

# 微博影响力评价研究进展

李军, 陈震, 黄霁崑

清华大学 计算机科学与技术系  
信息技术研究院网络安全实验室  
信息国家实验室  
北京 100084

e-mail: ljun852@gmail.com

**摘要:** 微博作为一种广泛使用的信息媒介平台, 其多样特性满足了人们信息获取、人际交流等多方面的新需求, 微博已成为人们日常生活的重要组成部分之一。微博影响力作为微博研究的一个热点, 对于优化微博的信息整合、传输, 以及了解微博发展方向、促进信息传播和网络发展具有非常重要的理论和实际意义。本文分析了微博信息传播的特点, 总结了目前主要的微博影响力评价方法, 并提出了针对中国微博的影响力评价模型和方法。

**关键词:** 微博媒体; 影响力; 新闻传播; 舆论安全

## Micro-blog Impact Evaluation Study

LI Jun, CHEN Zhen, HUANG Ji-wei

Department of Computer Science and Technology

Research Institute of Information Technology (RIIT)

Tsinghua National Lab for Information Science and Technology (TNLIST)

Tsinghua University, Beijing, China, 100084

e-mail: ljun852@gmail.com

**Abstract:** Microblogging(we call it Weibo) is a widespread use of information media platform, which variety of features meet people's information acquisition, interpersonal communication such as a wide range of new demands. Microblogging has become an important part of daily life. The research on microblogging influence is a hot topic currently, which has a very important theoretical and practical significance on optimization of microblogging information integration, transmission and understanding the microblogging direction of development. The microblogging influence will promote information dissemination and network development. This paper analyzes the characteristics of the microblogging information dissemination, and summarizes the main microblogging influence evaluation method. Finally, we present the influence of evaluation models and methods for Chinese Weibo.

**Keywords:** Weibo Media, Influence, Journalism and Communication, consensus safety

---

**基金项目:** 本文研究受国家 973 项目 (编号: No.2011CD302600, No.2011CB302805, No.2012CB315800) 和国家自然科学基金 A3 重点基金项目 (编号: No.61161140320) 资助。

**作者简介:** 李军 (1980-), 男, 山东, 硕士生, 主要研究领域为网络存储, 网络安全; 陈震 (1976-), 男, 浙江, 副研究员, 工学博士, 主要研究领域是网络安全, 云存储, 社交媒体; 黄霁崑 (1987-), 男, 北京, 博士生, 主要研究领域计算机网络和计算机系统的性能评价。

本文通信联系人: 李军, 北京, 100084, 清华大学信息技术研究院网络安全实验室

## 1. 引言

微博作为一种便捷的媒体交互平台，在全球范围吸引了数亿用户，已成为人们进行信息交流的重要媒介。用户通过微博可以进行信息的记录和交流、娱乐消遣，以及社会交往等[21]。微博用户作为微博的重要组成部分，其使用微博的动机、方式和行为模式都是微博研究领域的研究热点，微博用户影响力（简称微博影响力）作为影响力研究在微博领域的延伸，吸引了大量研究者开展研究，本文通过对当前主要微博影响力研究方法进行分析、归纳和分类，并针对中国微博提出了影响力评价方法。文章的将在第2章节介绍微博及其特点等研究背景，在第3章节详细介绍微博影响力评价方法研究情况，并在第4章节阐述我们提出的微博影响力评价方法，最后进行全文总结。

## 2. 研究背景

微博，即微型博客，又称微博客、迷博、一句话博客等。维基百科关于微博的定义是，微博客（Micro-blogging或Microblog）是一种允许用户及时更新简短文本，并可以公开发布的博客形式。它允许任何人阅读或者只能由用户选择的群组阅读。这些消息可以被很多方式传送，包括短信、实时消息软件、电子邮件或网页。微博也可以发布多媒体，如图片或影音剪辑。我们通常所使用的微博每次发布的信息限定140个字符左右。2006年，伊万·威廉斯（Evan Williams）、比孜·斯通（Biz Stone）和杰克·多尔西（Jack Dorsey）联合创办了Twitter网站，成为世界上最早提供微博服务的网站，目前其活跃用户数量已经超过1亿。在国内，2007年，原校内网创始人王兴首先建立了饭否网，并将微博概念引入中国。自2009年下半年以来，新浪、腾讯、搜狐、网易以及国家重点新闻网站新华网、人民网等纷纷开启微博服务。截至目前，新浪微博、腾讯微博的注册用户量均已超过2亿。中国社科院发布的《2011年中国社会形势分析与预测》蓝皮书指出，微博正在改变着我国互联网舆论载体的格局，微博已逐渐发展成为介入公共事务的新媒体，成为网络舆论中最具影响力的一种。

关于影响力，一般认为是用一种以他人所乐于接受的方式，改变他人所乐于接受的方式，改变他人的思想和行动的能力。影响力在社会学、经济学和政治学等领域已开展了很多研究[9]。影响力在市场营销、时尚传播等方面发挥着潜移默化的作用。

微博以大量的活跃用户、大量的流动信息，已成为一种重要的、广泛应用的信息承载和传输媒介，对我们的日常工作和生活所产生的影响也越来越大[18]。微博影响力是用户在微博中重要性的衡量，用户影响力作为用户的重要指标，是微博关系的基础，用户影响力越大，所受到的关注程度也就越高，对网络的影响作用也就越大。对用户影响力进行衡量，能对网络中的用户进行合理排序，为微博的拓展应用以及进一步发展提供理论基础。比如，当我们想寻找网络中满足一定条件的比如具有相同兴趣或者相同出游计划的用户时，满足条件的用户可能会很多，如何在最短时间选取最合适的结果，用户影响力就是一个非常重要的指标。再如，随着微博的发展，未来更多的商业将被移植到网络上来。当我们想在微博上推广某种商品或服务时，如何选取首批推荐使用产品的人群会直接影响到推广活动的效果及代价。此时，影响力比较大的用户往往成为活动的首选人群。通过他们的介绍和宣传，能将商品更快、更有效的传播给网络中其他用户[1, 2, 10, 14]。

微博近几年得到了迅速发展，微博用户群体也日益扩大。对微博用户行为及其影响力开展研究，对于优化微博的信息整合、传输，对于我们深入了解微博发展方向、促进网络经济发展具有非常重要的理论和实际意义。

## 3. 微博影响力评价方法

### 3.1 微博影响力相关定义

文献[3, 17]采用了韦氏字典（Merriam-Webster dictionary）对影响力的定义，认为影响力是直接或间接产生影响的能力。文献[19]认为，用户影响力本质上是用户间的相互作用，一个用户能够对其他用户产生的作用越大，其影响力就越大。有的研究者采用了社交影响力（Social Influence）的概念，社交影响力是指个

体的思想、感受、态度或行为因应其他个体或组织而发生改变。文献[18]认为社交影响力是使人改变其行为以使其更接近他的朋友的行为。文献[8]将影响力细分为3种，即粉丝、回复、转推影响力，分别对应具体的用户行为。总的来看，大多数研究者基本上沿用了传统领域中关于影响力的定义和概念。

### 3.2 微博影响力评价影响因素

微博影响力的构成涉及多个因素，比较直观的一个因素是“粉丝数”（follower），比如，在新浪微博上，姚晨的粉丝数已超过了1510万（截至2011年12月）。我们认为粉丝数从一定程度上能折射出用户的影响力，但更多的是反映“人气”的旺盛与否，或者说用“吸引力”来表述更为准确，并且粉丝数的背后有可能存在“僵尸粉丝”，在考虑这一因素时需要注意这一点。

再则，可以考虑的因素是“用户行为”，微博用户的主要行为有：发布（*post*）、回复（*posted*）、跟随/被跟随（*follow/followed*）、转推/被转推（*retweet/retweeted*）、提及/被提及（*mention/mentioned*）、评论/被评论（*comment/commented*，这是中国微博所特有的）。从影响力的定义可以看出，行为可以影响他人或是被他人影响，这些行为作为进行用户影响力评价的可参考因素，很多研究者从微博用户行为角度着手开展了研究（将在3.3中介绍）。

### 3.3 微博影响力评价方法

目前，在微博影响力研究方面，主要的评价刻画可以归纳为以下4种具有代表性的类型：基于PageRank的评价方法、基于用户行为权值的评价方法、基于PageRank和用户行为权值的评价方法、基于URL追踪的评价方法。

#### 3.3.1 基于 PageRank 的评价方法

PageRank[13, 20]由Google创始人拉里·佩奇（Larry·Page）和谢尔盖·布林（Sergey·Brin）提出，其基本思想是利用网页的超链接结构确定网页的重要性，即若网页 $v$ 的一个链接指向 $u$ ，就认为 $v$ 投了网页 $u$ 一票，PageRank排名系统会根据网页 $B$ 收到的投票数量来评估该网页的重要性。

这里对PageRank做一简要介绍，其计算公式是：

$$R_{(u)} = C \cdot \sum_{v \in B(u)} R_{(v)} / N_{(v)}$$

其中， $u$ 和 $v$ 表示两个不同的网页， $R_{(u)}$ 和 $R_{(v)}$ 分别表示 $u$ 和 $v$ 的PR值， $B(u)$ 为 $u$ 的所有反向链接（即由外部其他网页指向 $u$ 的链接）数量， $N_{(v)}$ 为 $v$ 的所有正向链接数量， $C$ 为规范化因子，以使网页的PR值是常数。显然这是一个递归公式，一个网页的等级值会均匀地分配给它的每个正向链接，表明PR值是众多网页中形成的一个分布概率，所有网页的PR值之和等于1。网页的PR值是由网络的超链接结构所产生的一个关于网页重要性指标的等级值，所有网页的PR值都可以通过其他网页的PR值和超链接的具体数量来计算得到，即将所有链接到它的网页的PR值除以各自正向链接数的商进行总体求和。

文献[5]根据PageRank设计了基于用户和链接结构的话题相似性的影响力排序算法。文献[19]借鉴PageRank，认为好友数量是用户影响力的一个重要指标，用户的好友数量越多，其影响力也越大，越容易影响到其他人，而所花费的代价也越小。其UserRank模型的重要参数是用户朋友对其影响力的贡献值。作者在PageRank模型基础上提出了用户影响力评估模型UserRank，其计算公式基本与PageRank一致。

#### 3.3.2 基于用户行为权值的评价方法

Meeyoung Cha等在文献[3]中，对用户行为和与之相应的影响力进行了深入研究。作者关注了三种用户行为：被跟随（文中用*indegree*来表述，其实就是图的入度）、转推（*retweet*）和提及（*mention*），并分析了这三种行为所表征的用户影响力类型，文章认为最能被跟随的用户往往传播了大量的公开图片和信息源；最能被转推的用户往往是内容聚合服务商、商务人士和新闻网站；最能被提及的用户是往往是名人或焦点人物。文章采用的比较用户影响力的方法是分别按照前述三种行为进行影响力测量，再将所得数值通过斯

皮尔曼等级相关系数 (Spearman's rank correlation coefficient) [11]进行两两比较来得出排名结果。

文献[8]将社交影响力划分为3种影响力5种排序准则, 分别是粉丝数量影响力、回复影响力、转推影响力, 按粉丝数、消息数(回复和转推)、响应用户数(回复者和转推者)的排序准则。通过采用斯皮尔曼等级相关系数和肯德尔等级相关系数 (Kendall Tau rank correlation coefficient) [12], 对5种排序准则进行了计算比较, 认为拥有最多回复者的用户集合是最稳定的, 将按照回复者数量进行社交影响力排序作为标准。

### 3.3.3 基于 PageRank 和用户行为权值的评价方法

在文献[4]中, 作者发现一个被许多话语权威跟随的用户有可能成为话语权威; 一个被许多话语权威转载的内容有可能是有用的内容; 一个发布了许多有用内容的用户有可能成为话语权威。基于这样的认识, 文章借鉴ObjectRank方法(基于PageRank) [15], 设计了TURank (Twitter User Rank)方法, 通过构造基于链接分析的用户-内容图 (User-Tweet Graph) 来进行用户排名, 它关注用户的发布/被回复、跟随/被跟随、转推/被转推行为[16], 并给出了用户行为权重计算公式。在计算时, 人工赋予各用户行为权值的初值。

### 3.3.4 基于 URL 追踪的评价方法

在文献[6]中, 作者指出口碑信息通过许多级联进行传播, 大多是由普通个体引发的, 也是普遍被接受的[7]。口碑相传受到少数具有关键影响力人物的驱动, 而最具影响力的用户同样也是费效比最好的。其提出的计算微博影响力的方法是, 从URL所在的受欢迎的种子节点开始, 通过一系列跟随者的报告来追踪URL传播情况, 直到传播终止。采用三种方法来分配相应的影响力: 将全部影响力贡献分配给第一个发送URL的人; 将全部影响力贡献分配给最近以来(最后一个)发送URL的人; 将影响力贡献平均分配给所有发送URL的人。

对上述4种类型方法做一小结:

基于PageRank的方法, 归结到微博特性上, 就是从用户的粉丝(数量、质量)着手来衡量微博影响力, 这类型方法简单、实用, 需要的参量少, 便于操作实现。需要注意到是, 参量少是一把双刃剑, 在便于操作实现的同时, 也带来了客观性、准确性上的问题, 因为“僵尸粉丝”的存在, 其客观性、准确性会受到一定程度的削弱。

用户行为权值计算方法, 将用户行为列入考察范围, 考虑的因素更全面一些。但有的方法如[3]和[8]在衡量时也仅选取一个参量(如粉丝数)来进行度量, 其客观性、准确性也会受到一定影响。URL追踪的方法在实现上不难, 但[6]中所描述的方法的问题在于如何抉择使用哪个方法。

PageRank+用户行为权值计算方法, 典型代表就是TURank, 它发掘了话语权威的作用, 通过构造用户-内容图较好地揭示了用户和信息、用户和用户以及信息和信息之间的相互关系, 避免了只以粉丝数量(被跟随)来进行用户排名的所造成的客观性、准确性上的缺失, 为取得一个较为准确、合理的排名结果提供了一个很好的机制。但也要注意, TURank的用户行为初值(即图中边的初值)是人工赋予的, 边的权值设置是排名结果准确合理与否的重要因素, 要使TURank有效, 要在分析数据集中边的数量对比情况的基础上来设定合理的边权值。

还要注意一点, 在微博影响力评价方法中, 特别是在进行微博用户行为间关系分析比较时, 统计学的方法(如斯皮尔曼等级相关系数和肯德尔等级相关系数)被广泛应用, 为发掘、建立微博用户行为间相互关系提供了可信的支持。例如, 转推(*retweet*)和提及(*mention*)这两种用户行为, 从直观上看, 用户在进行转推时会提及发布者, 在文献[3]中, 将这两种用户行为的相关数据通过斯皮尔曼等级相关系数进行计算比较, 所得结果很接近, 从而说明并验证了这两种行为间的紧密联系。由此可见, 统计分析是微博影响力分析不可忽视的环节。

## 4. Weibo User Rank

通过上一节, 我们了解了当前4种主要的微博用户影响力评价方法, 可以看出, PageRank的设计思想简单、实用, 可操作性强, 在排名方面可以作为一种基础参考。在3.3.3中介绍的TURank将PageRank与用户行

为权值相结合，特别是所构造的用户-内容图较好地描述微博主要关系。因此，我们以TURank为基础提出针对中国微博的影响力评价模型WURank——Weibo User Rank。

TURank是针对Twitter提出的，Twitter与中国微博最大的不同在于“评论”功能。在Twitter中，用户不能对其粉丝发布的内容进行评论。在中国微博，大多数运营商如新浪微博、腾讯微博等，都向用户提供了“评论”功能，用户可以评论其他用户发布的内容，这使得信息中间人的相互作用更加复杂。

4.1 基本定义

我们给出两个定义：

定义1：用户微博影响力：可以表征为用户发布一则消息并使其传播到其他用户的能力。

定义2：中国微博影响力的量化：用户微博影响力可以量化表示为某一用户与该用户发布的微博在所有用户和微博范围内的排名得分。

为了体现中国微博的特点，我们将微博标记为Weibo，在下面的图中代表所用户所发布的内容。

4.2 模型设计

参考目前已有相关研究工作，结合中国微博的特点，我们设计的微博影响力评价模型如图1所示。

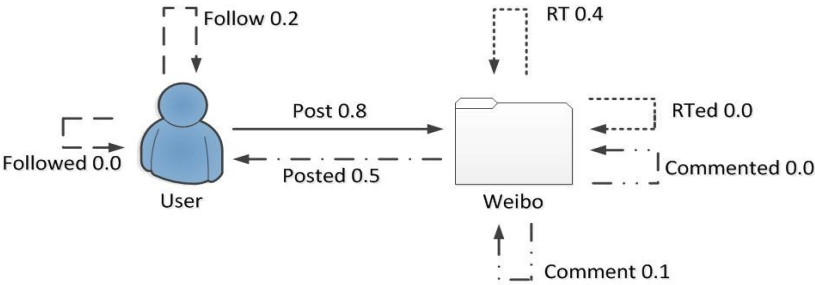


图1 User-Weibo Schema Graph

图1展示了用户-微博机制图（User-Weibo Schema Graph）的基本构成，节点包括所有用户和信息（即用户所发布的内容，我们称之为Weibo），边的含义为发布/被发布、回复/被回复、跟随/被跟随、转推/被转推、评论/被评论。微博中的每个行为对应一对节点间不同方向的两条边。模型中每条边的权值代表了该微博行为对于用户或消息的影响力的贡献：例如用户通过自身影响力直接影响所发布微博的影响力，其权值定义为0.8，对应图1中 $Post$ 边；而微博由于被其他用户或微博转推（RT）或评论（Comment）可以提高自身影响力，并对发布该微博的用户进行影响力的正反馈（ $Posted$ ）。为了保证模型的合理性和算法的收敛性，这里要求每个用户或微博节点的出边权值之和必须为1。

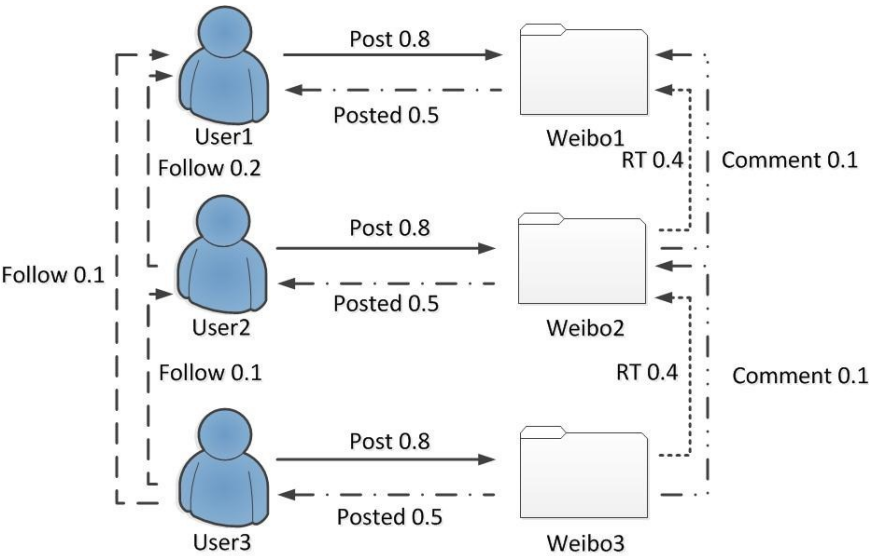


图2 User-Weibo Graph

图2展示了用户-微博关系图 (User-Weibo Graph)  $UWG = (V, E)$ ,  $V$ 包括所有用户节点和微博节点,  $E$ 代表用户操作行为, 边的权值 $w$ 的计算公式是:

$$w(e) = \frac{w(e_s)}{OutDeg(u, e_s)}$$

其中 $e_s \in E_s$ , 是和 $e$ 同一类型的边,  $OutDeg(u, e_s)$ 是节点 $u$ 的出边 $e_s$ 的数量。

#### 4.3 计算方法

根据用户-微博关系图, 我们来计算用户得分以进行排名。用户得分计算基于如下公式, 该公式基于ObjectRank[16]。

$$r = dAr + \frac{(1-d)}{|V|}e$$

其中 $r$ 是ObjectRank得分向量,  $d$ 是随机跳的概率,  $A$ 是转移矩阵。

用户得分可以通过迭代的算法得到, 该算法伪码如下:

**WURank**

$r^0 \leftarrow [1, \dots, 1]$

$\alpha \leftarrow 0$

**Repeat**

$\alpha \leftarrow \alpha + 1$

**foreach**  $r_i^\alpha \leftarrow r^\alpha$

$r_i^\alpha \leftarrow \sum_{e=(j,i) \in E} w(e) r_j^{\alpha-1} + (1-d)/|V|$

**end**

$r^\alpha \leftarrow r^\alpha / \|r^\alpha\|_1$

**until**  $\|r^\alpha - r^{\alpha-1}\|_1 < \epsilon$

**return**  $r^\alpha$

**end**

## 5. 结束语

微博作为一种广泛应用的信息媒介, 已成为人们获取信息、交换信息的重要途径。如何利用微博扩散信息、宣传产品、提高知名度已成为个人、团体特别是商业营销关注的重点问题。而微博影响力作为用户的重要衡量指标, 是微博关系的基础, 用户影响力越大, 所受到的关注程度也就越高, 对网络的影响、对信息的传播作用也就越大。通过影响力大的用户进行信息扩散、商品推介和宣传, 无疑将收到事半功倍的效果, 这对于商业营销来说意义重大。国内外已经开展了针对Twitter及其用户影响力的研究, 我们的工作细致分析对当前微博影响力评价方法的基础上, 通过借鉴已有方法的精髓, 针对中国微博的特点, 提出了中国微博影响力评价模型。这一模型充分考虑了中国微博的信息流和功能特点, 较好地揭示了用户、微博之间的两两相互关系, 便于理解, 易于实现。我们希望该模型可以对微博的理解与分析、用户及信息的影响力和价值的评价, 以及消息传播模式的挖掘分析等工作具有一定的借鉴和指导意义。

## References:

- [1] Akshay Java et al., Why We Twitter: Understanding Microblogging Usage and Communities[C]. 9th WEBKDD and 1st SNA-KDD Workshop 2007. 7-9.
- [2] Dejin Zhao et al., How and Why People Twitter: The Role that Micro-blogging Plays in Informal Communication at Work[C]. GROUP 2009 of ACM. 247-249.
- [3] Meeyoung Cha et al., Measuring User Influence in Twitter: The Million Follower Fallacy[C]. AAAI 2010. 11-13.
- [4] Yuto Yamaguchi et al., TURank: Twitter User Ranking Based on User-Tweet Graph Analysis[C]. WISE 2010. 243-246.
- [5] Jianshu Weng et al., Twitterrank: Finding topic-sensitive influential twitterers[C]. WSDM 2010. 7-9.
- [6] Eytan Bakshy et al., Everyone's an Influencer: Quantifying Influence on Twitter[C]. WSDM 2011. 67-69.

- [7] D.J.Watts. Influentials, networks, and public opinion formation[J]. Journal of Consumer Research, 2007: 7-11.
- [8] Shaozhi Ye, S. Felix Wu. Measuring Message Propagation and Social Influence on Twitter.com[C]. SocInfo 2010. 223-228.
- [9] Gabriel Weimann. The Influentials: People Who Influence People[M]. State University of New York Press, Albany, NY, 1994: 468-470.
- [10] Edward Mischaud, Twitter: Expressions of the Whole Self[M]. LSE-MEDIA@LSE, 2007: 7-9.
- [11] Jerrold H. Zar. Significance Testing of the Spearman Rank Correlation Coefficient[J]. Journal of the American Statistical Asso, vol. 67, No. 339, Sep.1972: 578-581.
- [12] Pranab Kumar Sen. Estimates of the Regression Coefficient Based on Kendall's Tau[J]. Journal of the American Statistical Association, Vol. 63, No. 324, Dec, 1968: 1379-1382.
- [13] Page Lawrence, Brin Sergey, etc. The PageRank Citation Ranking: Bringing Order to the Web[R]. Technical report, Stanford Digital Library Technologies Project (1998), <http://ilpubs.stanford.edu:8090/422/>
- [14] Goyal, A et al., Learning Influence Probabilities In Social Networks[C]. WSDM 2010. 241-243.
- [15] Andrey Balmin, et al., Objectrank: Authority-based keyword search in databases[C]. VLDB 2004. 565-569.
- [16] Danah Boyd et al., Tweet, Tweet, Retweet: Conversational aspects of retweeting on twitter[C]. HICSS-43, IEEE 2010. 4-10.
- [17] Fabrício Benevenuto, Tiago Rodrigues, Meeyoung Cha, Virgílio Almeida. Characterizing User Behavior in Online Social Networks[C]. IMC 2009. 54-58.
- [18] Louis Yu et al., What Trends in Chinese Social Media[C]. SNA-KDD'2011. 2-4.
- [19] 刘耀庭. 社交网络结构研究[D]. 浙江大学, 2008 年 6 月.
- [20] 袁鹏亮. 基于 Pagerank 网页排名技术计算方法的探讨[J]. 农业网络信息, 2007 年第 12 期: 171-172.
- [21] 王 娟. 微博客用户的使用动机与行为[D]. 山东大学, 2010 年 4 月.