analysis of the net of opinion leader , the authors find that the opinion leaders play important roles in the key nodes and their function is not weakened by the increasing of information sources in the micro – blogging.

【Keywords】 Micro – blogging Opinion leader Retweet relationship HITS algorithm

1 引 言

“意见领袖”最早由美国传播学学者拉扎斯菲尔德(Lazarsfeld) 在其著作《人民的选择》中提出^[1],他认为信息是按照“媒体——意见领袖——受众”的模式传播的。后续的研究则认为,意见领袖实际上兼具“信息中间人”和“有影响力的人”两种角色^[2]。最近几年,互联网的发展,特别是微博等社交网络的快速发展改变了人们获取信息与相互沟通的方式。在社交网络的信息传播交流中,意见领袖的作用更加明显。越来越多的研究表明,意见领袖在政治选举^[3]、突发事件传播^[4]、网络口碑效应^[5]等社会现象中具有重要作用。随着微博日益成为网络舆论阵地,研究微博等社交网络中的意见领袖对网络干预^[6]、微博营销^[7]等问题具有极大的意义。

微博平台上充斥着社会名人、机构、网络红人和其他微博大户等账户。他们并不能与意见领袖直接划等号。随着这些账号越来越多,识别微博中的意见领袖就显得更加重要。在微博平台中,用户的一次“转发”行为代表信息的一次传播。用户想表达的信息和用户的影响力可以沿着用户的“转发”网络传播开来。因此,本文提出一种

收稿日期: 2013 – 04 – 22

收修改稿日期: 2013 – 05 – 30

基于转发关系的意见领袖识别算法,同时验证算法的有效性,并进一步分析意见领袖影响力的来源和他们在信息传播过程中的作用。

2 相关研究

目前常用的识别意见领袖的方法包括指标打分法和社会网络结构挖掘法两类。

(1) 指标打分法是以意见领袖的特征为基础。通过对意见领袖的特征分析,建立相应的指标体系并打分。Vergani^[3]选择了社会群体认同、专业知识和社会资本三个指标。丁汉青等^[8]构建的指标包括中心性、活跃性、吸聚力、传染性。刘志明等^[9]根据微博的特点,建立了相对比较复杂的指标体系,包括影响力、活跃度两个一级指标,以及被转发数、被评论数、被提及数、原创数、自回帖数、回复他人帖子数、活跃天数 7 个二级指标。建立指标体系后,可以通过层次分析法^[9,10]确定权重,进行加权平均得到最后结果,也可以直接构造评分模型,通过评价函数得到最后结果^[11]。

(2) 社会网络结构挖掘法一般是根据用户之间的关系构建社会网络,再利用网络结构算法分析用户之间的链接关系,计算用户的重要性排名,排名靠前的就是意见领袖。Weng 等^[12]据此提出基于 PageRank 的 TwitterRank 方法来实现某一特定主题下有影响力的 Twitter 用户发现。肖宇等^[13]在 PageRank 的基础上加入情感权重,提出 LeaderRank 的意见领袖发现算法。Zhang 等^[14]研究了类 PageRank 算法 Z-Score、PageRank 算法和 HITS 算法在 Java 论坛中的专家(Expertise)识别,发现这些算法的效果和人工判断几乎一样好。

指标法更复杂更全面,也更常用。但是指标体系的构建存在一些问题,比如指标选取有较大争议;指标权重在很大程度上由研究者主观确定,导致指标法难以复制重现。比较起来,因为信息传播与网络结构有显著性关系^[15],社会网络结构挖掘在客观性方面具有更大的优势。同时,指标法可以为社会网络结构法的改进提供方向和依据。

但是目前使用网络结构挖掘算法的文献也存在一些不足。比如很多文献都着眼于对 PageRank 的改进^[12-14]。但是在微博平台中,用户多次转发其他用户微博是用户活跃度和信息扩散能力的体现,并不会因

为多次转发导致影响力贡献分散,所以 PageRank 的均分“投票”的思想并不适用^[16]。作为另外一种网络结构挖掘算法——HITS 算法的相关文献则很少,也鲜有对 HITS 算法的改进^[14],而原始的 HITS 算法难以适应微博环境。

其他使用网络结构挖掘算法的文献多集中于对传统的论坛的研究,或者仅仅关注于微博用户通过“关注”建立的社会网络,并未研究用户通过“转发”建立的网络。有研究表明,相对于粉丝数,微博的被转发情况更能准确地衡量用户的影响力^[17]。考虑到意见领袖作为信息中间人的角色,研究转发的意义更大。

因此本文参考指标法中影响力和活跃度两个指标,提出一种改进的 HITS 算法研究微博转发关系网络,实现意见领袖的发现。在此基础上,构建意见领袖网,分析意见领袖在网络中的特征和作用,以验证算法的有效性。

3 HITS 算法改进

HITS(Hypertext Introduced Topic Search)算法由 Kleinberg 提出^[18],最初应用于 Web 网页排序。但由于其本质是一种网络图的结构挖掘,因此 HITS 算法能够应用于很多新的领域,如期刊评价^[19]。

HITS 算法将网页分成中心网页(Hub Pages)和权威网页(Authority Pages)。权威网页,就是普遍认可的针对特定主题的重要网页;而中心网页则是指向很多与特定主题相关的权威网页的网页。权威网页和中心网页之间存在相互加强的依赖关系:一个好的权威网页应该被很多好的中心网页所指向;而一个好的中心网页则应该指向很多好的权威网页。HITS 算法利用它们之间相互加强的关系,通过迭代计算得到网页排序结果。在第 i 次迭代中,原始 HITS 算法计算过程如下^[18]:

I 操作(计算权威值):

$$a(p)_i = \sum_{q: (q,p) \in E} h(q)_{i-1} \quad (1)$$

O 操作(计算中心值):

$$h(p)_i = \sum_{q: (p,q) \in E} a(q)_{i-1} \quad (2)$$

对应到利用微博平台上用户之间的关注、转发等关系,同样可以通过 HITS 算法得到用户排名结果。其中,权威值可以认为是意见领袖作为“有影响力的人”的体现,中心值则是意见领袖作为“信息中间人”的体现。

为了将 HITS 算法引入到微博平台上的意见领袖识别, 还需要解决三个问题: 网络图边的选择、边的取值以及边的权重确定。本文根据这三个问题, 对 HITS 算法进行相应的改进, 以适应在微博环境下的意见领袖发现。不同于网页, 微博平台的特征主要体现在以下几个方面:

(1) 邻接图中边的含义不同。在微博平台上, 邻接图的边可以表示用户间的“关注”关系, 也可以表示用户间的评论或转发微文的关系。而网页之间的关系则是单纯的网页间的链接。

(2) 邻接图中边的取值不同。在微博中, 评论关系和转发关系都可能多次出现, 分别表示多次评论和多次转发。考虑到这可能是用户影响力、活跃度的体现, 邻接矩阵中边的取值不能单纯取为 0 或 1。而在某个网页中多次出现另一个网页链接则没有这么多的含义。

(3) 权重不同。微博用户的粉丝数代表该用户的信息能传达到的人数, 因此粉丝数是一种天然的权重系数, 而网页的权重则较难确定。

3.1 邻接边的选择

大多数研究意见领袖的文献都选择以用户间的“关注”关系作为邻接图的边。这样虽然能够简单明了地确定用户的影响力, 但是有一定的局限性。本文选择“转发”关系作为邻接边的原因如下:

(1) 微博用户可能因为共同的利益、关注点, 或者突发性事件等原因, 通过相互转发微博而形成一个临时群体。经过对微博进行抽样发现, 这样的临时群体的成员之间并不一定存在直接的关注关系。因此关注关系在这里有很大的局限性。

(2) 在微博平台上, 某用户微文被其他人转发, 表示这条微文得到了别人的重视, 并且有必要进行回应或扩散。因此, “转发”本身就意味着这条微文的影响力得到了别人的肯定。

(3) “转发”使用该用户的信息得到扩散, 达到更多的受众。这个过程对用户的影响力起到了放大作用。“评论”关系和“关注”关系则不能明显地扩散用户的影响力。

因此, 选择以“转发关系”来构建邻接图的边, 得到边的合集 E 和用户集 V 。对应到微博平台, 可以认为权威用户的微文应该被很多好的中心用户转发, 而

一个好的中心用户应该转发很多好的权威用户的微文。

3.2 考虑转发次数的 HITS 算法改进

在原始 HITS 算法中, 没有特别考虑一个网页中多次出现另一网页的链接的情况。但是在微博平台中, 用户 p 多次转发用户 q 的一条或多条微文是很常见的。这既是用户 p 活跃度的体现, 也是用户 q 影响力的体现。因此, 考虑到多次转发的邻接矩阵可以表示为:

$$L_{pq} = \begin{cases} m & \text{用户 } p \text{ 转发用户 } q \text{ 微文 } m \text{ 次} \\ 0 & \text{用户 } p \text{ 未转发用户 } q \text{ 的微文} \end{cases} \quad (3)$$

因为用户自己转发自己的微文可以视为扩大自己影响力和活跃度的方式之一。因此在公式 (3) 中, $L_{pp} = m$, m 为用户 p 转发自己微文的次数。

3.3 考虑用户权重的 HITS 算法改进

在微博平台上, 一条微文被一个粉丝数为 100 的人转发和被一个粉丝数为 100 万的人转发, 其意义是完全不同的。因此用户的粉丝数可以作为权重考虑到算法中来。

考虑用户权重的 HITS 算法第 i 次迭代过程为:

I 操作:

$$a(p)_i = \sum_{q: (p,q) \in E} h(q)_{i-1} \times w(p,q) \quad (4)$$

O 操作:

$$h(p)_i = \sum_{q: (p,q) \in E} a(q)_{i-1} \quad (5)$$

其中 $a(p)_i$ 表示第 i 次迭代后用户 p 的权威值, $h(p)_i$ 表示第 i 次迭代后用户 p 的中心值, $w(p,q)$ 表示用户 q 对用户 p 的权重。为了排除微博中大量存在“僵尸用户”的影响, 并缩小用户粉丝数的数量级差异, 找到粉丝数不是特别多的意见领袖, 通过公式 (6) 计算 $w(p,q)$ 的值:

$$w(p,q) = \sqrt{L_{pq} \times f(q)} \quad (6)$$

其中 L_{pq} 通过公式 (3) 计算得到, $f(q)$ 表示用户 q 的粉丝数。在引入粉丝数作为权重的时候, 中心值计算公式并未改变, 原因如下:

(1) 权威值计算的过程中已经考虑了粉丝数权重, 在中心值的计算过程中再加入将会过分放大粉丝的作用, 识别出来的就仅仅是“微博红人”而不是“意见领袖”;

(2) 微博中存在通过大量转发一些红人的微文来提升自己影响力的用户, 他们很少发表自己的意见, 这部分用户需要被弱化;

(3) 粉丝多的用户在一个社交圈子中被同时关注的可能性很高,多次转发对该用户影响力的提高有限。

在每次迭代完成后对权威值和中心值进行归一化处理,最终得到计算公式如下:

I 操作:

$$a(p)_i = \frac{\sum_{q: C(q,p) \in E} h(q)_{i-1} \times w(p,q)}{\sqrt{\sum_{y: y \in V} \left[\sum_{x: C(x,y) \in E} h(x)_{i-1} \times w(y,x) \right]^2}} \quad (7)$$

O 操作:

$$h(p)_i = \frac{\sum_{q: C(q,p) \in E} a(q)_{i-1}}{\sqrt{\sum_{y: y \in V} \left[\sum_{x: C(x,y) \in E} a(x)_{i-1} \right]^2}} \quad (8)$$

由于考虑了权重、转发关系等特点,改进后的 HITS 算法可以称为基于加权转发的 HITS 算法(HITS Base on Weighted Retweet, HITS-BOWR)。

4 数据准备

意见领袖大多都是话题依赖的^[9,13]。因此,本文仍然以话题为例来研究意见领袖。为了减少“网络水军”的干扰,避开娱乐、商业性话题,选取“阶梯电价”这一民生话题为例进行研究。利用新浪微博 API,以“阶梯电价”作为关键字,最后共随机获取 38 193 个微博用户和 53 563 条微文。微文时间跨度为 2012 年 4 月 20 日 4 时到 2012 年 7 月 3 日 17 时,而同一时段用新浪微博的搜索功能搜到 335 943 条微文,实际获取比例为 15.75%。

为了将用户限制在该话题下,避免 HITS 算法中存在的“主题漂移”现象,笔者取消了原算法中从根集扩充到基础集的过程,以直接获取到的用户集构造邻接矩阵。

由于新浪目前不提供直接获取转发路径上用户信息的接口,因此需要通过编程从微文中分离出转发路径上的中间用户。通过分析转发功能发现,转发内容是以“……//@ 用户昵称: ……”的格式表现的。因此,可以使用符号“//@”和“:”,利用正则表达式截取字符串的子串得到中间用户昵称。从获取的 53 563 条微文中,选出 10 421 个用户的 10 461 条转发的微文,以此构建出 10 940 条不重复的有向图的边。

微博平台不同于其他社交网络的一个显著性特点是媒体的大量参与。本文并未对用户集中的媒体用户进行特别处理。这是因为经过几十年的发展,意见领袖的概念已经从最初的“信息传播中间者”,扩大到“提供建议、施加影响的人”。而媒体用户的参与,使得意见领袖的信息传播作用相对弱化。另一方面,互联网的快速发展使得广大受众可以直接接触媒体。媒体为了满足受众的需求,开始发表评论,而不再是单纯的信息源^[4]。从这个角度来看,媒体与意见领袖在功能上开始有一定的重复。在很多基于微博平台意见领袖识别的研究中,并没有将媒体用户从意见领袖名单中除去^[9,11]。

5 实证分析

5.1 基本信息

以“阶梯电价”为关键字随机获取的用户集中,粉丝数最多的是微博客服,为 31 040 143 人;转发其他用户次数最多的是茶语 2000 和真爱网,均为 12 次;被转发次数最多的用户是南方都市报,次数为 477 次;其他情况如表 1 所示:

表 1 用户集基本统计情况

名次	粉丝数前 10 名用户		转发次数前 10 名用户		被转发次数前 10 名用户	
	用户名称	粉丝数	用户名称	次数	用户名称	次数
1	微博客服	31 040 143	茶语 2000	12	南方都市报	477
2	姚晨	20 352 767	真爱网	12	上海发布	233
3	何炅	18 162 651	高达无双 3	11	雷颐	216
4	王力宏	15 825 836	松鹤 73	11	环球家电网	215
5	头条新闻	12 807 946	正义的呼声 A	10	新浪财经	205
6	黄健翔	11 816 256	雅各布 Z	10	陈有西	202
7	冷笑话精选	10 308 402	零点电力研究中心	10	头条新闻	202
8	郎咸平	9 267 700	阳光的吉米小窝	10	每日上海	183
9	微博搞笑排行榜	9 090 762	趣稚部落格	10	叶匡政	169
10	任志强	8 110 875	马水旺	10	时代报	141

网页、博客、社交网络和社交网络都属于“无标度网络”^[20 21]，其主要特性是网络上的度分布服从幂律分布。与以往研究关注关系的文献不同，本文重点研究的是转发关系，在这种关系中，用户的被转发数是另外一种出度。根据网络中用户间的转发关系，可以得到出度分布如图 1 所示：

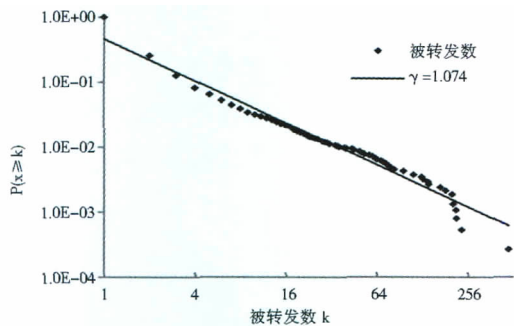


图 1 转发网络的用户度分布

从图 1 可以看出，基于转发关系的用户度分布仍然服从幂律分布，用户转发网络也是一种无标度网络。但是与一般幂律指数 $2 < \gamma < 3$ 不同^[21]，此网络的幂指数 $\gamma = 1.074$ 。这说明转发网络虽然也属于稀疏矩阵，但是网络中的出度比较大的节点较多，权威用户也可能比较分散。进一步统计后发现，在网络图的所有 10 940 条转发关系中，前 100 名被转发者所占比例为

49.70%，而前 100 名转发者所占比例仅仅为 3.69%。

5.2 HITS - BOWR 算法结果与分析

HITS 算法是收敛的，并且收敛速度很快。Kleinberg 认为迭代次数为 20 已经足够^[18]。根据公式 (7) 和公式 (8)，设置迭代次数为 30 次。图 2 显示了任志强、头条新闻、米瑞蓉、新浪财经 4 位用户的权威值和中心值收敛情况：

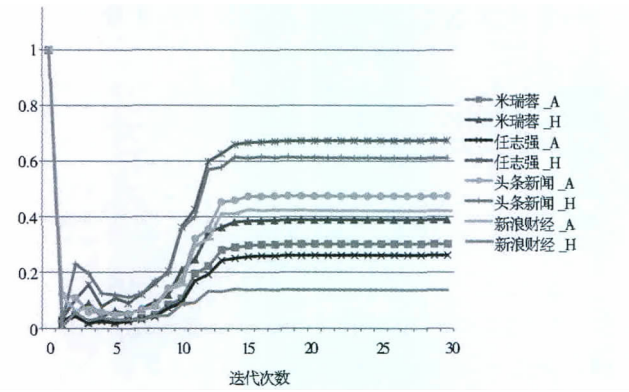


图 2 算法收敛情况

可以看到，HITS - BOWR 算法从第 15 次迭代开始收敛，权威值前 10 名用户和中心值前 10 名用户如表 2 所示。这些用户可以被认为意见领袖，本文分别称其为权威用户和中心用户。

表 2 前 10 名权威值和中心值用户

名次	权威值 TOP10			中心值 TOP10		
	用户名称	权威值	粉丝数	用户名称	中心值	粉丝数
1	头条新闻	0.4737	12 807 946	任志强	0.6741	8 110 875
2	新浪财经	0.4209	2 486 743	头条新闻	0.6108	12 807 946
3	毛振华	0.3375	1 006 862	米瑞蓉	0.3892	41 079
4	网络新闻联播	0.3143	142 624	新浪财经	0.1357	2 486 743
5	新浪房产	0.3019	926 355	中国新闻周刊	0.0486	2 084 967
6	米瑞蓉	0.3019	41 079	资讯大全	0.0160	4 566
7	任志强	0.2614	8 110 875	开进布拉格	0.0061	10 751
8	微天下	0.2430	1 485 652	李鸿文	0.0017	18 061
9	新浪评论	0.1718	368 409	陈有西	0.0011	120 799
10	中国新闻周刊	0.1509	2 084 967	雷颐	0.0004	105 268

从表 2 可以看出，上榜的用户粉丝数差别比较大：权威值上榜的用户中粉丝最少的为米瑞蓉，粉丝数为 41 079；中心值上榜的用户中粉丝最少的为资讯大全，粉丝数仅为 4 566。比较而言，前 10 位用户权威值差别不大，权威值比较分散，这一现象实际上与图 1 的结论一致；而中心值的分布则非常集中，只有 4 位用户超过 0.1，这表明虽然信息源较多，但是信息渠道其实比

较有限。这初步说明 HITS - BOWR 算法是一种比较有效的意见领袖识别算法。

微博中用户可以分为三类：媒体微博、微博名人（或个人认证用户）和草根用户（非认证）^[22]。从表 2 可以看出，权威值前 10 名的用户中，媒体类微博有 7 个，占绝大多数。其余三位则是名人，没有草根用户。前 10 名中心值用户中，媒体类微博则只有 3 名，认证

从图 3 中可以看出,此网可以分为两个社区:以李鸿文、陈有西、雷颐为核心的左下角社区,超过 1 100 个节点;以头条新闻、新浪财经和任志强为核心的右上角社区。左下角的社区占全部节点总数和转发总数的 60% 以上。但主要的权威用户和中心用户却集中在右上角的社区。另外,微博被任志强转发的用户只有两个,分别是米瑞蓉和毛振华,他们也因此进入权威值前 10 名。这说明 HITS - BOWR 算法的一个重要特征,即对转发质量的考虑优于转发数量。

转发网络还解释了米瑞蓉、资讯大全、开进布拉格等粉丝仅为几千到几万的用户能够进入中心值前 10 名的原因。资讯大全和米瑞蓉作为仅有的一条通路连接了两个社区,开进布拉格则是网络中最大环路的开端。HITS - BOWR 算法在识别关键节点上有很好的表现。

采用构建意见领袖精简网的方法进一步分析意见领袖的其他特征。通过对图 3 迭代删除度为 1 的节点,直到不能再删除为止^[23],可以得到包含 64 个节点的精简网,如图 4 所示:

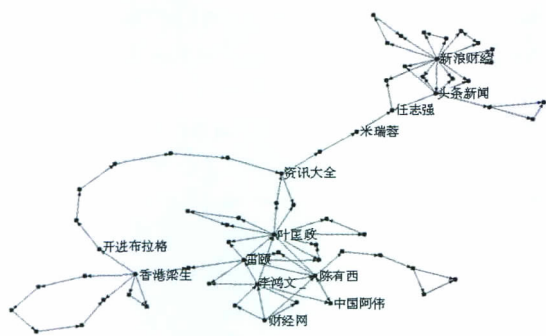


图 4 精简意见领袖网

从精简网中可以发现: 15 个意见领袖中,有 10 个最终留在了精简网上。另外,香港梁生的权威值排名为第 13 位;财经网权威值排名为第 17 位;叶匡政的中心值排名为第 12 位;中国阿伟中心值排名为第 13 位。他们可以算作准意见领袖。如此高的意见领袖比例体现了意见领袖在信息传播中的中枢地位,也再一次证明,意见领袖的作用并没有因为微博中信息源的增多而削弱。

从网络图中还可以发现领袖的另外一个特征: 他们大多都处于回路中(不考虑路径方向)。很多用户除了直接与意见领袖连接外,还会通过其他普通用户

与意见领袖连接,这种回路说明意见领袖在用户中有持续的影响力,如香港梁生的回路。另一种回路则是以新浪财经与头条新闻所在的回路为代表,即意见领袖之间除了与其他用户建立连接外,他们之间也有直接连接。考虑到箭头方向和用户背后的关系,可以认为这是意见领袖喜欢通过相互交流相互转发来扩大影响力。

6 结 语

以微博话题“阶梯电价”下的用户转发关系构建网络,通过改进后的 HITS - BOWR 算法挖掘微博转发网络中的意见领袖,并进一步构建意见领袖网络图,以识别意见领袖特征,验证分析算法的有效性。研究结果表明,HITS - BOWR 算法能够有效地发现意见领袖。通过相关性分析发现:意见领袖的中心值与粉丝数有很大的相关性,这说明粉丝数是用户成为中心用户的一个重要条件。而且意见领袖的确兼具“信息中间人”和“有影响力的人”两种角色。通过对意见领袖网的分析发现,意见领袖在关键性节点上占据重要位置;意见领袖的作用并没有因为微博中信息源的增多而削弱;意见领袖之间存在通过相互交流转发来扩大自己的影响力。

本文在确定邻接边权重的时候,只考虑粉丝数,没有考虑意见领袖的号召力等更直接的因素。未来研究可以通过情感识别,分析意见领袖的号召力等指标。另外,因为资源限制,本文重点研究了话题“阶梯电价”中的意见领袖识别和特征分析。从结果来看,得到的意见领袖特征都是可以解释并理解的。这些特征也没有局限在此话题下,而是呈现一定的一般性特点。未来还可以进一步结合其他话题,验证并研究意见领袖一般性特征和他们在信息传播过程中的作用。

参考文献:

- [1] Lazarsfeld P F, Berelson B, Gaudet H. The People's Choice [M]. New York: Columbia University Press, 1948.
- [2] Rogers E M, Shoemaker F F. Communication of Innovations: A Cross-cultural Approach [M]. New York: Free Press, 1971.
- [3] Vergani M. Are Party Activists Potential Opinion Leaders? [J]. Javnost - The Public 2011, 18(3): 71 - 82.
- [4] 姜珊珊,李欲晓,徐敬宏. 非常规突发事件网络舆情中的意见

- 领袖分析[J]. 情报理论与实践, 2010, 33(12): 101-104. (Jiang Shanshan, Li Yuxiao, Xu Jinghong. Analysis of Internet Public Opinion Leaders in Unconventional Emergencies[J]. *Information Studies: Theory & Application*, 2010, 33(12): 101-104.)
- [5] 元志润. 在网络口碑沟通中的意见领袖(Opinion Leader)研究——以韩国超级博客(Power Blog)为例[D]. 上海: 复旦大学, 2011. (Yuan Zhirun. Opinion Leaders in Online World - of - Mouth——Power Bloggers in Korea[D]. Shanghai: Fudan University, 2011.)
- [6] Borondo J, Morales A J, Losada J C, et al. Characterizing and Modeling an Electoral Campaign in the Context of Twitter: 2011 Spanish Presidential Election as a Case Study[J]. *Chaos*, 2012, 22(2): 023138. doi: 10.1063/1.4729139.
- [7] Rui H X, Liu Y Z, Whinston A. Whose and What Chatter Matters? The Effect of Tweets on Movie Sales[J]. *Decision Support Systems*, 2013. In Press.
- [8] 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): 82-91. (Ding Hanqing, Wang Yaping. Analyzing “Opinion Leader” Attributes in SNS Cyberspace: An Investigation of Douban.com[J]. *Journalism & Communication*, 2010(3): 82-91.)
- [9] 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. (Liu Zhiming, Liu Lu. Recognition and Analysis of Opinion Leaders in Microblog Public Opinions[J]. *Systems Engineering* 2011, 29(6): 8-16.)
- [10] 丁雪峰, 胡勇, 赵文. 网络舆论意见领袖特征研究[J]. 四川大学学报: 工程科学版, 2010, 42(2): 145-149. (Ding Xuefeng, Hu Yong, Zhao Wen, et al. A Study on the Characters of the Public Opinion Leader in Web BBS[J]. *Journal of Sichuan University: Engineering Science Edition* 2010, 42(2): 145-149.)
- [11] 王君泽, 王雅蕾, 禹航. 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. (Wang Junze, Wang Yalei, Yu Hang, et al. Study on the Recognition Model for Opinion Leader on Micro-blog[J]. *Journalism & Communication*, 2011(6): 81-88.)
- [12] Weng J S, Lim E P, Jing J, et al. TwitterRank: Finding Topic-sensitive Influential Twitterers[C]. In: *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining*. New York: ACM, 2010: 261-270.
- [13] 肖宇, 许炜, 夏霖. 一种基于情感倾向分析的网络团体意见领袖识别算法[J]. 计算机科学, 2012, 39(2): 34-37. (Xiao Yu, Xu Wei, Xia Lin. Networking Groups Opinion Leader Identification Algorithms Based on Sentiment Analysis[J]. *Computer Science*, 2012, 39(2): 34-37.)
- [14] Zhang J, Ackerman M S, Adamic L. Expertise Networks in Online Communities: Structure and Algorithms[C]. In: *Proceedings of the 16th International Conference on World Wide Web*. New York: ACM, 2007: 221-230.
- [15] Yoganarasimhan H. Impact of Social Network Structure on Content Propagation: A Study Using YouTube Data[J]. *Quantitative Marketing & Economics*, 2012, 10(1): 111-150.
- [16] 张伟哲, 王佰玲, 何慧. 基于异质网络的意见领袖社区发现[J]. 电子学报, 2012, 40(10): 1927-1932. (Zhang Weizhe, Wang Bailing, He Hui, et al. Public Opinion Leader Community Mining Based on the Heterogeneous Network[J]. *Acta Electronica Sinica*, 2012, 40(10): 1927-1932.)
- [17] Kwak H, Lee C, Park H, et al. What is Twitter, A Social Network or a News Media? [C]. In: *Proceedings of the 19th International Conference on World Wide Web*. New York, NY, USA: ACM, 2010: 591-600.
- [18] Kleinberg J M. Authoritative Sources in a Hyperlinked Environment [C]. In: *Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms*. 1998: 668-677.
- [19] 苏成, 潘云涛, 袁军鹏. 基于 HITS 算法的期刊评价研究[J]. 编辑学报, 2009, 21(4): 366-369. (Su Cheng, Pan Yuntao, Yuan Junpeng, et al. HITS for Journal Ranking[J]. *Acta Editor-logica*, 2009, 21(4): 366-369.)
- [20] 朱恒民, 李青. 面向话题衍生性的微博网络舆情传播模型研究[J]. 现代图书情报技术, 2012(5): 60-64. (Zhu Hengmin, Li Qing. Public Opinion Propagation Model with Topic Derivatives in the Micro-blog Network[J]. *New Technology of Library and Information Service*, 2012(5): 60-64.)
- [21] Java A, Song X, Finin T, et al. Why We Twitter: Understanding Microblogging Usage and Communities[C]. In: *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*. New York, NY, USA: ACM, 2007: 56-65.
- [22] Cha M, Benevenuto F, Haddadi H, et al. The World of Connections and Information Flow in Twitter[J]. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 2012, 42(4): 991-998.
- [23] Valente T W, Fujimoto K. Bridging: Locating Critical Connectors in a Network[J]. *Social Networks*, 2010, 32(3): 212-220.
- (作者 E-mail: mtxycc@126.com)