

## 【百年馆庆征文】

基于学术 BBS 的信息交流研究<sup>\*</sup>

——以北大中文论坛的汉语言文学版为例

邱均平,熊尊妍(武汉大学中国科学评价研究中心,湖北 武汉 430072)

[摘要] 本文首先简述了信息交流的概念并强调其在非正式信息交流的现实意义,然后以北大中文论坛的汉语言文学版块为例,借助 Ucinet 社会网络分析软件分析了该论坛参与者的中间中心性以得出核心发帖群体;同时,利用 SPSS 的因子分析法来发现该论坛的核心主题。

[关键词] 信息交流; 社会网络分析; 因子分析; 学术 BBS

[中图分类号] G350 [文献标识码] A [文章编号] 1005-6610(2008)08-0003-06

## The Information Communication Research Based on Academic BBS

——Take the Chinese Language and Literature  
Section of Chinese Forum of Peking University as Example

Qiu Junping, Xiong Zunyan

(China Scientific Evaluation Research Center, Wuhan University, Wuhan 430072, China)

[Abstract] First, this paper introduces the concept of information communication and emphasizes its practical significance in the informal information communication. Then, the paper takes the Chinese language and literature section of Chinese forum of Peking university as an example, analyses the centralization of participants by using the soft of Ucinet social network analysis in order to show the core post group. Meanwhile, it discovers the core theme of the forum by using the factor analysis of SPSS.

[Keywords] Information communication; Social network analysis; Factor analysis; Academic BBS

写在前面的话:欣逢天津图书馆建馆一百周年之际,谨以此文向贵馆表示热烈的祝贺。对于贵馆百年来辛勤耕耘取得的巨大成绩表示由衷的敬佩。祝愿天津图书馆拥有更加辉煌灿烂的明天!

## 1 引言

美国社会学家门泽尔(H. Menzel)根据信息交流的载体类型不同,把科学交流分为“正式”和“非正式”过程。<sup>[1]</sup>与正式信息交流相比,非正式信息交流具有高度的选择性、直接性、针对性和实效性;并且反馈速度快,连续性强。非正式信息交流所具备的这些特点使其成为正式信息交流渠道有力的补充,进一步完善了信息交流理论,并能较好的适应现代信息技术条件下以网络为载体的信息交流环境。

\* 本文系国家自然科学基金项目“网上学术信息的分布与变化规律研究及其应用”(项目编号:70673071)研究成果之一。

网络社区正是互联网上非正式信息交流的一个重要组成部分。关于网络社区的定义,刘毓民认为<sup>[2]</sup>,是一群拥有个别兴趣、喜好、经验的人,或是学有专长的专业人士,通过各种形式组成一个社区,让参与该社区的会员彼此之间能借此进行沟通、交流和分享信息。与大多数普通学术论坛的较少人气相比,北京大学中文论坛的汉语言文学版发帖较为活跃,用户在学术问题探讨方面参与性较强,所以以其作为对网络非正式信息交流的研究对象,通过对北大中文论坛的汉语言文学版的社会网络分析来深入信息交流理论的研究。

## 2 研究方法

### 2.1 社会网络分析法

“社会网络分析”(Social Network Analysis,简称 SNA)的基本观点是将个人或组织之间的社会联系所构成的系统视为一个个“网络”,并认为整个社会就是由这些网络所构成的大系统。<sup>[3]</sup>从知识管理的角度来看,社会网络分析通过个体之间的关系映射来识别知识流程,弄清谁从何处寻求获取信息和知识,谁能与他们分享信息和知识。<sup>[4]</sup>

本文将从以下两个方面对社会网络中的信息交流进行考察:

#### (1)网络密度

网络密度(Network Density)<sup>[3]</sup>指群体成员间彼此的联系程度,亦即团队成员间人员互动连带的平均程度。也就是说,群体的网络密度值越高,成员的互动程度也越高,互动程度越高,产生的信息与资源交换自然也会比较多。网络密度可用讨论网络中实际拥有的连接线数 1 与最多可能存在的连线总数 n 之比来计算<sup>[5]</sup>:密度=2l/n(n-1)。

#### (2)中心性

中心性(Centralization)的指标一般有三种:点度中心性、中间中心性和接近中心性。中心性包括中心度和中心势两个概念,中心度刻画单个行动者在网络中所处的核心位置,而中心势刻画的则是一个网络所具有的中心趋势。<sup>[5]</sup>点度中心性测量的是一个点与其它点发展交往关系的能力;中间中心性刻画的是一个点与其它点控制网络中其他行动者之间交往的能力,它依赖于行动者与网络中的所有行动者之间的关系,而不仅仅是与邻点之间的直接关系。<sup>[6]</sup>故本文以点度中心性和中间中心性为代表。

### 2.2 因子分析法

因子分析法是从研究变量内部相关的依赖关系出发,把一些具有错综复杂关系的变量归结为少数几个综合因子的一种多变量统计分析方法。它的基本思想是将观测变量进行分类,将相关性较高,即联系比较紧密的分在同一类中,而不同类变量之间的相关性则较低,那么每一类变量实际上就代表了一个基本结构,即公共因子。对于所研究的问题就是试图用最少数目的不可测的所谓公共因子的线性函数与特殊因子之和来描述原来观测的每一分量。因子分析法主要用于:①减少分析变量个数;②通过对变量间相关关系探测,将原始变量进行分类,即将相关性高的变量分为一组,用共性因子代替该组变量。<sup>[6]</sup>本文的研究主要用到因子分析法的第二个应用,即通过论坛用户的发帖情况对用户进行分组,以便从一个宏观的角度来考察用户的信息交流内容,也可从一个侧面反映出汉语言文学领域的热门话题。

## 3 数据来源

对社会网络进行研究,目的在于分析网络环境下人们非正式信息交流的特点和内在规律,为进一步完善信息交流理论提供实证的支持。本文主要考察北大中文论坛的汉语言文学版块在 2007 年 11 月 14 日到 12 月 14 日之间的发帖情况。通过分析得到的数据,考察各节点(即发帖者)之间的关系,探讨该论坛交流的活跃程度如何?论坛的话题主要有哪些成员引导?这些成员在论坛中所占的比例是多少?其他成员在论坛中处于什么样的地位?论坛成员的讨论话题集中在哪些方面?是否可以通过他们的发帖情况用统计软件直接进行话题分类等,从而揭示学术性 BBS 信息交流的特点和规律。

笔者通过对北大中文论坛的汉语言文学版块在 2007 年 11 月 14 日到 12 月 14 日之间的发帖情况用 Excel 表格进行了统计,如图 1 所示。列代表发帖者,行代表回帖者,即对任意一对注册用户 X 和 Y,

在互动关系矩阵中包含两个涉及到他们关系的数据单元格(X, Y)和(Y, X),分别代表 X 回复 Y 的帖子数和 Y 回复 X 的帖子数。其中,某参与者发给其他人的帖子总数,叫做这个参与者的点出度;而他获得的其他人的回帖总数,叫做这个参与者的点入度。本文剔除了一些无实际内容的回帖,如“顶”、“飘过”等。同时,回帖数为零的主题贴也不计算入内,因为这样的帖子不能反映论坛成员之间的信息交流情况。最后得出这七天在汉语言文学板块的发帖有 544 份,发帖者 235 人(包括回帖在这一个月之内,而主贴在这一个月外的作者),而 235 位发帖者就是本次调查的研究对象和标本。

	A	B	C	D	E	F	G	H
1		连登岗	连登岗	燕赵学子	尘封记忆	tungchan	冷风商洛	长相守
2	陈立民	1	1	0	0	0	0	0
3	连登岗	1	1	0	0	0	0	0
4	tungchan	0	0	18	0	3	0	0
5	燕赵学子	0	0	8	0	0	0	0
6	徐欣路	0	0	12	0	0	0	0
7	天月明	0	0	1	0	0	0	0
8	啸啸生	0	0	0	1	0	0	0
9	尘封记忆	0	0	0	1	0	0	0
10	旷世风华	0	0	0	1	0	0	0
11	pku	0	0	0	1		0	0
12	青藤书屋	0	0	0	1	0	0	0
13	余少平	0	0	0	2	0	0	0
14	王某	0	0	0	1	0	0	0

图 1 发帖数据统计(Excel 表的部分截图)

## 4 数据分析

### 4.1 网络密度分析

在一个讨论区中,密度值大的网络密度高,表明个体接触到的群体规模大,亦即个体所接收的回帖内容范围广,个体从中获取的信息量大。反之,如果仅有少数个体之间有联系,则该网络密度低,个体接收回帖的范围小,从中获取的信息量也小。<sup>[6]</sup>根据密度测度公式可以得出汉语言文学版块在这一个月中的网络密度为 0.0099。统计结果较现实的交往而言,数值是很低的,其原因也是多方面的。首要的原因是,在学术论坛中,可供讨论的话题多,样本数量也较多,因此人们可选择的讨论对象范围广,选择易分散。其次,与普通论坛所不同的是,学术论坛里讨论的话题通常带有一定的专业性,且不能引起普遍关注,所以在这样的信息交流过程中,受制于人们的受教育程度和知识领域的有限性,人们通常会选择自己擅长或感兴趣的领域进行有限深入,这也是导致许多学术论坛的信息交流不畅通、不活跃的主要原因。

### 4.2 中心性分析

把 Excel 表中的成员发帖信息导入到 Ucinet 中的 Spreadsheet 中,保存为 Ucinet 中的数据格式(简称为 Ucinet 表)。接下来进行中心性分析,在 Ucinet for Windows 窗口中打开菜单“Network→Centrality→Degree”,导入刚才保存的 Ucinet 表,由于该表中的社群矩阵属非对称矩阵,所以在“Treat data as symmetric”的选项中选择“No”;在“include diagonal values?”中选“No”,排除对角线上的数值(即发帖者回复自己的帖子数),再运行对该表的分析可得出点度中心度的分析结果。如表 1 所示,其中点出度最高的是 tungchan,点入度最高的是 ritter。

继续从“Ucinet for Windows”窗口中打开菜单“Network→Centrality→Freeman Betweenness→Node Betweenness”,导入 Ucinet 表,运行对该表的分析可得到中间中心度的分析结果。中间中心度最高的用户是 ritter,而在中间中心度最高的 13 名成员之后,其他成员的中间中心度均为零。

表 1 中心度排名

点出度(度数)	点入度(度数)	中间中心度(度数)
1 tungchan(30)	1 ritter(50)	1 ritter(271.0)
2 徐欣路(27)	2 沈炯(46)	2 八爪鱼多多(160.5)
3 薄守生(15)	3 燕赵学子(36)	3 zz_jjll(57.0)
4 刘晓林(12)	4 中言(28)	4 tungchan(42.5)
5 豫仙(10)	5 尘封记忆(22)	5 燕赵学子(31.0)
6 ZL-JJ520(10)	6 bnnnnnnnnnn(22)	6 中言(20.0)
7 光不度(9)	7 长相守(20)	7 南园颯(12.0)
8 zz_jjll(8)	8 随意居(19)	8 AbyssThinIce(10.0)
9 伊犁老狼(8)	9 八爪鱼多多(14)	9 余少平(9.5)
10 yima(7)	10 难说会道(12)	10 长相守(6.5)
11 wlren(7)	11 jodyjia(11)	11 柳漫(6.0)
12 shenj(7)	12 prometheus(10)	12 海中轩(4.0)
13 清风出袖(6)	13 love884(9)	13 andy2007(3.0)
14 无为(6)	14 tungchan(8)	
15 AbyssThinIce(6)	15 柳漫(8)	
16 ritter(6)	16 brooke(8)	
17 phagspa(6)	17 闪闪亮(8)	
18 汉文字爱好者(6)	18 shuishui81(7)	
19 长相守(6)	19 zz_jjll(7)	

如表 1 所示,ritter、tungchan、zz\_jjll 和长相守这四名成员在点出度、点入度和中间中心度的排名中均靠前。可见他们在北大中文论坛的汉语言文学版块中处于领导者的地位,可称为“精英型”成员。<sup>[7]</sup>他们不仅通过回帖积极参与论坛的各种讨论,所发的主题贴也能得到论坛成员的积极回复。在网络论坛中,成员获得中间中心度的必要条件是要在论坛中发布帖子并获得其他学习者的回复。它反映出成员对整个汉语言文学版的群体信息交流具有较强的影响力。因此,他们对论坛中的信息交流的贡献也是最大的,每个论坛都需要有这样的领导者才能使论坛产生一定的凝聚力,从而避免论坛的吸引力不足导致的结构松散。

沈炯、燕赵学子、中言、尘封记忆、bnnnnnnnnnn 等点入度较高的成员发帖的内容吸引了众多成员,但他们的影响力不及前者,主要是因为他们对其他讨论参与得很少,属于“实力型”成员。而徐欣路、薄守生、刘晓林等点出度较高的成员积极关注许多其他成员的帖子,但自己的发帖吸引力不够或者从未发过主题贴,影响力也不及“精英型”成员,属于“活跃型”成员。而论坛的其他大部分成员都较少参与论坛的讨论交流。

根据 Ucinet 的分析结果,就整个网络来说,其出度中心势指数为 0.730%,入度中心势指数为 1.135%,而中间中心势指数为 0.490%。中间中心势远远小于出度中心势和入度中心势的值。这说明整个网络的中心势程度很低,虽然中心势过高会形成少数人控制论坛讨论话题的局势,不利于人们在论坛中参与信息交流的过程。但中心势太低,会使得论坛缺乏凝聚力,论坛的吸引力不足,这也是许多 BBS 人气不旺的重要原因。

#### 4.3 因子分析

群体中常常会产生小群体的状况,它与成员的性格,文化背景和爱好等诸多因素有关,而学术论坛中的小群体的出现是与成员的研究兴趣有关的。本文采用统计学中的因子分析法来分析论坛上发帖者之间的相关性,并在此基础上找出论坛中产生的不同群体。

在一个论坛中,那些回复同一主题帖数越多的用户,我们可以认为他们的相关性越高;相反,所回复

的同一主题贴数越少的用户,相关性越小。例如,若现有 A、B、C、D 四个不同的主题帖,成员 x 和 y 都对 A 和 B 进行了回复,而 z 对 C 和 D 进行了回复,我们可以从中推测出 x 和 y 均对 A 和 B 的内容比较感兴趣,而 z 则对 C 和 D 的内容比较感兴趣,因而可以进一步推测出 x 和 y 的相关性高于 x 和 z 的相关性。因此可以通过论坛成员的发帖主题对其感兴趣的话题进行引分类。

由于论坛发帖数据庞大,其中有许多主题帖仅有少量的回复,不足以反映成员之间的相关关系,所以为了保证本次数据分析结果的有效性,笔者剔除了回帖数少于 6 的发帖信息和经过此次调整后发帖数变成零的用户数据。最后得出参与分析的主题帖发帖者有 25 名,回帖者有 168 名。

在 SPSS 软件中选择方差最大法(Varimax)因子旋转方法,它使每个因子上具有最高荷载的变量数最小,因此可以简化对因子的解释。在“Option”对话框中,在“Suppress absolute values less than”后填 0.50,突出荷载较大的变量,以便得出相关性分析的结论。

旋转后的因子负荷矩阵结果如表 2 所示。

表 2 因子负荷矩阵图

Rotated Component Matrix(a)												
	Component											
	1	2	3	4	5	6	7	8	9	10	11	12
随意居	.867											
jodyjia	.832											
燕赵学子	.707	.568										
中言		.880										
shuishui81		.853										
八爪鱼多多			.882									
闪闪亮			.878									
大内密探零零幺				.782								
tungchan				.774								
love8848				.520								
尘封记忆					.775							
长相守					.637							
珠源汉子						.763						
axuan						.604						
prometheus							.767					
brooke							.765					
ritter								.755				
沈炯								-.508				
柳漫								.507				
难说会道												
连登岗									.836			
hume										.763		
bnnnnnnnnn											.611	
zz_jjll											.812	
水镇过客												.949

在一个论坛中,那些回复同一主题帖数越多的用户,我们可以认为他们的相关性越高;相反,所回复同一主题贴数越少的用户,相关性越小。汉语言文学的研究体系可分为语言和文学两个方向,语言研究包括语音、语法和词汇三个部分,而文学研究包括古代文学、近代文学、现代文学、当代文学等。因此笔者也按照汉语言文学研究体系来对论坛讨论者的讨论话题进行分类,分类结果如下:

- 语音:燕赵学子、长相守、hume、bnnnnnnnnnn、shuishui81、闪闪亮、love8848;
- 语法:燕赵学子、tungchan、八爪鱼多多、zz\_jjll、ritter、shuishui81、中言、水镇过客、love8848;
- 词汇:连登岗、八爪鱼多多、zz\_jjll、ritter、jodyjia、大内密探零零幺、柳漫、难说会道;
- 由于涉及文学内容的讨论较少,所以本文将古代文学和现代文学等全部归为文学类:连登岗和 axuan。
- 除此之外,论坛的讨论中还涉及少数民族语言、词频统计等话题,本文将这部分讨论单独归为一类,包括:长相守、尘封记忆、沈炯、prometheus、brooke、珠源汉子、随意居。

结果显示,SPSS 把论坛讨论者分为 12 个组,其讨论的主题如下:第一、三、四组为语法,第二组为语法和语音,第五组为少数民族语言,第六、九组为古代文学,第八组为词汇和语法,第十组为语音,第十一组为词频统计,第十二组为专题探索(偏语法)。以 SPSS 的分组作为标准来对笔者的分类进行比较后发现,笔者的分组与其大部分都是吻合的,仅有少数几个分组有误差,比例不超过 20%。

## 5 总结与展望

本文通过社会网络分析法实现了对论坛信息交流的研究。通过社会网络分析法中的网络密度和中心势等方法分析了论坛交流的活跃度;用中心度的统计总结出论坛的活跃成员;而因子分析法则归纳了论坛所讨论的主要话题。本文的新颖之处在于用因子分析法对论坛的讨论主题进行分类,为 BBS 上的信息交流实证研究提供了一种可行的研究方法,但也存在不少问题有待于进一步研究。

(1)由于论坛用户访问和话题讨论的随机性特点,对 BBS 的实证研究必须基于大量的数据统计才能得出较为有效的结果。本文在进行数据收集时完全是手工操作,效率比较低,而且容易出错,获取的数据量也十分有限。可以开发一个程序,以便于把 BBS 访问者之间的交流关系直接转化为反映交流关系的社群矩阵。

(2)在本文的研究中可以发现,BBS 上的“精英型”成员对 BBS 论坛的发展起着至关重要的作用,他们是论坛社会网络中的关键节点,因此,如何突出论坛“精英型”成员的积极作用也可以作为信息交流理论的研究内容。

(3)我们通过前面的研究发现 BBS 信息序列的传递依然遵循它的传递规律,依然可以通过有效的引导达到相应的舆论效果。社会网络分析法中的中心性研究可以与情报学原理之一的小世界原理相结合进行研究,从而进一步深化对信息交流的理论研究。

(4)信息检索和文献计量是社会网络分析法目前在图书情报学的主要应用领域,社会网络分析可能在 LIS 的三个领域得到发展:社会知识网络嵌入、社会网络的互联性和信息传播管制以及网络关联桥的研究。在情报学领域,信息学家对出版物研究、引证和同被引网络、合作关系及其它社会互作用形式通过互联网都构成无标度的社会网络,而在所有这些研究中社会网络分析具有广泛的应用前景。

### [参考文献]

- [1]黄水清,王超.网络环境下非正式信息交流的概念、类别与特点[J].图书馆杂志,2004(6):8-11.
- [2]郭玉锦,王欢.网络社会学[M].北京:中国人民大学出版社,2005:234.
- [3]李培林,覃方明.社会学:理论与经验[M].北京:社会科学文献出版社,2005:102-115.
- [4]褚建勋.基于复杂网络的知识传播动力学研究[D].合肥:中国科学技术大学,2006.
- [5]刘军.社会网络分析导论[M].北京:社会科学文献出版社,2004:116-131.
- [6]胡勇,王陆.网络协作学习中的社会网络分析个案研究[J].开放教育研究,2006(5):56-60.
- [7]彭小川,毛晓丹.BBS 群体特征的社会网络分析[J].青年研究,2004(4):39-44.

[作者简介]邱均平(1947—),男,教授,博士生导师;熊尊妍(1985—),女,硕士研究生。