

一个基于微博的局域社交网络结构和社会关系分析*

A Data Analysis of Social Relations and Structure based on Local Social Network in Micro Blog

涂师师¹, 贾 贞^{1,2**}TU Shi-shi¹, JIA Zhen^{1,2}

(1. 桂林理工大学理学院, 广西桂林 541004; 2. 广西空间信息与测绘重点实验室, 广西桂林 541004)

(1. College of Science, Guilin University of Technology, Guilin, Guangxi, 541004, China;
2. Guangxi Key Laboratory of Spatial Information and Geomatics, Guilin, Guangxi, 541004, China)

摘要:选取新浪微博中标签为桂林理工大学(桂林理工大学片区)的部分注册用户关系数据,构建一个基于微博的局域社交网络,应用社会网络分析 UCINET 软件对该社交网络的基本结构特性、网络成员的派系、地位等做了实证分析,并通过一条微博转播的实际数据验证了分析结果的可信性。

关键词:社会网络分析 互联网 微博 UCINET

中图分类号:TP393.092 文献标识码:A 文章编号:1005-9164(2013)01-0075-04

Abstract:Based on the relational data of the registered users labeled as “Guilin University of Technology” (Guilin University of Technology area) in the Sina Micro Blog, a local area social network was constructed. Its basic structure characteristics, each network member's cliques, status, and so on were analyzed by means of the social network analysis software-UCINET, then the credibility of the results was verified through some examples.

Key words:social network analysis, internet, micro blog, UCINET

社会网络分析^[1]是通过研究网络关系,把个体间关系、微观网络与大规模的社会系统的宏观结构联系起来,运用图论、统计分析等定量分析方法来考察社会网络的结构特性和网络上各种事件的相互联系与传播的规律^[2]。微博是一种基于用户关系的信息分享、传播以及获取的平台,人们根据教育背景、兴趣爱好、职业背景等方面的差异有选择地对其他用户进行关注而成为其“粉丝”,从而形成一个虚拟的“圈子”。新浪微博是目前国内人气最高的微博服务网站之一,在这个微博网络中如果把传播主体(注册用户)当成节点,把用户之间的关注关系抽象为节点之间的连

线,那么微博用户及其关注关系就构成一个社会网络(以下称微博网络)^[3]。目前,研究人员主要从以下三个方面对微博网络进行研究:(1)微博网络的整体拓扑结构分析;(2)微博网络的媒体传播特征;(3)微博网络的社会网络分析。研究发现,微博网络属于典型的无标度网络,表现出小世界现象,微博用户的关注数分布呈现长尾现象,用户的被关注数遵循幂率分布^[4]。文献^[5]分别以新浪、腾讯、搜狐三大微博网站的注册用户作为研究对象,根据用户之间的关注关系构建了复杂网络^[6,7],分别给出该网络的度分布、聚集系数以及平均路径长度等统计特征,得出微博用户关系网络具有无标度特性以及小世界效应的结论。研究微博网络的媒体传播特征时是把微博视为一种新型媒体,从传播学角度来研究微博对信息的传播影响。文献^[8]对微博媒体的特征和传播媒体的应对措施做了相关研究,认为微博网络作为中国互联网的新生事物,以其便捷的发布、广泛的互动传播性,已成为各大传统媒体争相抢占的另一信息发布平台。社会

收稿日期:2012-09-26

修回日期:2012-11-08

作者简介:涂师师(1990-),女,硕士研究生,主要从事网络传播研究。

*国家自然科学基金项目(批准号:61164020)、广西自然科学基金项目(批准号:2011GXNSFA018147)和广西空间信息与测绘重点实验室课题(批准号:1103108-24)资助。

**通讯作者。

网络分析方法近几年来发展迅速,在组织关系网络,虚拟网络等领域中得到了广泛应用。其中运用该方法来研究微博网络的中心性、微博社区网络结构等问题已成为主流的研究方向。

在传播学理论中,传播的类型有人内、人际、群体、组织、大众等多种形式,因此,基于互联网的信息交流,人际交往而形成的社会网络^[8]带来的传播效果也有多重形式。一方面,人们可以通过网络与自己圈子之内或者圈子之外的人交流,相互促进;另一方面,由于互联网的公开性和匿名性,出现了网络暴力的蔓延,病毒疫情的相互传播感染。如何使网络健康发展是长期以来的研究热点,从社会网络分析的角度来研究该问题,其中一个比较通行的办法就是找出社会网络核心节点,再通过控制核心节点,来实现对信息的获取和控制。社会网络分析中的“中间人分析法”是寻找该类核心节点的一种方法,它从中间人在不同的群体中扮演不同的角色的角度出发,得到不同性质的中间人^[1]。

本文收集了新浪微博中标签为“桂林理工大学”(桂林理工大学片区)的部分注册用户关系数据,构建了一个局域微博社交网络。应用社会网络分析 UCINET 软件^[5,9],从社会网络分析视角对微博网络的基本结构和社会关系进行了相关分析,得到一些分析结果,期望该结果能在整体微博网络的研究中起一定的示范作用。

1 数据获取及局域微博社交网络构建

1.1 数据获取

因为人的思维方式具有不均匀性,人们的眼界里总是更关注所谓的“巨人”,因此在微博网络中“名人”相对于“草根”来说受到更多的关注,在信息及其网络舆情的传播中,起到更为重要的领袖作用^[3]。为了更好地理解数据的实际意义,我们从身边出发,选择新浪微博中桂林理工大学片区,对标有桂林理工大学标签的注册用户进行了部分的抽取。该片区注册用户以桂林理工大学的学生和教职工为主,大多数为粉丝数低于 6000 的新浪认证用户和非认证的“草根”用户,并无实际的“名人”。所以,研究分析的起点选择从目前该片区中粉丝数即受关注度最高的“桂林理工大学学生会”(已受新浪官方认证)的微博开始,采取滚雪球的方式获取数据^[3]。先从桂林理工大学学生会的关注用户中搜索被关注数(即粉丝数),把超过 600 的记录下来,再将记录下来的用户采取同样的方法观察记录,从而得到一组用户的信息。该方法选取 25 人,分别是:1 桂林理工大学学生会,2 桂工之

声,3 林申镗-vincent,4 桂林理工大学招办,5 桂林理工大学社团联,6 桂林理工大学管理学院,7 达文_fat-fat-fat,8 lolo 罗罗罗罗,9OPQ 歌唔安官,10 段_炼,11 晓毅其实是个好孩子,12 心 kong,13 桂林理工大学地学院,14 楚楚令,15 兔子小姐鸣啦啦啦,16 桂林理工大学自行车旅游协会,17 桂林理工大学自管委,18 熙语 kimi_chan,19 霜刀红影之变节,20 桂理工孙巍,21 周国桥 v,22 桂工葫芦娃,23 随手拍桂工校园,24 桂工 idea 精英汇,25 哦李哦李。为了使数据更均匀,再以粉丝数低于 200 的用户为基准,用同样的方法得到另外 25 个个人的信息,分别是:26 桂工制造 MIG-studio,27 兰-翼,28 菠萝离我远远远远,29 茶廷儿 tracie,30 屁-狐狸,31 AngelaRed,32 同病相恋,33 尹以微事 Vi,34 飞扬一季,35 MyPrayer3434,36 你有点 2222,37 i-murmur,38 他们叫我动物,39 程辉得以,40 冰冷咖啡一号,41 桂工阿三,42 花见客,43 桂林理工雁山动感社团,44 唱唱歌其实也是极好的,45 桂林理工大学人力 10-1 班,46 灵魂游离的天使,47 凌日飞桂林理工大学,48 JJ 南部之星,49 double 佳哥,50 桂工赵小丽。

1.2 局域微博社交网络构建

把上述 50 个微博用户看成 50 个节点,它们之间的联系抽象为节点之间的连线,那么上述 50 个微博用户构成一个局域微博网络。由于新浪微博的用户群十分庞大(最新数据为两亿),本次抽取的用户虽然只是整个微博网络的一个子网络,隶属于整个网络之中,但是仍具有一定的代表性。

2 局域微博社交网络结构和社会关系分析

运用社会网络分析方法和 UCINET 软件对这个小型的社交网络进行相关的数据分析^[1,10]。

2.1 局域社交网络结构分析

2.1.1 网络规模

选取的案例最终共取得 50 个用户的信息,将 50 个用户之间关注与被关注的关系(表 1)用一个邻接矩阵的形式表达出来。在这个方阵中行表示关注者,列表示被关注者,1 表示关注这种关系存在,0 则表示不存在。由此得到相互关注情况的关联矩阵和关注结构网络(图 1),这样可以较为直观地了解该片区微博网络结构和整体概貌。

2.1.2 网络密度及成员间的距离

用 UCINET 软件对上述数据计算,得出该微博网络整体密度为 0.2002。而网络密度^[1]描述的是网络成员之间彼此互动的紧密程度,如果整体网是无向关系网,其中有 n 个行动者,那么其中包含的关系总

数在理论上的最大可能值是 $n(n-1)/2$ 。如果该网络中包含的实际关系数目为 m ，那么该网络的密度就是实际关系数除以理论上的最大关系数，即等于 $m/(n(n-1)/2) = 2m/(n(n-1))$ 。如果该整体网是有向关系网，并且其中有 n 个行动者，那么其中包含的关系总数在理论上的最大可能值是 $n(n-1)$ ，所以该网络的密度等于 $m/(n(n-1))$ ^[1]。网络密度数值 0.2002 并不是很大，说明网络整体不够紧密；同时，我们通程序计算得到所有节点平均距离为 1.768，意味着任意两个节点用户通过平均 1.768 个中介用户可以互相得到联系。此外，还得到以距离为度量的凝聚力指数（即该网络基于距离的紧凑程度）为 0.616。凝聚力指数取值在 0~1 之间，数值越大，表明整体网络越具有凝聚力。对照其他社会网络的凝聚力指数可以看出，该微博网络的凝聚力不高。这两点均可以从图 1 得到直观的印证。

表 1 微博关注情况(部分数据内容省略)

Tablet 1 The results of Micro-blog attention (part of the data content have been omitted)

| 用户 User | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | ... |
|---------|---|---|---|---|---|---|---|---|---|----|-----|
| 1 | — | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | ... |
| 2 | 1 | — | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | ... |
| 3 | 1 | 1 | — | 0 | 1 | 0 | 0 | 0 | 1 | 1 | ... |
| 4 | 1 | 1 | 0 | — | 1 | 1 | 0 | 0 | 0 | 0 | ... |
| 5 | 1 | 1 | 0 | 1 | — | 1 | 1 | 1 | 1 | 1 | ... |
| ... | | | | | | | | | | | ... |

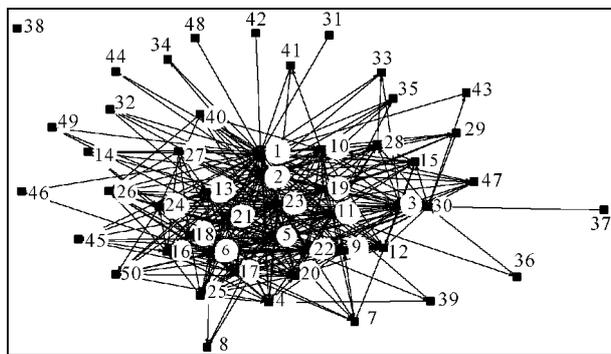


图 1 微博关注结构网络
Fig. 1 The network of Micro-blog's attention

2.1.3 派系分析

仅仅通过网络密度和成员间的距离只能大致了解该网络的结构概貌，对网络进行类似“帮派”、“小组”的研究可以得到该网络的小团体情况，从而更清楚地了解该社会网络内部结构，为此，对微博网络进行派系^[1]分析。经过多次试验，选取派系成员最小值为 8 进行派系分析，结果见图 2~图 4。

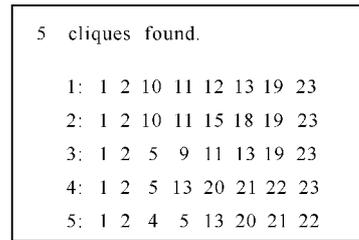


图 2 派系分析结果
Fig. 2 The results of clique analysis

| | 1 | 2 | 3 | 4 | 5 |
|----|-------|-------|-------|-------|-------|
| 1 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| 3 | 0.750 | 0.750 | 0.750 | 0.375 | 0.250 |
| 4 | 0.375 | 0.250 | 0.500 | 0.875 | 1.000 |
| 5 | 0.750 | 0.625 | 1.000 | 1.000 | 1.000 |
| 6 | 0.500 | 0.375 | 0.750 | 0.750 | 0.750 |
| 7 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 8 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| 9 | 0.750 | 0.625 | 1.000 | 0.625 | 0.500 |
| 10 | 1.000 | 1.000 | 0.750 | 0.500 | 0.375 |
| 11 | 1.000 | 1.000 | 1.000 | 0.625 | 0.500 |
| 12 | 1.000 | 0.750 | 0.750 | 0.500 | 0.375 |
| 13 | 1.000 | 0.750 | 1.000 | 1.000 | 1.000 |
| 14 | 0.250 | 0.250 | 0.125 | 0.125 | 0.125 |
| 15 | 0.750 | 1.000 | 0.625 | 0.375 | 0.250 |
| 16 | 0.500 | 0.250 | 0.375 | 0.375 | 0.250 |
| 17 | 0.500 | 0.375 | 0.625 | 0.750 | 0.625 |
| 18 | 0.750 | 1.000 | 0.625 | 0.750 | 0.625 |
| 19 | 1.000 | 1.000 | 1.000 | 0.625 | 0.500 |
| 20 | 0.500 | 0.500 | 0.625 | 1.000 | 1.000 |

图 3 各成员所属各派系的情况(部分数据内容省略)
Fig. 3 The results of every member in the various cliques (part of the data content have been omitted)

由图 2 可见，5 个派系的成员都较为集中，但是有很大一部分用户成员游离在这 5 个派系之外。由图 3 可以进一步了解所有用户的派系所属情况，可见派系的分布不是绝对的，用户所属各个派系的不同数值表明该用户与各派系的亲疏关系，如 4 号用户“桂林理工大学招办”属于第 1 个派系的指数为 0.375，而属于第 5 个派系的指数为 1，因此，4 号用户与第 5 个派系较为紧密。进一步分析第 5 个派系的成员(分别为：桂林理工大学学生会，桂工之声，桂林理工大学社团联，桂林理工大学地学院，桂林理工孙巍，周国桥 v，桂工葫芦娃)可以发现，前三者是由桂林理工大学内的几个官方学生组织，后 4 个是学校中行政机构或领导阶层的老师等。这 7 位均通过新浪加 V 认证，最明显共性就是他们是该微博片区中较为官方化、公开化和实名化群体的用户代表，发言主要围绕校园新闻和学校建设，具有一定权威性、指引性及控制力度，在该片区影响力也会比较大，因此他们作为一个派系

出现也是在情理之中。同时,由图 3 可见,37,8,31,32,36,37 等大部分用户的派系所属指数都为 0,表明他们不隶属于任何派系,即他们是该网络的较为孤立的成员,从图 1 也可以看出这些用户均处于关系图的边缘。以上数据还显示,在得到的 5 个派系中,重复现象很严重,说明这些派系彼此之间并不是特别独立,没有明显分裂格局的帮派现象。比如 1 号用户同时属于 5 个派系用户,说明该用户在该微博网络中人脉广,处于较为核心的低位。由此可见,该微博网络分布离散,用户之间联系不够紧密,呈现两极化的分布情况,这与复杂网络研究中呈现的微博网络具有“长尾现象”的结论一致,也与 2.1.2 节中提到的从该网络的相互交流整体来看比较稀疏的说法相一致。

2.2 局域微博社交网络社会关系分析

中间人大致可分为五种类型:协调人,顾问,守门人,代理人,联络人。用 UCINET 软件对该微博网络 50 位成员的中间人进行分析(分组情况采用成分分析的结果),结果如表 2 和图 5 所示。

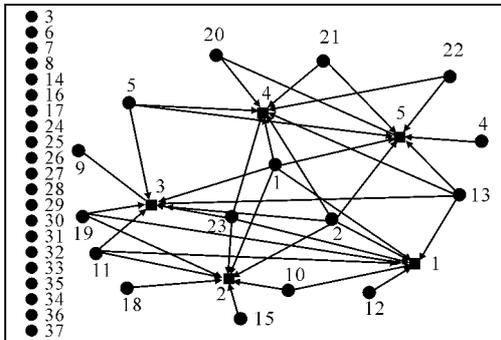


图 4 派系分析网络

Fig. 4 The network of clique analysis

表 2 各成员中间人分析结果(部分数据省略)

Table 2 The results of broker analysis (part of data content have been omitted)

| 用户 User | 协调人 Coordinator | 守门人 Gatekeeper | 代理人 Representative | 顾问 Consultant | 联络人 Liaison | 总数 Total |
|------------|--------------------|-------------------|-----------------------|------------------|----------------|-------------|
| 1 | 657 | 186 | 0 | 0 | 0 | 843 |
| 2 | 348 | 49 | 0 | 0 | 0 | 397 |
| 3 | 55 | 0 | 0 | 0 | 0 | 55 |
| 4 | 19 | 0 | 0 | 0 | 0 | 19 |
| 5 | 215 | 0 | 20 | 0 | 0 | 235 |
| 6 | 128 | 0 | 17 | 0 | 0 | 145 |
| 7 | 3 | 0 | 0 | 0 | 0 | 3 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 96 | 9 | 0 | 0 | 0 | 105 |
| 10 | 117 | 8 | 0 | 0 | 0 | 125 |
| 11 | 202 | 15 | 0 | 0 | 0 | 217 |
| 12 | 29 | 0 | 0 | 0 | 0 | 29 |
| 13 | 109 | 39 | 0 | 0 | 0 | 148 |
| 14 | 3 | 0 | 0 | 0 | 0 | 3 |
| 15 | 9 | 1 | 0 | 0 | 0 | 10 |

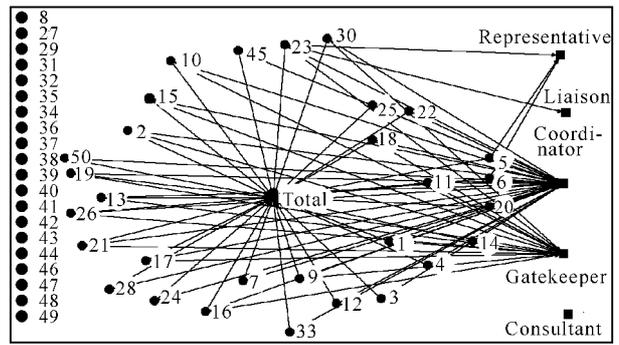


图 5 中间人分析结构网络

Fig. 5 The network of broker analysis

从表 2 和图 5 结果可知,该微博网络中中间人角色的担当者主要集中在协调人和守门人上,其中 1 号桂林理工大学学生会,2 号桂工之声,5 号桂林理工大学社团联,11 号晓毅其实是个好孩子,23 号随手拍桂工校园 5 位用户较多担任协调人的角色。协调人处理的是同一分组内最为普遍最为简单的一种信息传播方式,即控制组员相互之间的传播。因此我们可以判断上面 5 位在同一分组中的消息流通方面起到较为核心的作用。另一方面,所谓守门人,是分组之外的一个用户,通过目标分组内的一个人把信息传播到目标分组群体内的过程中,目标分组中的这个人即守门人。由表 2 和图 4 可知,桂林理工大学学生会,桂工之声,桂林理工大学地学院在各自的小组中起到守门人这样一个角色居多,即其他小组的成员要想获取小组之外或交际圈之外更远的信息,或者圈外信息想要更快更全面的覆盖某一个交际圈,主要依靠以上几个守门成员。虽然各个成员担任不同的中间人分量的情况不尽相同,但是,Total 统计显示该微博网络中中间人的整体情况中最有分量的中间人分别是桂林理工大学学生会,桂工之声,桂林理工大学社团联,晓毅其实是个好孩子,随手拍桂工校园。

3 实例验证

随机选取微博网络中一条具有较高传播率的微博(林申镔-vincent 在 2012 年 6 月 27 日发布的一条微博“要不要那么美,发现今天雨后的桂工那叫一个美...”)的真实传播数据,对以上的分析进行更直观的验证。通过近期较为热门的微博在线分析网站“WeiboReach 知微”(http://www.weiboreach.com/)分析知:该条微博被转发了 94 次,消息曝光量达 51160 人次,在传播过程中,关键的传播用户分别是桂林理工大学学生会,桂工之声,饭傻的丽酱饭,泓泓修身养息进行时,杨杨杨杨杨秘书,桂林理工大学

(下转第 81 页 Continue on page 81)

2.3.2 高效液相色谱法纯度检查

精密称取于 105℃ 干燥至恒重的对照品山茶苷 A 适量,加 80% 甲醇水溶液制成每 1ml 含 1mg 的溶液,色谱条件: Kromasil C₁₈ 色谱柱 (4.6mm × 250mm, 10 μm), 流动相为乙腈-0.2% 磷酸溶液 (21 : 79); 检测波长 330nm; 进样量 10 μl; 流速为 1ml/min, 柱温为室温。在该色谱条件下,对照品山茶苷 A 的 HPLC 色谱见图 2,用二级管阵列检测器 DAD 进行峰纯度检查为单一纯物质峰,用面积归一化法计算山茶苷 A 含量为 98.5%。改变流动相和波长,甲醇-0.2% 磷酸溶液 (46 : 54), 330nm 为检测波长和乙腈-0.2% 磷酸溶液 (21 : 79), 260nm 为检测波长分别检测,结果对照品为 1 个主峰,改变流动相分析未见有异常峰。

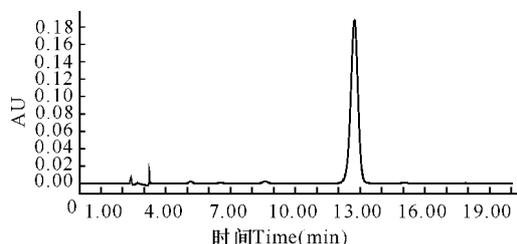


图 2 山茶苷 A 的 HPLC 色谱
Fig. 2 HPLC of Camellianin A

3 结束语

山茶苷 A 为亮叶杨桐的特征和有效成分,适合作为中药化学对照品。本实验提取、分离、纯化得到的山茶苷 A,符合中药化学对照品的相关要求,可以作为亮叶杨桐药材和含亮叶杨桐中成药质量控制,以及中药药效物质基础研究用的化学对照品。

参考文献:

- [1] 中国科学院中国植物志编辑委员会. 中国植物志: 第 50 卷[M]. 北京: 科学出版社, 1998: 28-29.
- [2] 陈美珍, 余杰, 余纲哲, 等. 野生石芽茶营养成分与药用成分的分析[J]. 天然产物研究与开发, 1996, 8(1): 85.
- [3] 袁萍, 王国亮, 钱少华, 等. 石崖茶活性成分分析[J]. 食品科学, 1999(8): 75.
- [4] 王英, 陈四宝, 倪洁, 等. 亮叶杨桐的化学成分研究[J]. 中国药科大学学报, 2003, 34(5): 407-409.
- [5] 王英, 叶文才, 殷志琦, 等. 亮叶杨桐的三萜皂苷类成分[J]. 药学学报, 2008, 43(5): 504-508.
- [6] 刘元, 李振麟, 张宁宁, 等. 亮叶杨桐的化学成分研究[J]. 中国民族民间医药杂志, 2012, 21(17): 62-63.
- [7] 刘元, 宋志钊, 李文琪. 亮叶杨桐指纹图谱研究[J]. 中国实验方剂学杂志, 2011, 17(6): 71-73.
- [8] 袁尔东, 肖仔君, 刘本国, 等. 亮叶杨桐叶总黄酮提取及抑菌活性的研究[J]. 现代食品科技, 2009, 25(3): 305-308.
- [9] 余杰, 陈美珍. 亮叶杨桐中类黄酮提取及其抗氧化、抑菌作用的研究[J]. 汕头大学学报: 自然科学版, 1997, 1(2): 52-58.
- [10] 袁尔东, 宁正祥, 刘本国. 亮叶杨桐叶中山茶苷 A 的分离纯化及其抗氧化性能的研究[J]. 食品工业科技, 2008(4): 212-214.
- [11] 刘本国, 战宇, 宁正祥. 亮叶杨桐叶黄酮类提取物的鉴定及其抗氧化活性研究[J]. 林产化学与工业, 2008, 28(1): 6-10.
- [12] 袁尔东, 王菊芳, 刘本国. 亮叶杨桐叶类黄酮的提取及其抗氧化活性研究[J]. 食品科学, 2009(14): 105-109.
- [13] 刘本国. 亮叶杨桐叶中类黄酮的提取、鉴定与修饰[D]. 广州: 华南理工大学, 2007.

(责任编辑: 邓大玉)

(上接第 78 页 Continue from page 78)

社团联等,其中有 3 个正好在本文选取的 50 个用户之内。综合来看,桂林理工大学学生会和桂工之声在该条微博中起到核心的作用,这与前面的中间人分析、派系分析等的结果基本吻合。类似的例子在新浪微博“桂林理工大学”片区中还有很多,在此不一一列举。

参考文献:

- [1] 刘军. 整体网分析讲义: UCINET 软件实用指南[M]. 上海: 上海人民出版社, 2009.
- [2] 周涛, 汪秉宏, 韩筱璞, 等. 社会网络分析及其在舆情和疫情防控中的应用[J]. 系统工程学报, 2010, 12: 742-753.
- [3] 平亮, 宗利永. 基于社会网络中心性分析的微博信息传播研究——以 Sina 微博为例[J]. 中国科技信息, 2010, 6: 163-164.

- [4] 尹书华. 基于复杂网络的微博用户关系网络特性研究[J]. 西南师范大学学报: 自然科学版, 2011, 12: 57-61.
- [5] Barabasi A L, Albert R, Jeong H, et al. Power-Law distribution of the world wide web[J]. Science, 2000, 287(5641): 2115.
- [6] Wasserman S, Faust K. Social network analysis: methods and applications[M]. Cambridge: Cambridge University Press, 1994.
- [7] Barabosi A. Statistical mechanics of complex networks[J]. Rev Mod Phys, 2002, 74(1): 47-97.
- [8] 黄艳. 微博的媒体特征以及传播媒体的应对[J]. 东南传播, 2011(1): 84-87.
- [9] 王陆. 典型的社会网络分析软件工具及分析方法[J]. 中国电化教育, 2009(4): 95-100.
- [10] 林聚任. 社会网络分析: 理论、方法与应用[M]. 北京: 北京师范大学出版社, 2009.

(责任编辑: 尹 闯)