

基于社会网络分析方法和数据挖掘方法的 网络论坛定量分析



重庆大学硕士学位论文

学生姓名：庄经纬

指导教师：周 明 教授

专 业：管理科学与工程

学科门类：管理学

重庆大学经济与工商管理学院

二〇一〇年五月

The quantitative analysis of BBS based on Social Network Analysis and Data Mining



A Thesis Submitted to Chongqing University
in Partial Fulfillment of the Requirement for
the Degree of Master of Management

By

Zhuang Jingwei

Supervised by Prof. Zhou Ming

Major: Management Science and Engineering

College of Economics and Business Administration Chongqing
University, Chongqing, China

May 2010

摘 要

近年来网络飞速发展，全国网民数量大幅度上升，网络论坛也成了网民表达自己观点，看法的重要场所。网络舆论这一新的舆论方式通过网络论坛等媒体走到前台。网络的匿名性，开放性，跨时空性，低受控性等特性，使得网络媒体和传统媒体有很大差异，管理和控制网络舆论显得尤为重要。但是网络舆论的传播规律深埋在纷繁复杂的网络中，很难掌握，因此对网络舆论的管理非常困难。

数据挖掘和社会网络分析等数据分析技术的兴起，给我们带来了机会。本文分别采用数据挖掘和社会网络分析对某大学网络论坛某段时间的发帖和回帖情况进行分析。一方面抽出两个月回帖量最大的 50 个帖子，找出各个帖子的生命周期相关特征并用凝聚层次聚类法给这 50 个帖子聚类，然后用关联规则分析寻找帖子类别和帖子其他特性之间的关系，从而寻找帖子的生命周期规律；另一方面抽取某天某板块的所有帖子，为回帖人和发帖人之间，回帖人和帖子之间分别建立社会网络分析模型，分析发(回)帖人在网络中所处的位置和帖子在网络中所处的位置，最后归纳这些特征得出网络舆论传播相关的规律，并通过总结这些规律，拟定出一个网络论坛管理方案。

这一套结合了数据挖掘和社会网络分析方法的定量分析模型能够描述某些网络论坛的舆论传播规律。

关键词：网络舆论，网络论坛，数据挖掘，社会网络分析

ABSTRACT

With the rapid development of the Internet, the amount of the netizen has increased drastically, and the BBS became to the place where netizen expressed their opinions. The web opinion came into our sight though the BBS. with the characteristics of anonymity, open, space-time crossing and low controlled, the Internet had large difference compared with traditional media, and the management of web opinion came out to be quite important. As the discipline of the spread of web opinions was deeply hide in the wide internet, and hard to grasp, the management of the web opinion was very difficulty.

The rise of the method of Data mining and Social Network Analysis brought an opportunity to us. This paper adopted Data mining and Social Network Analysis anglicizing the situation of post and reply of a BBS in a university during some time separately. In one aspect, we extracted the top 50th posts in two month, and made a model computing the feature of lifecycle of the posts, then made cluster with the method of Hierarchical Agglomerative Clustering, after that we find some rule between the cluster and other feature of the posts with method of Association analysis. In the other aspect, we extracted all posts in one day, made one mode between the users another mode between users and posts, and analysis the position of the users in the network, and finally concluded the rules about the spread web opinions. We gave some management comments about BBS according to the rules.

It is concluded that the method comprised data mining and social network analysis could depict some feature of the spread of web opinions

Keywords: web opinions, BBS, Data Mining, Social Network Analysis

目 录

中文摘要.....	I
英文摘要.....	II
1 绪 论.....	1
1.1 研究背景.....	1
1.2 研究的目的与意义.....	4
1.3 国内外研究现状.....	4
1.3.1 国外的研究现状.....	4
1.3.2 国内的研究现状.....	5
1.3.3 目前研究的不足之处.....	8
1.4 研究的主要内容和思路.....	8
2 社会网络分析方法和数据挖掘方法介绍.....	11
2.1 社会网络分析方法.....	11
2.1.1 用户——用户网络模型.....	11
2.1.2 用户——事件网络模型.....	13
2.2 数据挖掘方法.....	14
2.2.1 关联规则分析.....	14
2.2.2 聚类分析.....	17
3 用数据挖掘方法分析帖子的特性.....	19
3.1 数据来源.....	19
3.2 分析流程.....	19
3.3 构建特征模型和聚类分析.....	20
3.3.1 数据提取.....	20
3.3.2 构建特征模型.....	21
3.3.4 聚类分析.....	23
3.4 帖子属性和关联规则分析.....	25
3.4.1 帖子属性.....	25
3.4.2 关联规则分析.....	26
3.5 规律归纳.....	28
4 用社会网络分析方法分析用户和帖子的特性.....	29
4.1 数据来源.....	29
4.2 分析流程.....	29

4.3 用户——用户网络模型分析.....	30
4.3.1 数据提取和数据预处理.....	30
4.3.2 用户威望度和中心度分析.....	32
4.4 用户——事件网络模型分析.....	35
4.4.1 数据提取和数据预处理.....	35
4.4.2 事件中心度分析.....	36
4.4.3 事件之间的联系.....	38
4.5 归纳网络用户在网络舆论传播中的地位和规律.....	40
5 结果总结和管理建议.....	42
5.1 总结分析的结果.....	42
5.2 管理建议.....	43
6 结论与展望.....	45
致 谢.....	46
参考文献.....	47
附 录： A. 作者在攻读学位期间发表的论文目录.....	50
B. 部分图表.....	50

1 绪 论

1.1 研究背景

自 1997 年以来,网络在我国获得了飞速的发展。网络受到越来越多的人的喜爱,网民规模不断增加。据 CNNIC (中国互联网络信息中心) 发布《第 25 次中国互联网络发展状况统计报告》显示,截至 2009 年 12 月,我国网民规模已达 3.84 亿,互联网普及率进一步提升,达到 28.9%^[1]。2010 年 3 月 5 日,腾讯 QQ 同时在线人数突破 1 亿。这些无不说明网络已经深入了我们的生活,影响着我们的生活。

伴随着网络的发展,网络成为了继报纸、广播、电视之后的“第四媒体”。互联网的飞速发展使网络新闻的重要性突显。网上论坛、媒体言论和新闻跟帖内容丰富、异彩纷呈,无不体现出网络舆论举足轻重的作用。互联网走进了人们的工作和生活。网络舆论成为现代舆论的一个重要组成部分,并且它所占的比重越来越重大,无论是“华南虎事件”还是“躲猫猫事件”都是首先在互联网上迅速传播,随后才引起其它媒体的关注。网络舆论对人类的生存方式和社会状态产生了全方位的影响,从国家领导人到普通百姓,从国家大事到明星轶闻,从社会热点到个人琐事,网络不时掀起一阵阵的舆论热潮。可见网络舆论对人们生活、国家政治经济都存在着巨大的影响。

“家乐福事件”充分证明了网络舆论的巨大影响力。2008 年 4 月,出于对奥运圣火海外传递遭遇的抗议,中国网民自发掀起了一场声势浩大的“抵制家乐福”风潮。整个事件过程如下:法国站火炬传递结束后,国内所有形式的媒体都进行了广泛、持续的报道;网民在各大论坛及 QQ 群上积极讨论如何向法国表示抗议;4 月 10 日,“水婴”(网民)在“猫扑”社区上发布“抵制法国货,从家乐福开始”一帖,受到网民积极跟帖相应;不到一周时间,通过 QQ 群,MSN 和 BBS 论坛的持续传播,“抵制家乐福”这一号召在网络上迅速壮大,并发展成有组织的抵抗风潮。

舆论需要得到正确的引导和控制,一个国家如果不能有效地控制舆论那将是非常危险的,现今最大众化的舆论传播工具电视和报纸都得到了国家的严格控制,像中央电视台这样的部门绝对是国家重点控制的部门,因为这关系到政权的稳定和社会的和谐。就互联网现在这发展速度,它明天很有可能取代电视成为最大众化的舆论传播工具,但是国家对网络舆论的引导和控制明显做得没有对电视的控制那么好。导致网络舆论不容易控制的原因有很多,如网络传播的开放性、互动性、匿名性等特征造成对它的监控面临许多困难,对网络舆论传播规律的不熟悉让我们找不到重点。

网络应用中的交流沟通的手段主要有即时通信、电子邮件、博客和论坛/BBS

等四类。而在这四类沟通交流手段中，即时通信和电子邮件的沟通主要是私人之间信息的沟通，对网络舆论的影响不是很大。博客与网络论坛相比，相对 BBS 那种口无遮拦、随心所欲、良莠不齐的情绪化发言来，博客制作的日志更加审慎、仔细和周详，其单个文本的丰富性、讨论脉络的清晰度、论题的拓展空间都超过了 BBS 的网友发言帖子^[2]。可见，网络论坛对网络舆论影响较大，且最不规范，情绪化严重等。所以，本文主要研究网络论坛中网络舆论传播的规律和管理。

网络论坛（BBS）是英文 Bulletin Board System（简报版系统）的缩写，即电子公告版，是上世纪 70 年代末、80 年代初互联网业余爱好者发明的一种网络交往形式。它能为用户提供各种信息服务，并形成相对固定的网络公共环境。成熟的 BBS 涵盖了寄存函件、读取通告、参与讨论和交流信息等多种功能，其中最重要且最深入人心的功能仍然是作为用户发表观点的论坛^[2]。目前，全国主要门户网站新浪、搜狐、网易等都开设了自身的网络论坛，中央媒体网站如人民网、新华网、央视国际等，以及地方媒体网站如东方网、千龙网、红网等也都开通了自己的 BBS 论坛，另外还有许多大型商业 BBS 社群，如天涯社区，西陆社区、凯迪社区、西祠胡同、等。而目前网络论坛中，校园网络占据着非常重要的地位，主要是由于在现有的网民中，学生群体所占的比重最大，具《第 25 次中国互联网络发展状况统计报告》统计显示 2009 年学生网民占的比率高达 28.8% 在所有职业中排在第一位，比排在第二位的企业/公司一般职员 15% 高出了 13.8 个百分点。而各个学校为了能了解学生的动向而建立学校网络论坛无疑成为各类网络论坛中非常具有代表性的网络论坛。如教育科研网的水木清华站、木棉站等。

网络论坛给网民们脱去了身份的外衣，摘掉了旁人的目光，放下种种顾虑，提供一个自由言论的平台。使得网络论坛凝聚了民意的力量，关注民生、关注国家的一系列事情，通过一个虚拟的网络平台来影响着政府的公共决策。如孙志刚案最终促成了一部实行了 21 年的国家法规的废止；刘涌案使得最高人民法院再审，让其最终走上了黄泉；宝马撞人案网友的群起激愤终于“逼”得黑龙江省委副书记表态“依法办事、秉公处理”等^[3]。

在网络论坛舆论为国家公共决策等方面提供支持，显示出了不可忽视的积极力量的同时。网络论坛舆论也存在诸如色情化、情绪型舆论弥漫、网络道德行为失范、流言大量存在、群体的盲从与冲动、群体极化、网络侵权、违规问题等问题。所以，必须对网络论坛舆论进行有效的引导和管理，为网络论坛舆论的发展提供健康的环境。目前我国网络舆论管理的方法主要有网络立法管理、行政手段的监督、技术手段的控制和行业自律的约束，但是由于网络论坛舆论具有传播速度快、信息传播者匿名等特征和我国网络舆论管理机制的不完善，导致对网络论坛舆论的管理存在一定的不足。

传统舆论管理方法在网络论坛舆论管理的缺陷要求有一种新的方法能弥补这一不足。这就需要我们加强对网络舆论传播规律的理解。然后针对网络舆论在网络论坛中的传播扩散规律提出一种更加有效的管理制度，更好的控制和引导网络舆论。网络舆论在复杂的网络论坛中传播，数据量非常大规律也不明显，分析起来非常的困难。此时社会网络分析方法和数据挖掘方法的成熟为此提供了条件。

社会网络分析方法是在人类学、心理学、社会学、经验研究、数学及统计学领域中发展起来的，经历了 70 多年的历史，被应用于物理学、生命科学和社会科学等领域的研究。社会网络分析方法在经历了起源，诞生，黑暗时期，哈佛的复兴和成熟一系列的发展过程后，已经完全成熟。目前拥有了一个国际性的组织：社会网络分析国际网络，并设立了刊物《社会网络》。现在社会网络分析正式进入社会学量化研究的行列，成为社会科学研究的一种新的范式。社会网络分析主要是针对一个复杂的社会网络，对复杂社会网络中成员的行为，地位及成员之间的关系等方面进行分析，从而探讨复杂网络中成员之间的一些性质，为社会学，管理学等学科的研究提供方法论支持。而网络舆论是由很多网民之间通过交流，进行信息的传递形成的。在此，将网络舆论中的成员视为节点，成员之间的信息传递视为联系，这将构成一个庞大的复杂网络，为社会网络分析利用到网络舆论管理提供了基础。数据挖掘是通过自动或半自动的工具对大量的数据进行探索和分析的过程，其目的是发现其中有意义的模式和规律^[4]。数据挖掘现在已经相当成熟，其中聚类分析，关联规则分析等方法已经形成完备的体系，大量的运用在各行各业的数据分析中。

数据挖掘方法，简单的讲就是从大量的数据中挖掘或抽取知识。该方法能够从大量的，不完全的，有噪音的，模糊的，随机的实际应用数据中提取人们感兴趣的知识，这些知识是隐讳的，事先未知的，潜在有用的信息。数据挖掘有这些特性正好适合用来分析网络中纷繁复杂的数据，寻找网络舆论的传播规律。常用的数据挖掘方法有：关联分析，序列模式分析，分类分析，聚类分析，和孤立点分析等，我们可以从其中选取一种或几种来分析对象数据。

总之，网络的发展让网络舆论成为国家和社会非常关注的问题，而对于网络舆论的管理则需要我们清楚网络舆论的传播特性，网络舆论产生在纷繁复杂的网络大环境中，这里数据量大并且种类繁多，要从其中找出网络舆论的传播特性非常困难。但是社会网络分析方法和数据挖掘方法的兴起，给我们研究分析网络舆论的传播规律提供了契机。本文将以 BBS 这一网络舆论传播的重要平台作为研究对象，采用社会网络分析和数据挖掘方法分析网络舆论传播的特性，找出规律，并根据规律提出一些针对 BBS 的管理方案。

1.2 研究的目的与意义

本研究的主要目的是：通过引入社会网络分析方法和采用数据挖掘技术建立分析模型，收集处理网络论坛中各用户之间相互交流的情况，分析网络论坛中各用户之间的联系以及这种联系对网络舆论传播的影响，寻找网络用户在网络舆论传播过程中的作用和地位；分析网络论坛中帖子的生命周期和活跃周期，分析帖子本身的扩散规律。通过对这些规律的分析，捕捉网络舆论的传播规律，并针对这些规律提出有关网络论坛的管理的对策建议，以帮助提高网络论坛管理的针对性和有效性，为加强对网络舆论的控制提供分析方法。

其主要意义有：

首先，对网络舆论传播的分析有利于掌握网络舆论的内在传播规律。国内对舆论传播的研究很多都是在对报纸，广播，电视等传播媒体中的传播，而对网络舆论传播的研究还处于起步阶段，因此本研究对于网络舆论传播规律的研究对于完善舆论传播规律具有重要的理论意义。

其次，将定量方法与网络舆论传播结合起来给舆论传播研究提供了新的研究思路。传统的舆论传播研究限于人与人之间相互联系的数据很难获得，大多是定性研究，少有的定量研究方法都是通过计算机模拟人际之间的交流或采用调查问卷在一个很小的范围内做定量研究，然而网络论坛给我们提供了真实准确并且数量巨大的人际之间的交流情况，给定量研究舆论传播提供了一个新的平台。

最后，根据研究结果提出的网络论坛管理方法可以加强对网络舆论传播的控制。网络舆论的影响力与日剧增，然而网络舆论又不同于传统舆论，对网络舆论的管理需要采用新的办法，本研究探讨也是对这些新方法的一种尝试，给网络舆论的管理提供一种可以借鉴的方案。

1.3 国内外研究现状

1.3.1 国外的研究现状

对舆论的关注，国外将舆论的研究扩展到大众传播、公共关系等领域，并借鉴了社会学、心理学、政治学、传播学、新闻学、信息论、控制论、系统论等学科的理论和方法，并成立了不少舆论研究机构。

政治学和行政学是国外研究舆论学的传统之一。1762年，卢梭在政治学名著《社会契约论》中，首次将“public”和“opinion”结合起来，创造了公共舆论这个词。1887年，威尔逊在《行政学研究》中，阐述了行政管理和公共舆论的关系。1899年，法国社会学家埃尔·德·塔尔德写出了社会舆论研究专著《社会舆论与公众》^[5]。1922年，沃尔特·李普曼在其著作《舆论学》中写道：舆论基本上就是对一些事实从道义上加以解释和经过整理的一种看法^[6]。1926年，拉斯韦尔在《传播在社会

中的结构与功能》中传输了“5W”理论，成为传播学的奠基人之一^[7]。1981年，希尔德勃兰特在其出版的《权力与公民：1897-1921年对外事务中的舆论行政管理》中明确提出了“舆论行政管理”一词。同时，西方现有的一些政治学教科书中，也把舆论作为重要的研究内容。

关于言论自由方面，美国耶鲁大学法学教授爱默森（ThomaI.Emerson）的论述比较全面。他说，“在现代社会里，言论自由系统是个非常复杂的结构。它的核心是担保社会各个分子的一群权利。”这群权利包括：1、人人对任何事务皆有发表意见的权利，2、保持缄默之权利，3、听取他人之观点与阐述事实的权利以及基于此而演绎和调查的权利，4、为了使个人意见表达更为广泛有效而进行集会、结社、出版的权利。对言论自由的功能他提出四个方面：个人自我成就之达成，真理之追求，决策之参与和维护社会变迁的稳定^[8]。

很多学者借用模型来表达描述舆论的传播：Paul A 用一种带有阈值判断的图理论来描述舆论和疾病的传播规律^[9]。Dietrich Stauffer 用一个扩展的离散模拟模型来研究一个给定的极端舆论如何在社会中传播^[10]。

也有学者研究具体的虚拟社区，如 Michael Chau 用博客蜘蛛从网站上抽取博客信息，包括链接，描述，分组等信息，然后用网络分析方法中的拓扑分析（度的分布，聚类系数，最短路径长度），中心度分析（度，紧密度，中间性），和团体分析（聚类，类之间的影响）来对该网站的博客群体进行分析，他分析的是一个厌恶网络博客中的 820 个博客，通过分析找到了两个最大的群体^[11]。Weitong Huang 采用文本挖掘的方法分析网络论坛用户的观点倾向^[12]。Feng Fu 抽取某网站博客和博客之间的连接情况，在他抽取的数据中，共有 200292 个节点和 901607 条连接，然后他分析了这个网络结构的一些统计属性，这其中包括度和度的分布，平均最短路径，度和度的相关系数等，他同时也抽取了校内网上用户的数据作了类似的分析，这些数据包括 396836 个节点和 7097144 条边，他将这两个网站上的数据进行比较分析^[13]。Lada A. Adamic 分别从两个学校的学校社区的学生的个人主页中抽取学生的基本信息和各个学生主页之间的联系信息，然后基于这些链接和个人主页的关键词来分析学生之间的联系^[14]。

1.3.2 国内的研究现状

在我国舆论的研究中，不同的学者从不同的角度进行研究。纵观我国学者的研究，主要的研究视角有：基础理论视角，心理学伦理学视角，传播学视角，管理模式视角。

基础理论视角：陈力丹的《舆论学》首先从认识舆论入手，介绍舆论学的发展以及舆论的八要素，随后分析舆论的形成，其中包括从各个研究视角下的舆论形成过程，并且还给出舆论的一般形成过程以及影响舆论形成的宏观因素；然后作者

介绍了舆论的三种存在形式（潜舆论、显舆论和行为舆论）以及舆论的不同信息形态并给出了大众媒介应该如何引导各种形态的舆论；接着书中总结了我国当代舆论的特征并介绍了接受理论用以帮助大众媒介引导舆论；随后书中大众传播媒介引导舆论的几种方式；作者最后研究了我国当前舆论导向中存在的问题^[15]。韩运荣的《舆论学原理，方法与应用》分别从原理，方法和应用三个部分介绍了舆论学。在原理部分，着重阐述舆论学的基本原理知识：舆论三要素（客体，主体和存在形式），舆论在社会生活中扮演的角色和功能，舆论的形成过程，舆论的引导和控制。在方法部分，对民意测验方法进行梳理，涉及舆论测量的方法，特点和功能。在应用部分，书中介绍了舆论调查课题的确定原理，调查问卷的设计和调查实施报告的撰写^[16]。

心理学伦理学视角：钟瑛在《网络传播伦理》中首先讨论了网络传播内容，行为以及效果等方面出现的一些不和谐因素。即：网络传播中虚假信息的泛滥，网络内容的控制与现有社会秩序和个人权利的冲突；网络传播行为在场景上和交往上都与传统的传播模式有相当大的不同；网络传播的效果也因为网络资源的分布不均造成了严重的地区差异和群体差异。接着书中讨论了网络伦理规范的构成和意义，解读了互联网行业自律公约和多家门户网站的栏目管理规范和条例，并分析了网络规范的主要特点和网络规范的价值构建。然后分析了网络 BBS 的议题的引导和网络新闻的舆论的控制，以及网吧现状及管制情况。最后讨论了网络传播中的道德选择，以及影响道德选择的心理因素和社会因素^[17]。刘正荣的《从非理性网络舆论看网民群体心理》一文中指出我国互联网上的公众舆论十分突出。通过互联网发表意见是公众的权利，是拓宽民主渠道的一种方式。但是由于网络舆论的匿名性和开放性等特点使得一些偏激的情绪化舆论得以形成，对于这种非理性的公众舆论必须针对网民群体的心理特征，通过立体的“议程设置”加以引导^[18]。刘毅的《从社会心理学视角分析网络舆论引导》一文指出网络舆论引导的实质就是舆论引导主体使网民“态度改变”的过程，文中从社会心理学视角分析了典型报道、深度报道、网络新闻评论和意见领袖等舆论引导方式，并提出对网络舆论的引导应多“疏”少“堵”，因势利导，顺势而为^[19]。

传播学视角：吴风的《网络传播学》一书首先介绍了网络传播的发展历史，然后分析了网络媒体的运营模式并剖析网络传播特征；紧接着，作者分析了网络传播的意义，即这是一场传播学上的革命并且对传统的传媒造成了相当大的打击；随后作者分别从符号学和文化学的视角审视网络传播；然后作者简要介绍了一下随着网络传播发展而兴起的网络舆论和网络论坛；最后书中谈了网络传播中出现的问题，这些问题包括社会伦理和社会责任意识方面的问题以及相应的法制建设^[20]。匡文波的《网络传播理论与技术》一书从介绍网络传播的概念入手，介绍了网络传播的优势和它给新闻传播带来的影响；紧接着讨论了网络传播的类型，包括网络人际传播、

网络群体传播、网络组织传播，网络大众传播；然后作者分析了网民的特征，从基本特征到网民的心理特征，上网动机以及网民的行为特征；随后作者归纳了网络传播带来的负面效应以及相应的管理对策；然后作者介绍了网络媒体的微观经营，包括网络广告，电子商务等方面的发展；接着作者介绍了移动网络传播，即借助手机媒体进行的传播活动这一最新的传播模式的发展情况；然后作者介绍了数字文献格式，包括 HTML、XML 和 PDF 格式；接着书中介绍了网页设计原理，网页设计工具，动画制作的知识和多媒体网络传播技术；最后作者介绍了网络信息检索技术^[21]。喻国明的《基于神经科学的传播学研究》一文是对于如何基于神经科学进行传播学研究的基本范式所进行的描述和阐述，它扼要介绍厚薄规和解读了基于神经科学的传播学研究的工具、路径和研究框架^[22]。

管理模式视角：王天意的《网络舆论引导与和谐论坛建设》一书首先介绍互联网在中国的发展历程以及它对中国的影响和挑战；紧接着介绍了中文网络论坛的发展，并对中文网络论坛分类，归纳特点和主要功能，给出了中文网络论坛发展中存在的问题；随后作者介绍了网络舆论，从概念特征到变迁过程，从网络舆论的功能到对社会的影响；然后书中重点讨论了网络牢骚和网络恶搞两种网络社会现象，对它们的特征，造成的影响和形成的现实根源做出分析；最后作者对如何构建和谐中文网络论坛做了研究，他先给出理论框架，然后讨论了现实路径，结合一些实践探索给出了意见和建议^[23]。燕道成的《国外网络舆论管理及启示》一文通过解读国外网络舆论的管理模式对国内的网络舆论管理提出一些建议。文中首先指出国外网络舆论管理的归属都是以传统管理模式为基础，即依据对新事物的认识总是以旧事物为参照这一传统惯例；随后文中提出没有哪个政府放弃对网络舆论的管理，说明对网络舆论的管理是必须的；接着作者发现由于各个国家的网络发展情况不同，文化不同，对网络舆论的管理也出现了许多不同的方式；并发现国外网络舆论的管理的具体方式是多元化管理，即集合政府立法管理，技术手段控制，网络行业与用户自律和市场机制调节等方法的管理；最后作者得到一些对我国管理的启示，即网络管理方式应多管齐下，网络舆论引导方式要整合疏导^[24]。

对于舆论传播的定量方法也有探讨，学者们一种思路是给定某种假设，让计算机模拟出一个舆论传播模型出来。肖海林假设人员态度的改变方式遵守少数服从多数规则，通过计算机模拟给出了二维四方格子影响领域在周期性条件下的舆论传播行为，模拟给出了模型的演化趋势，最后结论人员的移动将加快舆论的形成^[25]。刘常昱利用小世界模型构建人际关系网络拓扑，通过设计个体的局部相互作用规则，引入个体心理因素和外界影响，利用计算机仿真建立舆论传播模型，并分析该模型的建构过程^[26]。刘常昱引入个体意见的不确定性以及个体间的信任因子，建立了个体间的不对称影响函数，以此为基础实现了一个舆论涌向的 Agent 仿真模型^[27]。

1.3.3 目前研究的不足之处

第一，关于舆论研究国内外大多都是从理论出发，分析舆论传播的形成，要素，受众的心理因素等，这方面的研究做得很充分。缺乏对舆论传播的定量描述性分析，这是因为对传统意义上的舆论传播数据的采集非常的困难，这就形成了一个很尴尬的局面，一方面理论上的研究很充分，另一方面实证上的研究又非常缺乏。

第二，国内外不少学者通过构建各种模型来模拟舆论的传播，从中获取模拟的舆论传播数据，这确实是一个比较有效的定量分析舆论传播的思路。但是无论假设有多充分，模型有多完美，这毕竟是模拟出来的数据，没有真实的数据那种说服力。

第三，在舆论管理特别是网络舆论管理控制上，很多方法都是凭借经验归纳总结出来的管理办法。虽然在大多情况下，这些办法都非常有用，但是这些方法也有一些弊端，因为这些方法不是通过对实际数据的科学分析而归纳总结出来的，缺乏对舆论传播数据的精确分析，然后得出的管理方案。

1.4 研究的主要内容和思路

整个研究分为数据采集和预处理阶段，数据分析阶段和管理应用阶段。

在数据采集和预处理阶段，主要的工作是采集网络论坛用户之间相互发帖和回帖的记录即用户交流数据，以及将这些记录按照分析的要求进行预处理。由于这两个步骤是结合的非常紧密的，即在采集数据的时候就要提前做一些预处理，所以我将这两个步骤放在了一起。在这个阶段将要涉及对网络论坛后台数据库的数据提取和转换处理，以及一些数据格式上的转换。由于该数据库是 **Mysql**，因此在数据采集阶段要涉及对数据库 **Mysql** 的大量操作。在数据格式转换阶段，会涉及对 **Excel** 的一些操作。

数据分析阶段是整个研究的重点阶段，也是研究重要的创新之处。数据分析用的主要方法主要有社会网络分析方法和数据挖掘。社会网络分析方法涉及的分析软件是 **Pajek**。分析的具体思路是：一方面分别用两个模型即“用户——用户网络模型”和“事件——用户网络模型”（这两个模型是社会网络分析方法的两个基本模型，将在后面作介绍）处理分析用数据。然后分析用户——用户网络模型的网络性质：中心度，中心势，对于事件——用户网络分析事件的中心度，分析事件与事件之间的联系等特征，这些分析大多是描述性分析，将结合图形和帖子的实际内容做详细的说明，从中找出用户和帖子在网络中的地位，归纳出与此相关的规律。另一方面，从数据库中找到 50 个最火的帖子，构建模型描述每一个帖子的扩散时间分布情况，然后用凝聚层次聚类法将这些帖子聚类，同时，调查这些帖子所属的板块和涉及的主题内容，然后将帖子的这些特征加上聚类结果做关联规则分析，寻找它们之间的关系，由此得出帖子扩散时间分布的规律。

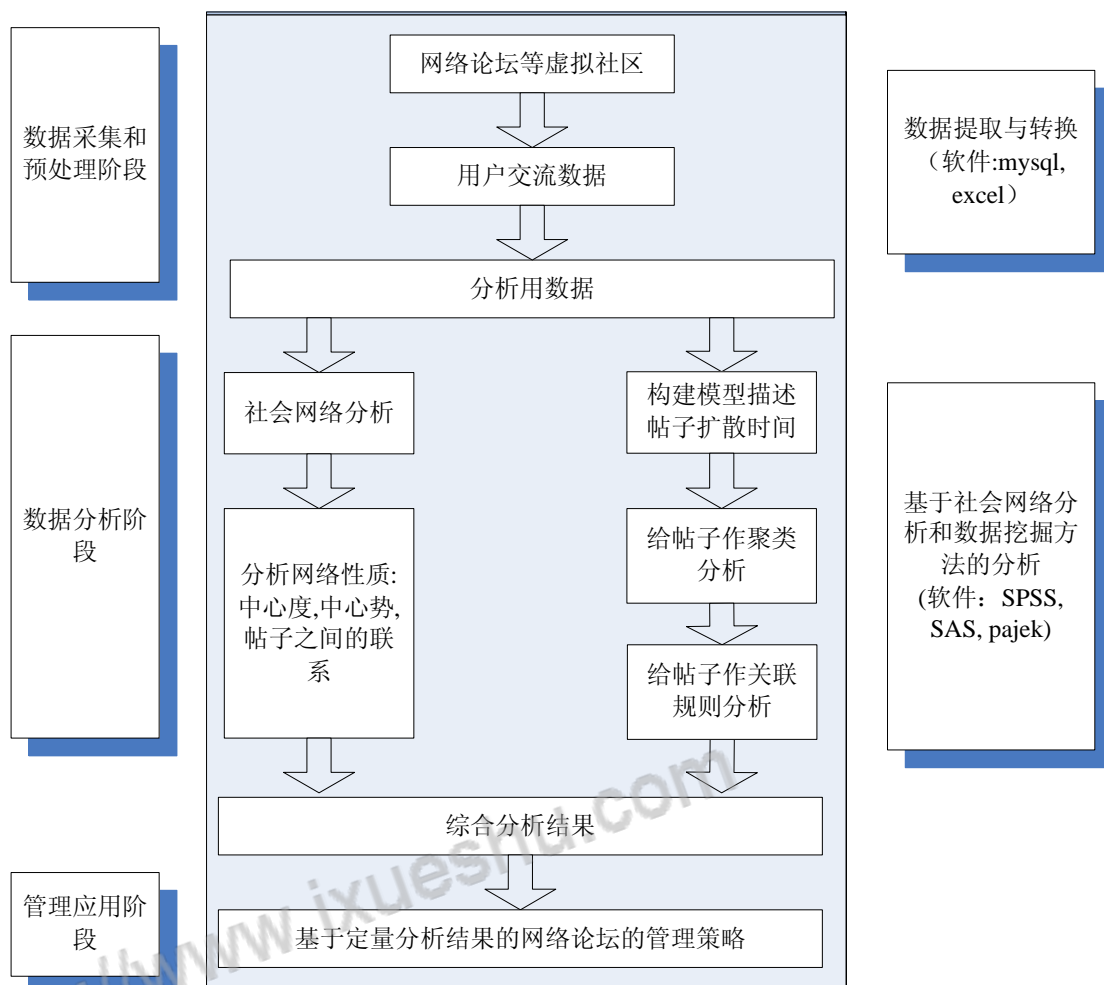


图 1.1 研究总体思路图

Fig.1.1 Structure of the whole study

管理应用阶段是针对分析结果提出一些管理上的建议。通过分析结果，让我们了解网络舆论在网络论坛中的形成和传播的一些规律。这些管理建议的目的是让网络论坛的管理者能够根据这些规律更早的发现网络舆论的出现，避免一些危害社会和谐的，虚假的，带有负面影响的舆论的扩散，同时做出正面的，和谐的，正确的舆论导向。

本章小结和以后各章介绍：本章首先介绍了研究背景，研究意义和研究目的，然后是国内外研究现状和存在的不足，最后给出了本研究的研究思路。第二章，主要介绍社会网络分析方法中的“用户——用户网络模型”和“事件——用户网络模型”和一些分析参数；同时对数据挖掘方法中的聚类分析和关联规则分析做了简要的介绍。第三章，用数据挖掘方法分析 50 个最火帖子的生命周期规律，从中找出一些舆论传播的规律。第四章，用社会网络分析方法分析某天某板块所有用户之间的

相互回帖情况，寻找用户在网络舆论传播中所处的地位，通过用户寻找帖子与帖子之间的关系，从中得出一些与用户相关的网络舆论传播规律。第五章，总结第三章和第四章的分析结果，并根据分析结果提出一套管理方案。第六章，结论和总结研究的不足之处，并指出以后的研究方向。

<http://www.ixueshu.com>

2 社会网络分析方法和数据挖掘方法介绍

2.1 社会网络分析方法

二十世纪 30 年代, Jacob Moreno 和哈佛大学的一组研究人员分别提出了社会网络模型来分析社会学中的现象和问题。早期的社会网络局限于一个小的团体之内, 往往仅包含几十个结点。但是, 随着现代的通信技术的发展, 越来越多的数据被收集和整合在一起, 建立一个大的社会网络成为可能。社会网络分析是一种应用性很强的社会学研究方法, 成功地解决了一些社会学问题上, 得到了广泛的关注。

社会网络分析是一套用来分析多个个体通过相互联系构成的网络的结构, 性质以及其他用于描述这个网络的属性的分析方法的集合。如社会网络分析方法提供了根据网络中节点的联系紧密情况将网络分层的方法^{[28][29]}, 网络中节点相互作用模式识别^[30], 将网络分块^[31], 给用户评级^{[32][33]}, 信息扩散, 对社会网络提供图形描述^{[34][35]}, 中心度的分布^{[11][13]}等。Stanley Wasserman 和 Peter J. Carrington 相继在^{[36][37]}中详细介绍了社会网络分析的各种模型及应用。

社会网络分析已广泛应用于分析国家贸易关系、社会人物关系、发现和理解社会结构、研究组织通信行为。作为 SNA 的新的分支, 犯罪网络 (Crime Network) 的研究受到各国政府和相关机构的特别关注。Jennifer J. Xu 设计了一个犯罪网络知识发现体系 Crime Net Explorer, 能够帮助执法机关高效准确地发现犯罪网络, 并将其结构可视化显示, 该体系集成了多种数据挖掘技术, 如实体抽取技术自动从警察犯罪记录中鉴别罪犯身份、地址; 聚类分析查找用相同方法作案的嫌疑犯或区分不同的犯罪组织; 偏差检测用于发现欺诈犯罪行为、网络入侵和其它的犯罪方式^[38]。邮件挖掘是反犯罪的重要技术, Stolfo, S.J, Hershkop, S. et al 提出了一种基于个性特征的邮件模拟系统, 该系统可以根据不同用户的性格特征来推测其对邮件收发的影响, 但随机性较大; 同时, 通过分析邮件通信流, 挖掘相关邮件主题, 获取个人相关信息, 为执法机关的侦察和分析提供决策支持^[39]。

下面我们介绍社会网络分析最重要的两个模型, 用户——用户网络模型和用户——事件网络模型

2.1.1 用户——用户网络模型

在网络论坛中, 我们把用户作为节点, 用户之间的联系作为节点之间的连线, 构成一个社会网络。用户之间通过相互回复对方的帖子产生联系, 一个用户对另一个用户的回复可以是一次也可以是多次, 并且是带有方向的, 因此这是一个带有数值的有向网络。如果用户 A 回复了 B 的帖子, 则产生弧 AB, 这条弧由 A 指向 B, 回复的次数就是这条弧的值。整个网络中节点的数量用 l 表示。每个结点用 v_i 表示,

i 从 1 到 5。

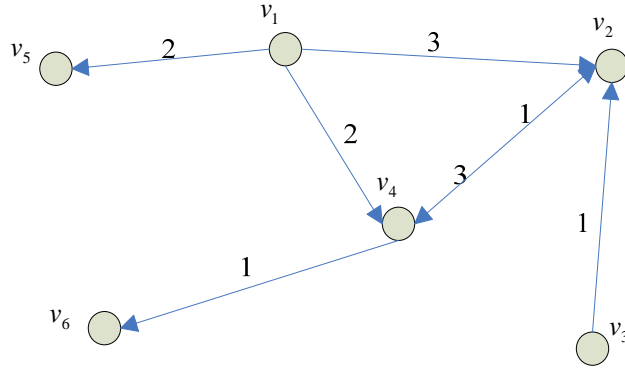


图 2.1 用户——用户网络示意图

Fig.2.1 One-mode net schematic diagram

图 2.1 给出了一个由 6 个用户构成的简单的“用户——用户网络”示意图。图 2.1 中，结点 v_1 到 v_2 有 1 条弧，值为 3，表示用户 v_1 回复 3 次用户 v_2 发的帖子。结点 v_2 与 v_4 之间是双向弧，值为 1 和 3，表示用户 v_2 回复 3 次 v_4 发的帖子，用户 v_4 回复了 1 次 v_2 发的帖子，以此类推。

威望度计算：在一个有向带值得网络中，一个结点的威望度是指这个结点的入度与所有节点的入度和的比值。

$$P_D(v_i) = \frac{x_{i-}}{\sum_{i=1}^l x_{i-}} \quad (2.1)$$

式(2.1)是威望度的计算公式，其中 x_{i-} 表示节点 v_i 的入度。入度是指所有指向该结点的弧上的值的和。如图 2 中，节点 v_4 的入度为 $2+3=5$ ，威望度为 $5/13=0.385$ ；节点 v_6 的入度为 1，威望度为 $1/13=0.077$ ；结点 v_3 的入度为 0，威望度为 $0/13=0$ 。一个结点的威望度越高，该结点所代表的用户的帖子被其他用户回复的次数就越多，该用户在论坛中所处的位置就越重要。

中心度计算：在一个有向带值的网络中，一个结点的中心度是指这个结点的出度与所有结点的出度和的比值。

$$C_D(v_i) = \frac{x_{i+}}{\sum_{i=1}^l x_{i+}} \quad (2.2)$$

式(2)为中心度的计算公式，其中 x_{i+} 表示节点 v_i 的出度。出度是指该结点指向其他结点的所有的弧上的值的和。如图 2 中结点 v_1 的出度为 $2+3+2=7$ ，中心度为 $7/13=0.538$ ，结点 v_4 的出度为 $1+1=2$ ，中心度为 $2/13=0.154$ 。一个结点的中心度越高，

表示该结点所代表的用户回复别人的次数就越多，说明该用户在论坛中越活跃。

2.1.2 用户——事件网络模型

在网络论坛中，用户之间除了相互回复这种联系以外，用户还因为同时参与一个事件而联系，而事件也因有相同的用户参与而联系。在论坛中，我们把贴子定义为事件，用户无论是发起贴子还是回复贴子都定义为参与了该事件。通过用户与事件之间的联系构建网络拓扑图，假如用户 v 参与了事件 e ，则在 v 和 e 之间连线，因为我们这里主要研究事件传播的广度，因此我们不考虑用户对同一事件的多次参与，又因为这种联系只在两种不同类型的对象（用户和事件）之间存在，所以方向已经没有意义，因此这是一个不带值的无向图。

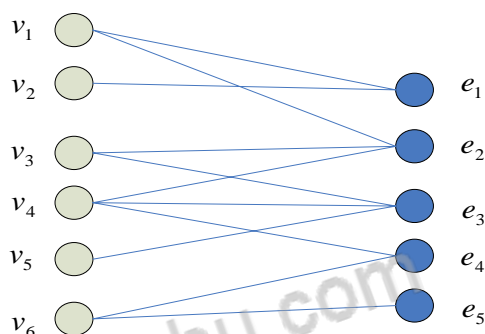


图 2.2 用户——事件网络示意图

Fig.2.2 Two-mode net schematic diagram

图 2.2 给出了一个由 6 个用户和 5 个事件构成的简单的“用户——事件网络”示意图。图 3 中，左边的 6 个圆球表示 6 个用户，由 v_i 标出， i 从 1 到 6；右边的 5 个圆球表示 5 个事件，由 e_i 标出， i 从 1 到 5。图中的线连接用户与事件，如 v_1 到 e_1 有一条连线，表示用户 v_1 参与了事件 e_1 ，以此类推。

事件的中心度计算：事件的中心度是指参与该事件的人数与总人数个数的比值。

$$C_E(e_i) = \frac{x_i}{n} \quad (2.3)$$

式(2.3)为事件中心度计算公式，其中 x_i 表示参与 e_i 事件的用户， n 表示该图中总的用户个数。如图 3 中 e_2 事件的参与人数为 3，事件中心度为 $3/6=0.5$ 。

事件之间的联系。在图 2-2 中，事件之间没有直接联系，但是却存在间接联系，我们认为拥有相同的用户参与的两个事件之间存在联系，拥有相同用户的数量越多则两个事件联系越紧密。如，事件 e_1 和 e_2 拥有一个相同的用户（ v_1 ），则它们的联系强度为 1；事件 e_2 和 e_3 拥有 2 个相同的用户（ v_3, v_4 ），则它们的联系强度为 2。通过这种方式建立只有事件之间的联系而把用户从网络中剥离出来得到新的“事件

——事件网络”模型，可以很方便的找出联系最紧密的事件。

2.2 数据挖掘方法

数据挖掘(Data Mining)就是从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中，提取隐含在其中的、人们事先不知道的、但又是潜在有用的信息和知识的过程。与数据挖掘相近的同义词有数据库中的知识发现(KDDKnowledge Discovery in Database)、数据分析、数据融合以及决策支持等。这个定义包括好几层含义：数据源必须是真实的、大量的、含噪声的；发现的是用户感兴趣的知识；发现的知识要可接受、可理解、可运用；并不要求发现放之四海皆准的知识，仅支持特定的发现问题。即所有发现的知识都是相对的，是有特定前提和约束条件，面向特定领域的，同时还要能够易于被用户理解。最好能用自然语言表达所发现的结果。

数据挖掘的任务是从数据中发现模式。模式有很多种，按功能可分为两大类：预测型模式和描述型模式。第一种是预测型模式，即可以根据数据项的值精确确定某种结果的模式。挖掘预测型模式所使用的数据也都是可以明确知道结果的。第二种是描述型模式，即对数据中存在的规则做一种描述，或者根据数据的相似性把数据分组。描述型模式不能直接用于预测。数据挖掘涉及多学科技术的集成，包括数据库技术、统计学、机器学习、高性能计算、模式识别、神经网络、数据可视化、信息检索、图像于信息处理和空间数据分析。

数据挖掘是一门发展非常迅速的技术，Dunham[40]，Geatz[41]等对数据挖掘技术进行了详细的介绍。数据挖掘方面的期刊也比较丰富，如：《IEEE Transations of Knowledge and Data Engineering》、《Data Mining and Knowledge Discovery》、《Intelligent Data Analysis》等。

这里主要介绍将在后面运用到的关联规则分析和聚类分析。

2.2.1 关联规则分析

关联规则分析首先由 Agrawal 等人提出^{[42][43]}，用来发现购物篮数据事务中各项之间的有趣联系。从那以后，进行了广泛的研究，以解决关联分析的概念，实现和应用问题。

关联规则是形如 $X \rightarrow Y$ 的蕴含表达式，其中 X 和 Y 是不相交的项集，即 $X \cap Y = \phi$ 。关联规则的强度可以用它的支持度和置信度量。支持度确定规则可以用于给定数据集的频繁程度，而置信度确定 Y 和在包含 X 的事务中出现的频繁程度。支持度(s)和置信度(c)这两种度量的形式定义如下：

$$s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \quad (2.4)$$

$$c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \quad (2.5)$$

关联规则的挖掘问题就是生成所有满足指定的最小支持度和最小置信度的关联规则。满足最小支持度和最小置信度的关联规则称为强关联规则。

关联规则的挖掘是一个两部的过程：

①找出所有的频繁项集：根据定义，这些项集出现的频繁性至少和与定义的最小支持计数一样。

②由频繁项集产生强关联规则：根据定义，这些规则必须满足最小支持度和最小置信度。

关联规则的算法有：搜索算法，该类算法只适合于项集数量相对较小的数据集的关联规则挖掘；分层算法，Apriori 算法是这类算法的典型代表，该算法由 Agrawal 等人在 1993 年首先提出，Dunkel 和 Soparkar 研究了 Apriori 算法的时间和存储复杂度^[44]，该算法扫描数据集的次数等于最大频繁项目集的项目数；深度优先算法，此类算法中最高效的是 J.Han 等人提出的 FP-growth 算法^[45]；划分算法，划分算法的基本思想是将整个数据集划分成可以存放在内存中进行处理的数据块，以节省访问外存的 I/O 开销。

在关联规则中有涉及到的数据维度，可以分为单维关联规则和多维关联规则，比如：“啤酒→尿布”就是单维关联规则，“性别=‘女’→职业=‘秘书’”就是多维关联规则，在第三章的分析中，研究使用的就是多维关联规则。

Apriori 算法是关联规则分析中最经典的算法。Apriori 算法是通过有候选项集的方法来产生频繁项集，Apriori 算法原理：任何一个频繁项集的子集必定是频繁项集。具有最小支持度的项目集也就是频繁项集。Apriori 算法对数据库进行多趟扫描来发现所有的频繁项目集。在第一趟扫描中，Apriori 算法计算数据库 D 中所有单个项目的支持度，生成所有长度为 1 的频繁项目集。在以后的每一趟扫描中，首先以前一趟中生成的频繁项目集为基础，连接生成所有新的候选项集，即潜在的频繁集的支持度，最后确定候选项集中哪些真正成为频繁项集。重复上述过程直到再也发现不了新的频繁项集。

Apriori 算法基本步骤如下：

- ①首先计算所有的 C_1 ；
- ②扫描数据库，裁剪其中的非频繁子集，生成 L_1 (1-频繁项集)；
- ③将 L_1 与子集连接生成 C_2 (候选 2-项集)；
- ④扫描数据库，裁剪 C_2 中的非频繁子集，生成 L_2 (2-频繁项集)；
- ⑤如此类推，通过 L_{k-1} (k-1-频繁项集)与自己连接生成 C_k (候选 k-项集)，然后扫描数据库，生成 L_k (频繁 k-项集)，直到不再有频繁项集产生为止。

由此分析可行，Apriori 算法的每一趟产生频繁项集它都可以分为连接和裁剪这两个步骤。连接采用迭代的链接方法求 C_k ，用 L_{k-1} 与自身连接产生候选 k -项集 C_k 。裁剪是在扫描数据库过程中，将候选 k -项集与数据库中的事务比较，统计候选项集在事务中的支持度，将不满足最小支持度的 k -项集删除，得到频繁 k -项集 L_k 。

Apriori 算法它可分为以下两个步骤进行：第一步，从事务数据库 D 中找出所有支持度不小于用户指定的最小支持度的频繁项目集；第二步，使用频繁项目集产生所期望的关联规则，产生关联规则的基本原则是置信度必须不小于用户指定的最小置信度。由于第二步较为容易和直观，因此挖掘出所有频繁项目集是该算法的核心，占据整个计算的大部分。

Aprior 算法的中心算法描述如下：

procedure Apriori(Dataset D)

① $L_1 = \text{find_frequent_1-itemset}(D);$

② for ($k=2; L_{k-1} \neq \phi; k++$) {

③ $C_k = \text{aprior_gen}(L_{k-1}, \text{min_sup});$

④ foreach transaction $t \in D$ {

⑤ $C_t = \text{subset}(C_k, t);$

⑥ for all candidates $c \in C_t$

⑦ $c.\text{count}++;$

⑧ }

⑨ $L_k = \{c \in C_k \mid c.\text{count} \geq \text{min_sup}\}$

⑩ }

(11) return $\bigcup_k L_k$

procedure apriori_gen($L_{k-1}, \text{min_sup}$)

① foreach itemset $l_1 \in L_{k-1}$

② foreach itemset $l_2 \in L_{k-1}$

③ if ($(l_1[1]=l_2[1]) \wedge (l_1[2]=l_2[2]) \wedge \dots \wedge (l_1[k-2]=l_2[k-2]) \wedge (l_1[k-1] < l_2[k-1])$)

④ then { $c = l_1 \cup l_2$

⑤ ifhas_infrequent_subset(c, L_{k-1}) then

⑥ delete $c;$

⑦ else add c to C_k

⑧ }

⑨ return C_k

procedure has_infrequent_subset($C_k; L_{k-1}$)

① foreach ($k-1$)-subset s of c

- ② if s is not in L_{k-1}
- ③ return TRUE;
- ④ return FALSE;

2.2.2 聚类分析

聚类分析将数据划分成有意义或有用的组。聚类分析仅根据在数据中发现的描述对象及其关系的信息，将数据对象分组。其目标是，组内的对象相互之间是相似的(相关的)，而不同组中的对象是不同的(不相关的)。组内的相似性越大，组间差别越大，聚类就越好。

聚类的方法通常有 K 均值算法^[46]，凝聚层次聚类^{[47][48]}，DBSCAN^{[49][50]}。

K 均值是基于原型的，划分的聚类技术。它试图发现用户指定个数(K)的簇。基本 K 均值的算法为：

- 1) 选择 K 个点作为初始质心。
- 2) repeat
- 3) 将每个点指派到最近的质心，形成 K 个簇。
- 4) 重新计算每个簇的质心。
- 5) until 质心不发生变化。

凝聚层次聚类：首先将每一个点作为单点簇；然后重复的合并两个最近的簇，直到产生单个的，包含所有点的簇。具体算法如下：

- 1) 将每一个点分为一个簇。
- 2) 计算邻近度矩阵。
- 3) repeat
- 4) 合并最接近的两个簇。
- 5) 更新邻近度矩阵，以反映新的簇与原来簇之间的邻近性。
- 6) until 剩下需要个数的簇。

步骤 2 中计算两个点的邻近度的方法很多，最常用的是计算欧几米德距离，平方欧几米德距离，夹角余弦，皮尔逊相关系数等。

步骤 4 中合并最接近的两个簇可以有多种选择方案：

组间连接，合并两类的结果使所有原本属于两类的两两点对之间的平均距离最小。

组内连接，两类合并为一类后，类中所有点与点之间的平均距离最小。

最短距离法，两类之间的最近点代表两类的距离。

最远距离法，两类之间的最远点代表两类的距离。

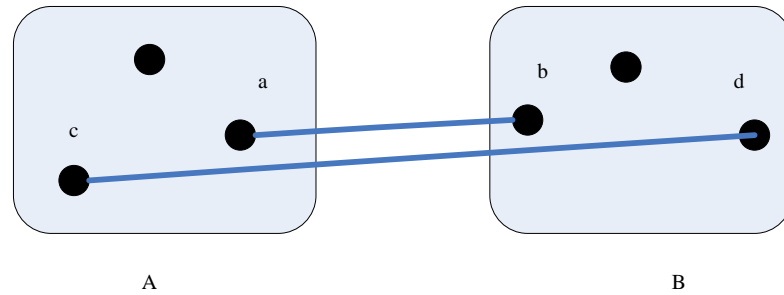


图 2.3 组间距离基于图的定义

Fig.2.3 The definition of the distance between groups based on diagram

在图 2.3 中，如果用 a, b 之间的距离代表组 A, B 之间的距离，则是最短距离法，如果用 c, d 之间的距离代表组 A, B 之间的距离，则是最远距离法。

除了这 4 种常用方法以外还有重心法，中位数法，ward 法。

DBSCAN 是一种产生划分聚类的基于密度的聚类算法，簇的个数由算法自动地确定。低密度区域中的点被视为噪音而忽略。因此 DBSCAN 不产生完全聚类。

本章小结：本章主要介绍分析中要用的两种方法，社会网络分析方法和数据挖掘方法。对于社会网络分析方法，主要介绍了“用户——用户网络”模型和用户——事件网络”；对于数据挖掘，主要介绍了关联规则分析和聚类分析。这些方法将在后面的分析中用到。

3 用数据挖掘方法分析帖子的特性

3.1 数据来源

分析的论坛是重庆大学民主湖论坛，这个论坛是重庆大学的校园论坛，数据的时间段是 2008 年 3 月到 4 月。数据收集的方式是直接操作论坛数据库，从中抽取论坛的发帖和回帖记录。该论坛在这段时间一共有 95 个板块，包括：黄桷树下，网络张贴栏，材料在线，网络一卡通，研究生院，毕业生专版，视点，08 欧洲杯等。由于该论坛是实名注册论坛，为了不伤害论坛用户的隐私权，分析的过程中都不会出现用户名，而是用用户编号代替。

3.2 分析流程

这一章主要是利用数据挖掘的相关方法分析帖子的生命周期特征与帖子的其他属性之间的关系。

首先从论坛数据中抽取 3-4 月回帖量最大的 50 个帖子，然后找出这 50 个帖子的生命周期特征，这些特征是：该帖子活跃了多久，回帖量在哪一天最多，哪一天第二多（这里的回帖只是他人的回帖，不包括自己对自己帖子的回复）。这三个特征都是能够反映帖子扩散情况的因素。然后根据这三个特征用凝聚层次聚类法给这 50 个帖子聚类，这其中包括计算帖子之间的相异度等。另一方面通过再次对数据库的查询，寻找着 50 个帖子所在板块和帖子的主题，并将这些帖子的基本属性与分类结果作关联规则分析。聚类分析用到软件 SPSS，关联规则分析用到软件 SAS。

最后根据整个分析过程和结果归纳出一些舆论传播相关的规律。本章分析流程图如图 3.1 所示。

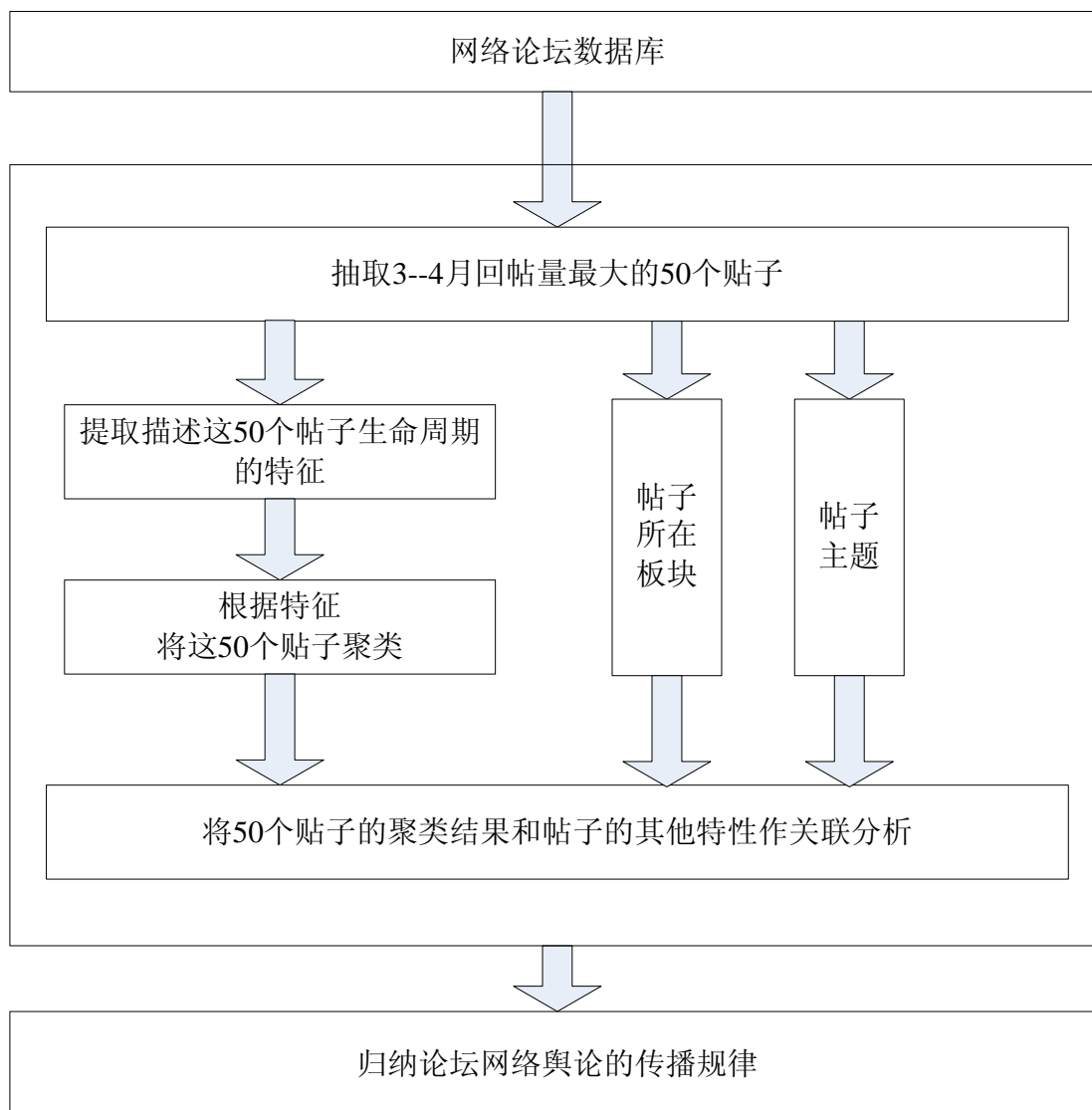


图 3.1 数据挖掘分析流程图

Fig.3.1 Flow chart of Data mining analysis

3.3 构建特征模型和聚类分析

3.3.1 数据提取

从数据库中提取 3-4 月回帖量最大的 50 个帖子。在数据库中发贴表为 thread，回帖表为 post。

表 3.1 post 表和 thread 表的结构

Table.3.1 The structure of table “post” and table “thread”	
post 表结构	thread 表结构
回帖标号	帖子编号
帖子编号	发帖人编号

续上表 3.1:

回帖人编号	所在版块
回帖内容	帖子主题
回帖时间	帖子内容
	发帖时间

数据库操作语句如下:

```
Select post.pid, post.tid as ptid, post.uid, post.message, post.dateline, thread.tid,
thread.uid, thread.subject, thread.dateline, thread.bid, thread.name, thread.message
count(*) as counts from post, thread
where post.tid=thread.tid
and post.dateline>1204243200
and post.dateline<1209427200
group by ptid
order by counts desc limit 50;
```

其中 dateline 为时间戳, pid 是回帖编号, tid 是发帖编号, uid 是用户编号, subject 是帖子标题, name 是板块标题, message 是回帖内容。找到回帖量最大的 50 个帖子后记录这 50 个帖子的帖子编号, 表 3-2 给出了这 50 个帖子中的 5 个帖子编号(整体的帖子编号见附录 B 表 b.1)。可以看出这些编号是一些没有规律的数据, 为了分析方便, 我们在后面用 1-50 作为它们的顺序号代替它们。

表 3.2 部分帖子编号

Table.3.2 The number of some posts

部分帖子编号
472239
468275
468360
466962
472434

3.3.2 构建特征模型

然后分析这 50 个帖子的特征。首先对于每一个帖子, 计算该帖发帖后每天的回帖量。比如对于帖子 465525 发帖后第一天的回帖数, 数据库查询语句为:

```
select count(*) from post where tid=465525 and
```

dateline<1205971200

and dateline>1205884800;

其中 1205884800 为发帖时间, 1205971200 为第一天的结束时间。反复更换由 dateline 记录的时间段, 由此统计出该帖自发帖后每天的回帖量表 3.3:

表 3.3 帖子 465525 每天的回帖量

Table.3.3 The quantity of reply of post 465525

第几天	回帖数	第几天	回帖数	第几天	回帖数
第 1 天	47	第 6 天	10	第 11 天	2
第 2 天	24	第 7 天	21	第 12 天	8
第 3 天	13	第 8 天	5	第 13 天	4
第 4 天	21	第 9 天	1	第 14 天	5
第 5 天	15	第 10 天	1	第 15 天	4

柱形图为:

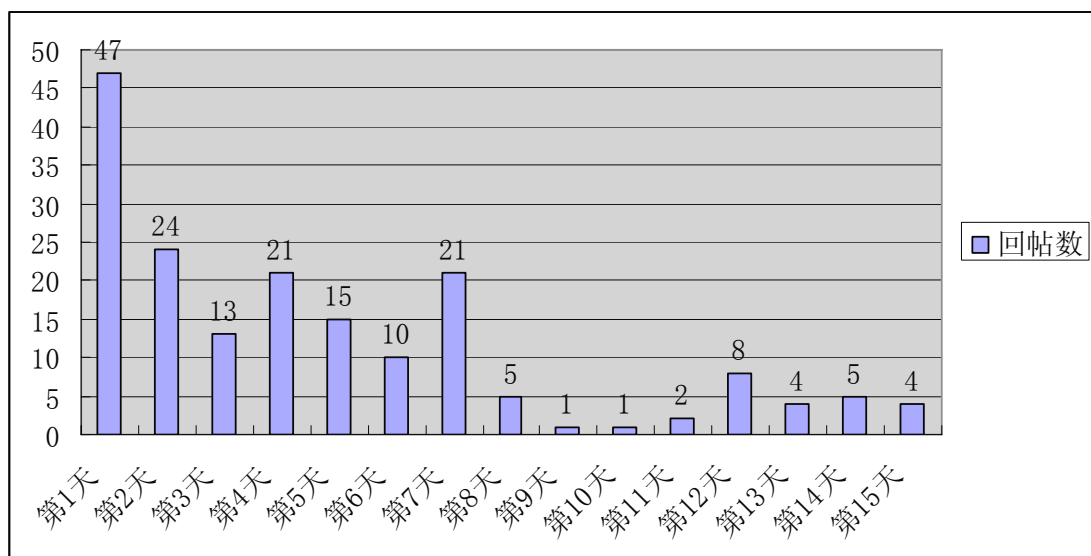


图 3.2 帖子 465525 每天的回帖量的柱形图

Fig.3.2 Bar diagram of quantity of reply of post 465525

由图 3.2 看出这个帖子的回帖是一个随天数增加而逐渐递减的过程。用这种方法统计出这 50 个帖子的每天回复量和相应的柱形图。(50 个帖子的回帖量分布柱形图见附件 B 图 b.1 和图 b.2)

然后将这 50 个帖子的回复规律进行聚类。聚类首先要计算相似性, 对于每个

帖子，抽取几个特征信息用于计算相似性。考虑到我们的研究是要研究网络舆论的传播规律，在这里我们重点调查这些帖子的爆发点和帖子的影响时间，我们选择的特征信息是：回帖持续的天数，峰值所在的天数(第几天出现回帖次数最多)，第二高峰所在的天数(第几天出现回帖次数第二多)。对于 tid=46552 的帖子来说，他的回帖持续天数是 15，峰值所在的天数 1，第二峰所在的天数是 2。表 3.4 给出了帖子 1-6 的特征值。(帖子 1-50 的特征值见附件 B 表 b.2)

表 3.4 帖子 1-6 的特征值

Table.3.4 The eigenvalue of post 1-6			
帖子顺序号	持续天数	峰值所在天数	第二峰所在天数
1	9	2	4
2	18	1	2
3	1	1	1
4	1	1	1
5	6	3	5
6	9	1	2

3.3.4 聚类分析

聚类方法选择层次聚类法。根据这些特征计算任意两个帖子的相异度，采用的度量方式是欧几米德距离平方和。见公式(3.1)

$$D_{a,b} = \sum_{i=1}^n (a_i - b_i)^2 \quad (3.1)$$

公式中 a, b 表示帖子编号，下标 i 表示特征。例如如帖子 1 和帖子 2 的相异度，按照这种方法计算每两个帖子的相异度： $D_{1,2} = (9-18)^2 + (2-1)^2 + (4-2)^2 = 86$ ，然后构造 50×50 相异度矩阵(矩阵太大在文中没有给出)。合并两类的方法选用的是组内连接，即两类合并为一类后，类中所有点与点之间的平均距离最小。类别个数设定为 5 个。层次聚类法的聚类步骤详见第二章。选用分析软件是 SPSS，操作步骤为：首先将数据倒入 SPSS，选择分析→聚类分析→分层聚类，将需要分析的变量选择到变量框，然后点击“统计量”设置聚类的个数，点击“方法”设置聚类的方法和间距测量方法，在这里可以对数据进行标准化处理，但是这里的数据都是天数，没有对原始数据进行标准化处理，然后点击“确定”开始聚类。图 3.3 给出了部分操作画面。表 3.5 给出了聚类的结果。

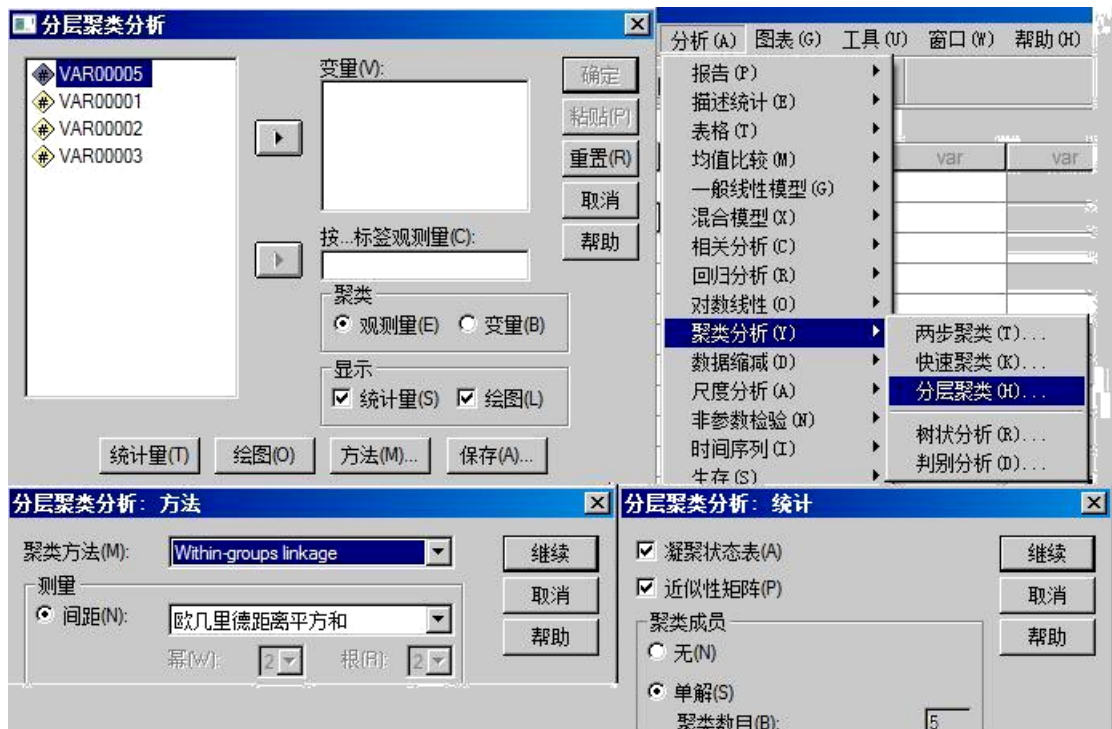


图 3.3 SPSS 分析操作过程

Fig.3.3 Operation procedure of SPSS

聚类结果为:

表 3.5 聚类结果

Table.3.5 Clustering results

类别	帖子序列号	所占比重
第一类	1, 2, 6, 13, 16, 19, 25, 29, 37, 39, 43, 44, 47, 48, 49, 50	32%
第二类	3, 4, 5, 7, 8, 9, 10, 11, 12, 14, 15, 17, 18, 20, 22, 26, 27, 30, 31, 32, 33, 34, 35, 36, 38, 40, 41, 42, 45	58%
第三类	21, 28, 46	6%
第四类	23	2%
第五类	24	2%

每一类帖子的特征平均值为: 第一类, 回帖持续天数为 11.69, 峰值所在天数为 1.69, 第二峰值所在天数为 2.56; 第二类, 回帖持续天数为 3.46, 峰值所在天数为 2.14, 第二峰值所在天数为 2.21; 第三类, 回帖持续天数为 22, 峰值所在天数为 3.3, 第二峰值所在天数为 5.67; 第四类, 回帖持续天数为 44, 峰值所在天数为 9, 第二峰值所在天数为 14; 第五类, 回帖持续天数为 34, 峰值所在天数为 29, 第二峰值所在天数为 1。

由占 58%的第二类帖子的平均特征我们可以看出，大多数帖子的影响力只有 3 天时间，即 3 天后它们就会消失，不会再有人关注。并且这些帖子的爆发点即回帖量最大的时间在第 2 天。占 32%的第一类帖子的平均特征显示这类帖子的持续时间稍长，能够达到 10 天左右，并且用户对他们的关注是随着时间的推迟，而逐渐降低，他们的爆发点是在第 1 天和第 2 天。占 6%的第三类帖子影响的时间更长，达到 22 天，并且它们的爆发点在第 3-5 天左右。第四，第五类帖子所占比率太低，并且都只有一个实例，不便于做进一步分析。

3.4 帖子属性和关联规则分析

3.4.1 帖子属性

为了进一步分析这些帖子的特征，研究重新查询数据库，查这些帖子所在板块和帖子所涉及的主题内容。表 3.6 给出了部分帖子所在板块和主题内容。(50 个帖子所在板块和主题内容详见附录 B 表 b.3)结果如下：

表 3.6 部分帖子所在板块和主题内容

Table.3.6 The plate and main idea of some post		
帖子号	板块	主题内容
1	黄桷树下	演讲报告
2	心语馨愿	贴图
3	黄桷树下	水贴
4	黄桷树下	水贴
5	视点	演讲报告
6	心语馨愿	爱情
7	黄桷树下	水贴

在主题内容中，将“电视情感”划分到“爱情”类，将“贴图激励生活”，“球赛贴图”，“运动会贴图”，“贴图 贴漫画”划分到“贴图”类，将“战争”，“火炬传递”，“家乐福事件相关”划分到“政治”类，将“食堂问题”，“饭馆问题”，“演讲报告”划分到“学校生活”类。由上表可以看出：在主题内容上，这 50 个帖子中，有 17 个水贴，所谓水贴就是没有什么实质主题，大多都在灌水和闲聊的帖子；有 15 个帖子是与爱情相关的；有 10 个帖子是贴图的帖子；有 5 个帖子是关于学校的演讲报告和食堂等涉及学校生活的问题；有 3 个帖子是关于战争，火炬等政治相关的帖子。

在板块分布上，有 26 个帖子来自“黄桷树下”，9 个帖子来自“心语馨愿”，3 个帖子来自“视点”，2 个帖子来自“回收站”，2 个帖子来自“鱼游天下”，其他在“爱国家”，“08 奥运”，“激情体育”，“好摄之徒”，“重大影像”，“轻松一刻”，“机械雄风”，“网络张贴栏”这 8 个板块中各有 1 个帖子。

3.4.2 关联规则分析

现在对于每一个帖子，有三个描述项：聚类的类别，所在板块，主题内容。比如对于帖子 3，它属于聚类中的第二类，板块分在“黄桷树下”，主题内容是“水贴”。又如帖子 5，它属于聚类中的第二类，板块分在“视点”，主题内容是“学校生活”。把帖子编号当作事务编号，把帖子的特性当作项集，采用关联规则分析这 50 个帖子的三个项之间的关联关系，分析这些属性之间的关联关系，最小支持度设定为 12%，最小置信度设定为 12%，分析软件选用 SAS。

表 3.7 帖子 3, 5 的属性示意表

Table.3.7 The attribute of post 3,5

TID	项集
3	{第二类, 黄桷树下, 水贴}
5	{第二类, 视点, 学校生活}

图 3.4 给出了项目的节点示意图，其中 SASHELP.BBS4 是数据导入节点，Association 是关联规则分析节点

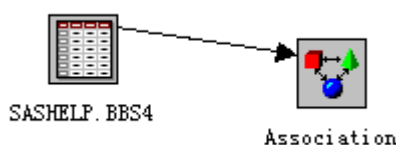


图 3.4 关联分析节点示意图

Fig.3.4 Diagram of Association analysis node

图 3.5 给出了节点 Association 设置示意图，注意把目标列的 Model Role 设置为“target”，把编号列的 Model Role 设置为“id”。

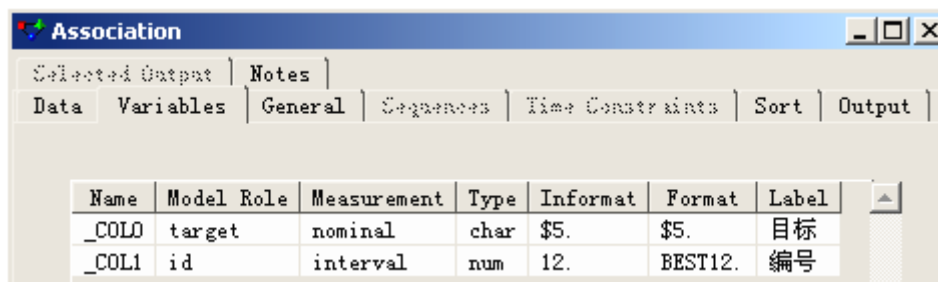


图 3.5 节点 Association 设置示意图

Fig.3.5 The setting of association node

分析结果如下：

表 3.8 关联规则分析结果

Table.3.8 The result of association rule analysis

规则编号	支持度 (%)	置信度 (%)	规则
1	44	84.62	黄桷树下→第二类
2	44	75.86	第二类→黄桷树下
3	34	100	水帖→黄桷树下
4	34	65.38	黄桷树下→水帖
5	32	94.12	水帖→第二类
6	32	55.17	第二类→水帖
7	16	88.89	心语馨愿→爱情
8	16	53.33	爱情→心语馨愿
9	14	46.67	爱情→第一类
10	14	43.75	第一类→爱情
11	12	67.67	心语馨愿→第一类
12	12	37.57	第一类→心语馨愿
13	32	94.12	水帖→黄桷树下&第二类
14	32	61.54	黄桷树下→水帖&第二类
15	32	55.17	第二类→水帖&黄桷树下
16	32	94.12	水帖&黄桷树下→第二类
17	32	100	水帖&第二类→黄桷树下
18	32	72.73	黄桷树下&第二类→水帖

规则 1 表示，在黄桷树下这个板块里的帖子有 84.62% 都属于第二类帖子，而這些帖子在整个样本中所占的比重是 44%。规则 13 表示水贴有 94.12% 都既在黄桷树

下板块并且属于第二类，这种帖子占到 34% 的比率。由这些规则我们可以看出第二类的帖子和黄桷树下这个板块，水贴主题三种联系比较紧密，相关度很高，并且所占比重最大。第一类帖子和心语馨愿板块，爱情主题联系紧密，相关度较高，所占比重比较大。

由于在种类中三，四，五类的帖子数量太少，在板块中，其他板块出现的次数也比较少，帖子主题中，水贴和爱情是出现频率最高的主体，其他主题有的出现频率太低，有的分布没有规律，所以在 12% 支持度和 12% 置信度的条件下没有相应的关联规则对应。

3.5 规律归纳

通过分析我们得出一些适用与重庆大学民主湖网络论坛的网络舆论的传播规律：

- ①大部分帖子的生命周期都比较短，只有 3 天左右的活跃时间。
- ②有部分帖子生命周期稍长，达到 10 余天，但是这些帖子最火的时间都是在前两天。
- ③小部分帖子生命周期长达 20 天以上，这类帖子最活跃的时间出现在第 3 天到第 5 天之间。
- ④极少部分帖子生命周期在 30 天以上，这类帖子活跃规律不明显。
- ⑤民主湖论坛中最火的板块是“黄桷树下”，其次是“心语馨愿”，除了这两个之外，“试点”板块稍强。
- ⑥论坛中最多的帖子是水贴，其次是与爱情相关的帖子，排在后面的帖子是贴图类的，学校生活类和政治相关类的。
- ⑦水贴，“黄桷树下”，第二类这三者联系很紧密，第二类就是生命周期比较短，只有 3 天左右活跃时间并且出现频率最高的帖子。即水贴基本上都发布在“黄桷树下”，并且大多生命周期短，最多活跃 3 天就消失；“黄桷树下”的帖子大多生命周期短，并且大多是水贴；生命周期短的帖子大多都是在“黄桷树下”，大多都是水贴。
- ⑧爱情相关的帖子，“心语馨愿”，第一类帖子这三者联系比较紧密。第一类帖子是指生命周期长达 10 天，并且最活跃的时间在前两天。即爱情帖大多在“心语馨愿”板块，并且生命周期长达 10 天，前两天最活跃；第一类帖子中大部分是爱情相关的帖子，并且多在“心语馨愿”板块；“心语馨愿”板块的帖子多是爱情帖，并且多是第一类帖。

4 用社会网络分析方法分析用户和帖子的特性

4.1 数据来源

前一章分析的民主湖论坛 2008 年是 3 月到 4 月最热门的 50 个帖子，这一章分析的是 2008 年 4 月 3 日民主湖论坛“视点”板块所有的帖子和回帖记录。所涉及的表同样是发贴表 thread 和回帖表 post。

4.2 分析流程

首先我们从数据库中抽取 4 月 3 日试点板块的所有用户的发帖和回帖记录；一方面，我们将这些数据转化为适用于“用户——用户网络”的格式然后导入 pajek 软件做描述分析，这里主要是给出各个用户之间相互联系的关系图，同时根据计算各个用户的中心度和威望度并用给出用户的中心度和威望度的图形描述。另一方面，将 4 月 3 日的这些发帖回帖记录转化为“用户——事件网络”的格式然后导入 pajek 做描述分析，这些分析主要包括两个思路，一是分析事件的中心度，二是分析事件与事件之间的联系。然后结合这些帖子的实际内容分析这些描述性分析。

最后归纳网络用户在网络舆论传播中的地位和规律。

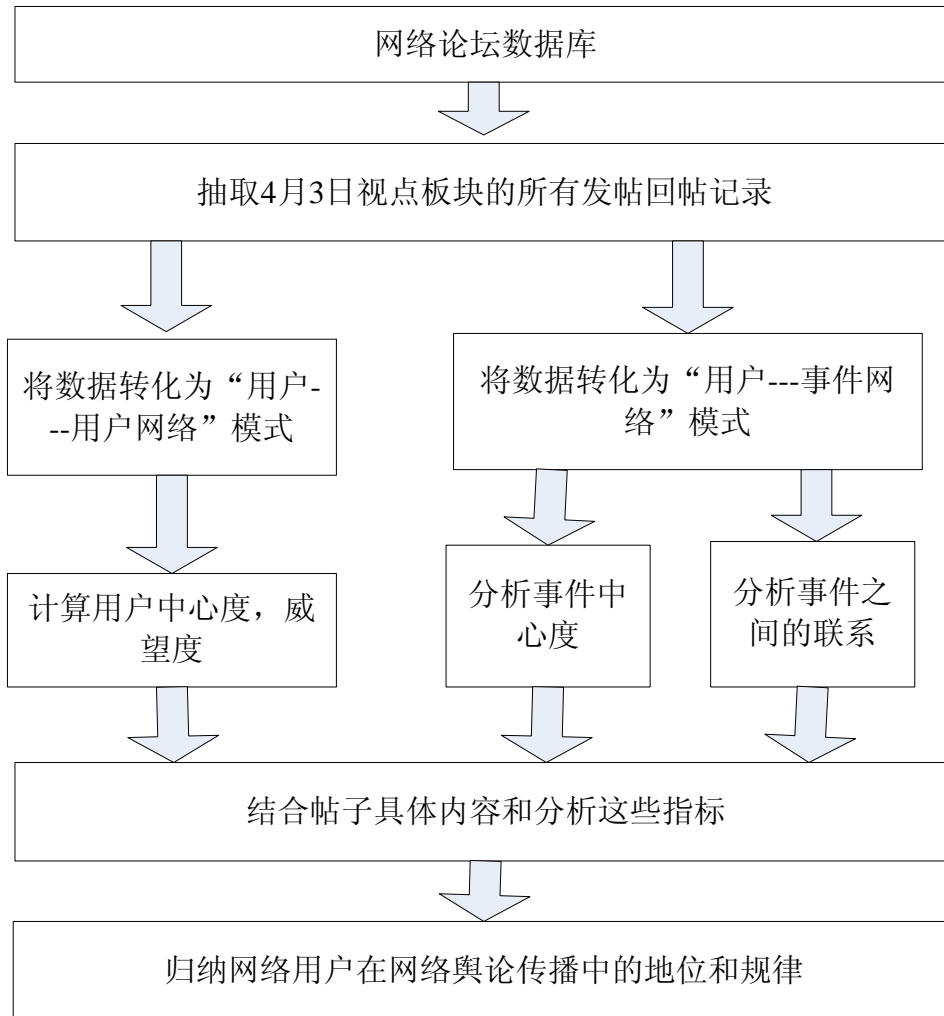


图 4.1 基于社会网络分析法的分析流程图

Fig.4.1 Analysis flow chart based on Social Network Analysis

4.3 用户——用户网络模型分析

4.3.1 数据提取和数据预处理

首先从数据库中提取 4 月 3 日“视点”板块用户之间相互回复的情况，即如果用户 2 回复了用户 1 的帖子，则用户 2 与用户 1 之间存在一个关系。先创建一个表来存放用户之间的联系，数据库查询语句如下：

```

create table shidian
(select post.uid as puid, thread.uid as tuid, count(*) as counts
from post,thread
where thread.tid=post.tid
and post.dateline>1207094400 and post.dateline<1207180800
and thread.name="视点" and post.uid<>thread.uid
group by tuid,puid order by counts desc);
  
```

新表名称为“shidian”。其中 1207094400 和 1207180800 是 4 月 3 日开始和结束的时间戳，uid 是用户编号，tid 是帖子编号。

表 shidian 的结构和部分数据见表 4.1:

表 4.1 表 shidian 的结构和部分数据

Table.4.1 The structure and some data of table “shidian”

puid	Tuid	counts
43536	64706	23
54640	50944	6
45036	50944	6
12514	33724	4
59241	64795	4
56246	64795	3
.....

其中 puid 和 tuid 是用户编号，counts 是计数项，表明前一个用户回复后一个用户的次数。

图中可以看到，用户编号用的都是数据库中的原始编号，其中有很多缺失，不是连续的，但是在用 pajek 软件分析的时候，软件只能识别连续的用户编号，所有必须把数据转换成用户编号用从 1, 2, 3.....的连续整数表示。数据转换操作语句如下:

```
create table uid43
(select distinct puid from shidian where puid not in (select tuid from shidian));
create table uid43new
(select distinct tuid from shidian);
insert into uid43new(tuid)
select puid from uid43;
alter table uid43new add(id int(10) key auto_increment);
update shidian
set tuid=(select id from uid43new where shidian.tuid=uid43new.tuid),
puid=(select id from uid43new where shidian.puid=uid43new.tuid);
```

转换步骤首先建立表 uid43 存放 shidian 表中在 puid 列中但是不在 tuid 列中的用户原始编号，建立 uid43new 表存放 shidian 表中 tuid 列中的用户原始编号，然后合并 uid43 表和 uid43new 表，合并后给这个合并的表增加 id 列，该列自然计数。最后

用这个 id 替换 shidian 表中的用户原始编号，结果如下：

表 4.2 转换后的表 shidian 部分数据

Table.4.2 Some data of table “shidian” after transformation		
puid	Tuid	counts
49	1	23
50	2	6
51	2	6
52	3	4
53	4	4
54	1	3
.....

4.3.2 用户威望度和中心度分析

表 shidian 有 482 行纪录，有 301 个用户参与其中，建立用户——用户网络。将数据倒入 Pajek 软件，构建网络结构图，如图 4-2 所示。

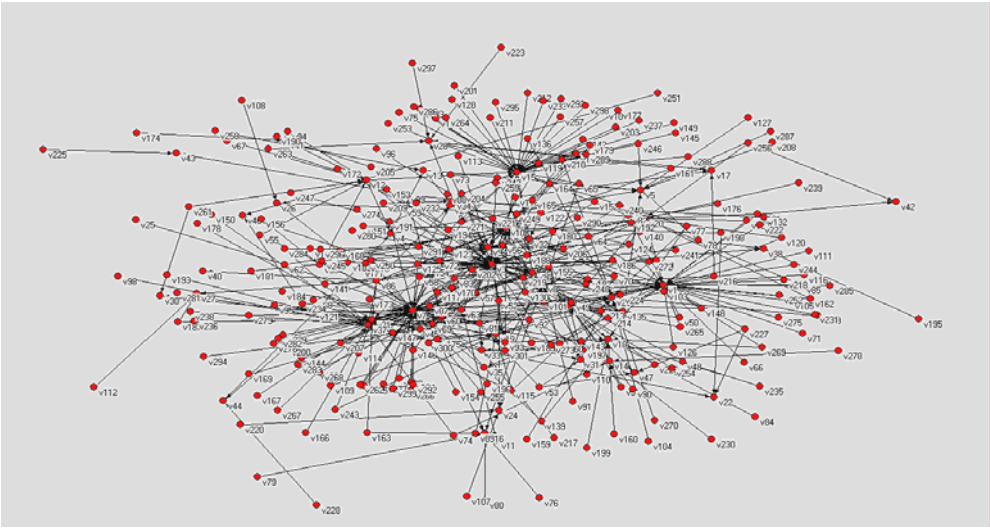


图 4.2 用户——用户网络图
Fig.4.2 One-mode net chart

图中的点表示用户，点之间的联系表示用户之间的联系，即用户之间的回复。从图中我们可以看到有的结点处于中心地位，有很多其他点与他联系，而有些结点处于边缘地位，只有很少的结点与他联系。由于用户太多，导致图形比较复杂，不能清楚的辨别处于中心的用户，下面我们计算各用户的威望度和中心度。

①用户威望度分析：根据第二章中威望度公式(1)计算出每个节点的威望度。表 2 给出了威望度最高的 30 个用户的威望度，用结点的大小表示威望度的大小，我们重新画图，得到图 4.3。

表 4.3 用户威望度

Table.4.3 Prestige of users					
用户编号	威望度	用户编号	威望度	用户编号	威望度
v3	0.153	v37	0.026	V28	0.01
v7	0.133	v21	0.024	V29	0.009
v1	0.11	v18	0.022	V17	0.007
v15	0.069	v19	0.021	V24	0.007
v2	0.045	v5	0.019	V13	0.005
v10	0.041	v23	0.019	V22	0.005
v9	0.04	v12	0.016	V26	0.005
v8	0.036	v16	0.016	V27	0.005
v20	0.028	v14	0.014	V31	0.005
v6	0.026	v4	0.01	V32	0.005

图 4.3 中，我们用较大的圆形表示威望度较大的结点，由于很多用户没有发贴，仅仅是跟贴，威望度为 0，在图中他们的大小就是 0。

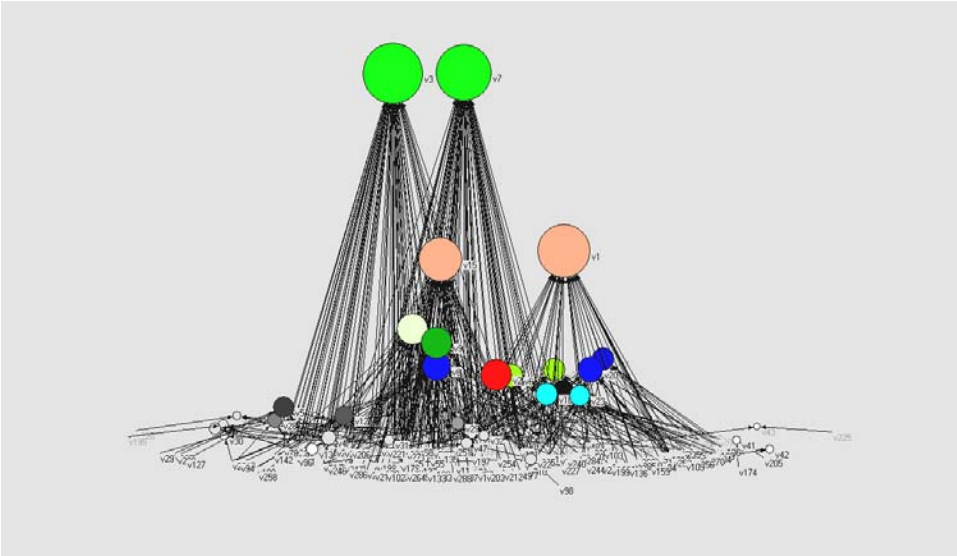


图 4.3 威望度大小图

Fig.4.3 Prestige chart

我们找出威望度最大的4个用户作进一步分析。威望度最大的四个用户为 v3(威望度为 0.153, 被回复了 89 次), v7(威望度为 0.133, 被回复了 77 次), v1(威望度为 0.110, 被回复了 64 次), v15(威望度为 0.069, 被回复了 40 次)。在这里威望度越高表明该用户的贴子被其他用户跟贴的次数越多, 我们从数据库中调出这四个威望度最高的用户在当天在该板块被回复的贴子。v3 在三天前(2008 年 3 月 31 日)发了一个题为“我为我和李开复博士用英语交流感到自豪”在(2008 年 4 月 3 日)得到了 89 次回复, 大家对这件事情进行了激烈的讨论, 原来在三天前李开复博士来重庆大学演讲, 到了观众提问阶段, 有一位同学用英语向李开复博士提问, 李开复博士也作了回答, 该同学回家后就在论坛该板块上发表了这个贴子, 说自己觉得比拿了奖学金还开心, 然后其它同学就展开评论, 有的人为她不能这样炫耀, 有的让她注意影响, 有的觉得她英文很厉害等; v7 在当天(2008 年 4 月 3 日)发表了题为“我们~不怕把事情闹大”内容为“一点小事情偏偏要紧到费”的贴子, 得到了 77 条回复, 也是对 v3 那个事件的讨论。v1 发帖希望给女生寝室安装窗帘得到其他用户的 64 次回复, v15 发帖希望给篮球场安装地灯, 得到了其他用户 40 次回复。由此我们可以发现当天在该板块的讨论热点问题, 以及处于讨论中心的热点用户。

②用户中心度分析: 根据公式(2)计算每个用户的中心度, 用结点的大小表示中心度的大小, 我们重新画图, 得到图 6。

图 4-4 中每个圆表示一个用户, 圆的大小表示中心度的大小, 圆所处的高度, 表示他所参与的贴子的多少。比如 v49 中心度最高, 但是他所涉及的贴子很少, 即他可能对同一个贴子回复了很多次(他对 v1 的“希望给女生安装窗帘”贴子回复了 23 次), 所以他在图中位置并没有 v36 高。

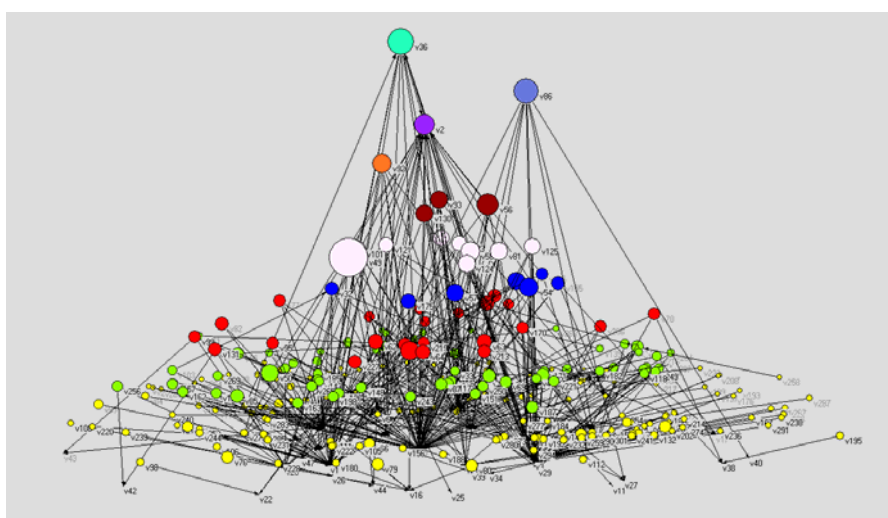


图 4.4 中心度大小图

Fig.4.4 Centrality chart

在这个模型中，中心度是描述用户回复其他用户的数量多少的（这里没有包括对自己贴子的回复），相当于用户的活跃程度。中心度最高的 4 个用户分别是：v49(中心度为 0.047，一共发言了 27 次)，v36(中心度为 0.021，一共发言了 12 次)，v86(中心度为 0.021，一共发言了 12 次)，v56(中心度为 0.017，一共发言了 10 次)。我们发现最大中心度要比最大威望度要小许多，并且威望度高的用户中心度不一定高，反之亦然。

4.4 用户——事件网络模型分析

4.4.1 数据提取和数据预处理

从数据中抽取 2008 年 4 月 3 日视点板块所有用户与帖子之间的联系，即如果用户 A 回复了帖子 B，则它们之间存在一个联系，回复的次数表示他们联系的强弱。首先从数据库中抽取数据建立一个表 `events`，操作语句为：

```
create table events
(select post.uid as puid, thread.tid as tid, count(*) as counts
from post,thread
where thread.tid=post.tid
and post.dateline>1207094400
and post.dateline<1207180800
and thread.name="视点"
group by puid,tid);
```

表 `events` 的结构和部分数据如下表所示：

表 4.4 `events` 表的结构和部分数据

Table.4.4 Structure and some data of table “events”

puid	Tid	counts
1896	473737	1
2548	473291	1
2548	473972	1
2548	474049	1
2548	474275	1
2548	474302	1
.....

其中 `puid` 是用户编号，`tid` 是事件编号。但是这些编号都是数据库中的原始编

号，在用软件分析的时候处理起来比较困难，因此对这些数据进行加工，操作语句如下：

```
create table id43(select distinct puid from evevts);
create table id43new(select distinct tid from evevts);
insert into id43new(tid)
select puid from id43;
alter table id43new add(id int(10) key auto_increment);
update evevts set puid=(select id from id43new where evevts.puid=id43new.tid),
tid=(select id from id43new where evevts.tid=id43new.tid);
```

转换的原理和上一个表相似，唯一不同的地方是，由于 puid 和 tid 没有可能相同因此少了一个排出相同的步骤，转换后 evevts 表的部分数据如表 4-5。

表 4.5 转换后的 evevts 表的部分数据

Table.4.5 Some data of table “evevts” after transformation

puid	Tid	Counts
59	1	1
60	2	1
60	3	1
60	4	1
60	5	1
61	6	1
.....

4.4.2 事件中心度分析

然后将 evevts 表转换成 pajek 能接受的格式，导入 pajek 软件，计算事件的中心度，软件操作步骤 Net→Partitions→Degree→Input，如图 4.5 所示：

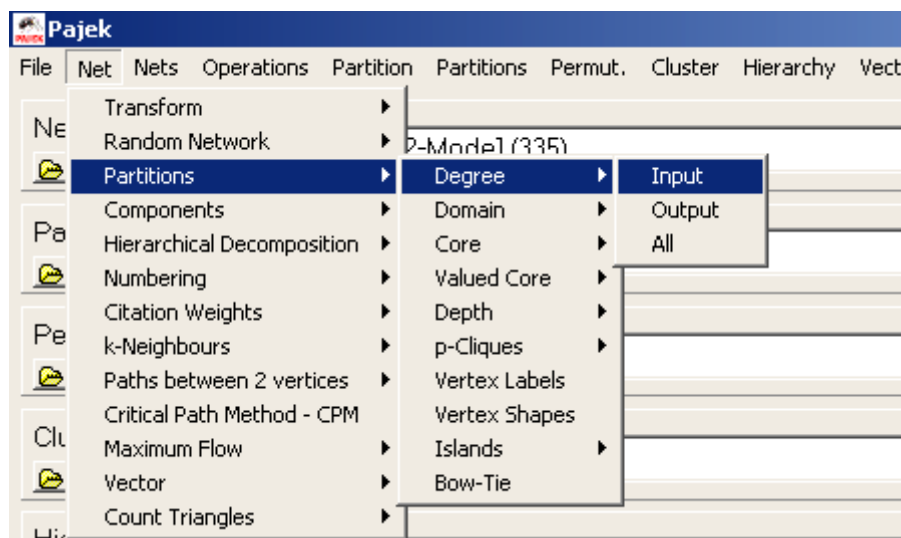


图 4.5 pajek 事件中心度计算操作图

Fig.4.5 Operation of pajek for computing events centrality

表 4.6 给出了事件中心度最大的 30 个事件的事件中心度。从表中可以看出有的事件如 e4, e8 的事件中心度非常大, 这说明这些帖子有很多用户回复, 而大部分事件如 e3, e11, e23, e36, e39 等的事件中心度很小, 只有 0.013, 说明回复这些帖子的用户非常少。

表 4.6 事件中心度

Table.4.6 The centrality of events

事件编号	事件中心度	事件编号	事件中心度	事件编号	事件中心度
e4	0.232	e5	0.047	e29	0.017
e8	0.232	e15	0.037	e35	0.017
e1	0.131	e33	0.037	e3	0.013
e12	0.128	e13	0.034	e11	0.013
e22	0.081	e10	0.03	e23	0.013
e7	0.074	e16	0.027	e36	0.013
e6	0.054	e21	0.027	e39	0.013
e2	0.051	e24	0.027	e40	0.013
e18	0.051	e14	0.024	e43	0.013
e28	0.051	e17	0.02	e9	0.01

用 pajek 作用户——事件网络图, 操作步骤 Draw→Draw-Partition-Vector,
用户——事件网络图如下图所示

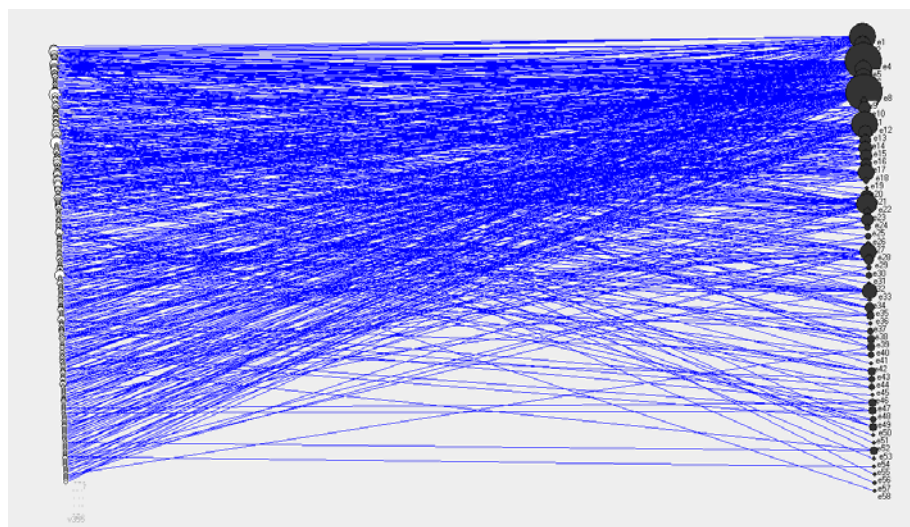


图 4.6 用户——事件网络图

Fig.4.6 Two-mode net chart

图 4.6 中，左边的圆球代表用户，圆越大说明该用户参与的贴子越多；右边的圆代表事件，圆越大说明参与该事件的用户越多，即事件中心度越大。图 7 中事件中心度最大的四个事件分别是：e4(69 人参与，事件中心度为 0.232)，e8(69 人参与，事件中心度为 0.232)，e1(39 人参与，事件中心度为 0.131)，e12(38 人参与，事件中心度为 0.128)。这四个贴子主题分别是 e4 “我为我和李开复博士用英语交流感到自豪!”, e8 “我们不怕把事情闹大，一点小事情偏偏要紧到费”，e1 “[建议]希望学校为女生宿舍统一装上窗帘!!”，e12 “[建议]建议篮球场地装灯（打篮球的进来回复)!”。与前面分析的威望度最高的四个用户发的贴子相同。这个模型能更直接的找出影响范围最广的贴子。

4.4.3 事件之间的联系

抽出的这 58 个事件，通过拥有相同的用户个数建立事件之间的联系，即如果两个贴子拥有相同的用户回复我们则这两个贴子之间存在联系，联系的强度就是这两个贴子拥有相同用户的个数，即把这个拥有事件和用户两种节点的图转化为只有事件一种节点的图，转化的操作步骤为 Net→Transform→2-Mode to 1 Mode→Rows，如下图所示：

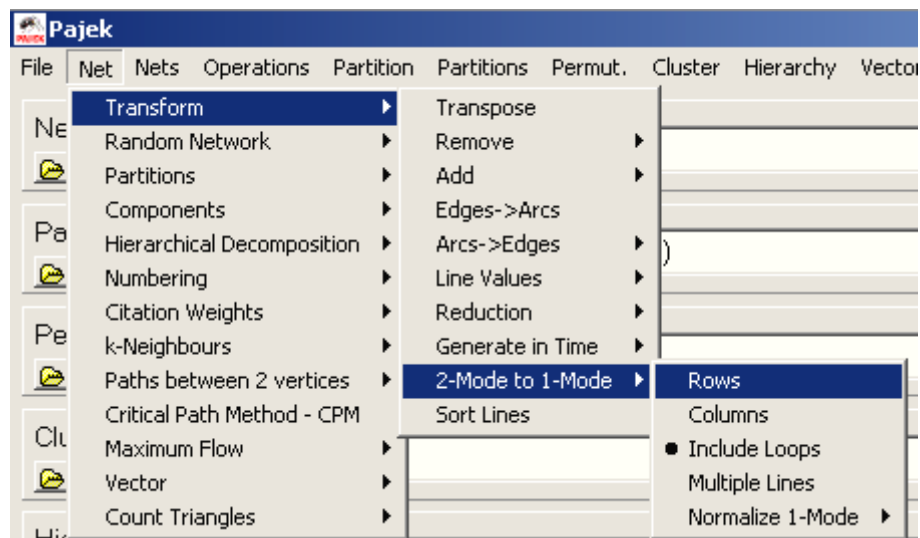


图 4.7 模型转换操作

Fig.4.7 Operation of mode transformation

然后构建事件——事件网络图形，图中的点表示帖子，点与点之间的连线是这两个帖子拥有相同的回复者，线越粗，这两个帖子拥有的相同回复者越多，则认为这两个帖子的相关性越强。图中可以明显的看出处于中心部位的几个帖子彼此之间拥有较多的相同回复者，而处于边缘地位的帖子，彼此之间很少有相同的回复者。

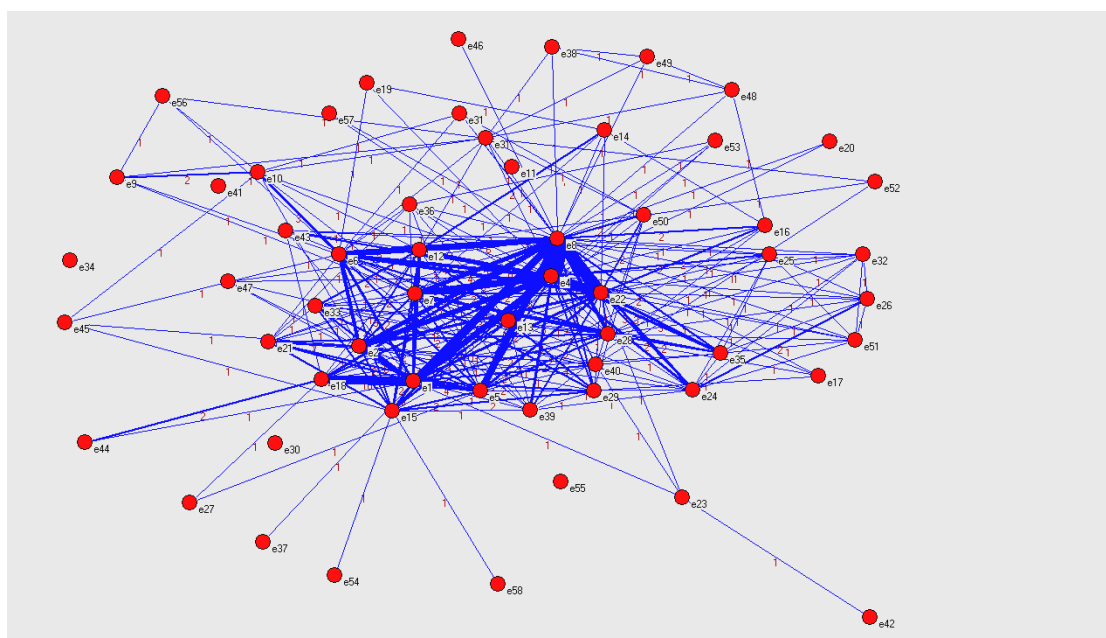


图 4.8 事件之间网络图

Fig.4.8 Net chart of events

表 4.7 给出了拥有相同用户数最多的 10 对事件。

表 4.7 事件之间的相同用户数

Table.4.7 The amount of same users between events

事件编号	事件编号	相同用户数
e4	e8	25
e8	e22	14
e1	e18	10
e4	e22	10
e1	e8	9
e4	e7	8
e5	e8	8
e1	e4	7
e6	e8	7
e1	e2	6

联系最强的两个是事件是 e4 和 e8, 拥有 25 个相同的回复者, 其次是 e8 和 e22, 拥有 14 个相同的回复者, 第三是有两组: e4 与 e22, e1 与 e18, 他们都拥有 10 个相同的回复者。e4 和 e8 前面说过了, 都是讨论用英语和李开复博士交流的事情, e22 也差不多, 贴子标题是“就‘我为我和李开复博士用英语交流感到自豪!’”说两句”, 因此这三个贴子联系紧密是很容易理解的, 他们都属于同一个话题。e1 前面也说过了是“[建议]希望学校为女生宿舍统一装上窗帘!!”, e18 的标题是“[原创]恳请学校给男生寝室安装窗帘”, 就标题来看这两个贴子也是非常相关的贴子, 属于同一种话题, 只是对象由“女生”变成了“男生”。这种模型能够从某种程度上探索事件之间的关联程度。

4.5 归纳网络用户在网络舆论传播中的地位和规律

通过对 2008 年 4 月 3 日“视点”板块发帖和回帖的分析, 得出以下几点网络舆论传播的特征:

①在特定的板块中有一种用户处于舆论中心位置, 他们的帖子拥有很多的关注, 他们的影响范围比其他一般用户要广, 可以用“威望度”寻找这种用户, 这种用户的“威望度”都比较高。

②在特定的板块中有一种用户非常活跃, 他们的关注范围非常广泛, 对很多帖子都感兴趣, 他们穿梭于各种帖子之间, 起着穿针引线的作用, 可以用“中心度”寻找这种用户, 这种用户的“中心度”都比较高。

③除了少数用户处于舆论中心地位和少数用户非常活跃以外, 其他大部分用户

都处于边缘地带，他们既不活跃，也不在引领舆论的中心。

④在论坛中有小部分帖子拥有相当大的关注程度，即有很多用户关注这类帖子，而大部分帖子受到的关注程度很小。可以用“事件中心度”来找出这种受到很多用户关注的帖子，即这类帖子的“事件中心度”很大。

⑤在论坛特定板块中有些帖子之间联系很紧密，有可能是主题相似，这些帖子拥有较多的相同的用户，可以通过计算相同用户的数量把这些帖子找出来。

5 结果总结和管理建议

5.1 总结分析的结果

通过前两章的分析，我们得到一些关于网络论坛的帖子生命周期的规律，帖子扩散规律，网络用户的特征，帖子之间相互联系的特征。这些规律和特征如下：

①大部分帖子的生命周期都比较短，只有 3 天左右的活跃时间。

②有部分帖子生命周期稍长，达到 10 余天，但是这些帖子最火的时间都是在前两天。

③小部分帖子生命周期长达 20 天以上，这类帖子最活跃的时间出现在第 3 天到第 5 天之间。

④极少部分帖子生命周期在 30 天以上，这类帖子活跃规律不明显。

⑤民主湖论坛中最火的板块是“黄桷树下”，其次是“心语馨愿”，除了这两个之外，“试点”板块稍强。

⑥论坛中最多的帖子是水贴，其次是与爱情相关的帖子，排在后面的帖子是贴图类的，学校生活类和政治相关类的。

⑦水贴，“黄桷树下”，第二类这三者联系很紧密，第二类就是生命周期比较短，只有 3 天左右活跃时间并且出现频率最高的帖子。即水贴基本上都发布在“黄桷树下”，并且大多生命周期短，最多活跃 3 天就消失；“黄桷树下”的帖子大多生命周期短，并且大多是水贴；生命周期短的帖子大多都是在“黄桷树下”，大多都是水贴。

⑧爱情相关的帖子，“心语馨愿”，第一类帖子这三者联系比较紧密。第一类帖子是指生命周期长达 10 天，但是最活跃的时间在前两天。联系紧密的含义参看第 7 条。

⑨在特定的板块中有一种用户处于舆论中心位置，他们的帖子拥有很多的关注，他们的影响范围比其他一般用户要广，可以用“威望度”寻找这种用户，这种用户的“威望度”都比较高。

⑩在特定的板块中有一种用户非常活跃，他们的关注范围非常广泛，对很多帖子都感兴趣，他们穿梭于各种帖子之间，起着穿针引线的作用，可以用“中心度”寻找这种用户，这种用户的“中心度”都比较高。

⑪除了少数用户处于舆论中心地位和少数用户非常活跃以外，其他大部分用户都处于边缘地带，他们既不活跃，也不在引领舆论的中心。

⑫在论坛中有小部分帖子拥有相当大的关注程度，即有很多用户关注这类帖子，而大部分帖子受到的关注程度很小。可以用“事件中心度”来找出这种受到很多用户关注的帖子，即这类帖子的“事件中心度”很大。

(13)在论坛特定板块中有些帖子之间联系很紧密，有可能是主题相似，这些帖子拥有较多的相同的用户，可以通过计算相同用户的数量把这些帖子找出来。

5.2 管理建议

基于对这些规律的理解，提出以下管理建议：

①管理的实施者必须是网络论坛拥有者。因为只有网络论坛的拥有者才能在第一时间掌握自己论坛的动向，只有网络论坛的拥有者能够轻松的访问论坛的后台数据库。

②论坛拥有者应该定期了解自己论坛各板块的活跃程度，并根据板块的活跃程度的不同制定不同的具体管理措施。因为根据分析，论坛中可能存在某些板块非常的活跃，这些板块的帖子更新特别快，并且用户集中程度也很大，而这种地方往往是重大事件的爆发地点，因此在这种板块就应该加大管理力度，而那些冷门的板块，由于去的用户比较少，可以根据自己的实际情况适当分配精力。

③调查各板块帖子的主题分布趋势。虽然各板块都有自己定义的主题，但是有时候网络论坛用户并不会按照规定把帖子发布在指定板块，这就需要论坛管理者适时了解自己论坛各板块的帖子的主题分布，这样可以对一些特殊内容帖子的板块投入更多的精力。

④了解本论坛一些重要种类帖子的生命周期和爆发周期。根据分析结果看出论坛中很大一部分帖子的生命周期都比较短，相对的影响力也就较弱，而有些帖子的生命周期较长，这种帖子的数量相对比较少，但是影响时间却比较长，对于这些帖子就需要加强观察和监控力度。

⑤因为有很大一部分帖子的活跃期和爆发期都在前两天，所以对帖子的监督应当做到迅速，全面。最好建立一个制度，每天都对各板块的重要帖子作一个统计，找出当天最活跃的那些帖子，并对这些帖子的内容作归纳，对于特殊的需要及时处理的帖子要在当天做出处理。

⑥要定期的对各个板块的用户作中心度和威望度分析，并记录各板块威望度较高的用户和中心度较高的用户。时时关注威望度较高的用户的发帖情况，最好是给对他们的帖子建立一种先审查后发表的制度，即威望度较高的人发帖必须经过工作人员审核内容之后才发布出去。在需要减缓舆论传播速度的时候，可以封锁中心度较高的用户的话语权，因为这类用户在舆论的传播中起到很重要的桥梁作用。

⑦论坛中很有可能出现相同主题的不同帖子，在发现需要特别关注的主题帖子时，不能仅仅关注本贴，还应同时查看是否有其他帖子是相同主题或内容的帖子，如果有相同主题或内容的帖子，就应该对他们进行集中关注，在寻找相同类型帖子的过程中，可以考虑通过寻找拥有相同用户数的方法寻找。

⑧作为第三方，可以设置一些对网络论坛拥有者的监督措施，例如让论坛拥有者定时提供该论坛的用户活跃参数和帖子活跃参数，要求论坛拥有者必须具有分析改论坛这些参数的能力，要求论坛拥有者限定帖子回复量的上限等。

6 结论与展望

通过对网络论坛的帖子和回帖规律的分析，以及网络论坛用户通过帖子相互联系相互交流规律的分析，找到舆论在网络论坛中的传播规律。本文尝试用了聚类分析、关联规则分析、社会网络分析等定量分析方法建立分析模型，并将其运用到网络舆论的研究中并揭示出网络舆论的传播规律。本文分析网络论坛帖子的生命周期和帖子主题以及帖子所在板块之间的关系，得出一些相关的关联规则，增加了我们对网络论坛帖子特征的理解，提供了一种新的研究网络舆论传播的思路。在研究用户与帖子之间的联系的过程中，可以通过调查他们拥有的相同的用户数量来分析帖子与帖子之间的联系，这给寻找帖子之间的关联提供了一种新的思路。研究提出的管理对策建议是严格按照分析过程和分析结果提出来的对策和措施，可以供网络管理者和监管部门借鉴。

本文也存在一些不足之处：第一，本文研究所使用的数据均是来自一个大学的校园论坛，该论坛的用户类型比较特殊并且单一，使得这些研究结论不一定适合于公用论坛。第二，在使用聚类分析和关联规则分析的阶段，由于研究设计不够合理，导致分析的数据量不大，不太符合数据挖掘对海量数据分析的特性，因此可能导致分析结果不具有代表性，但是这种分析思路却是一种成功的尝试。

在未来的研究中可以选择分析公用论坛，公用论坛更具有普遍性和代表性，同时分析的难度也会更大。在未来的研究中如果能够设计出更巧妙的分析方法可以尝试提取海量的数据，充分发挥数据挖掘处理海量数据的优势。希望网络论坛的管理者能够将这些方法和自己论坛的具体情况结合起来，在实践中对寻找网络舆论的传播规律和总结网络论坛的管理方法。

致 谢

本文的研究工作是在我的导师的精心指导和悉心关怀下完成的，在我的学业和论文的研究工作中无不倾注着导师辛勤的汗水和心血。导师的严谨治学态度、渊博的知识、无私的奉献精神使我深受的启迪。从尊敬的导师身上，我不仅学到了扎实、宽广的专业知识，也学到了做人的道理。在此我要向我的导师致以最衷心的感谢和深深的敬意。

在多年的学习生活中，还得到了许多领导和老师的热情关心和帮助。

在日常学习和生活中，我的同门张科给予了我很大帮助。

感谢重庆大学民主湖论坛提供数据支持

在此，向所有关心和帮组过我的领导、老师、同学和朋友表示由衷的谢意！

衷心地感谢在百忙之中评阅论文和参加答辩的各位专家、教授！

庄经纬

二〇一〇年四月 于重庆

参考文献

- [1] 中国互联网络信息中心(CNNIC). 第 25 次中国互联网络发展状况统计报告. [2010-01-15]
<http://www.cnnic.net.cn/html/Dir/2010/01/15/5767.htm>.
- [2] 王亚南. 论博客舆论传播及其现实问题[D]. 上海: 上海外国语大学. 2008.
- [3] 杜城. 时政网络论坛角色类型分析及其社会意义[D]. 湖北: 华中科技大学. 2007.
- [4] 袁卫, 王星等译. 迈克尔.J. A.贝里,戈登.S.利若夫著. 数据挖掘[M]. 北京: 中国财政经济出版社. 2004.
- [5] 刘伯高. 政府公共舆论管理[M]. 北京: 中国传媒大学出版社, 2008.
- [6] 阎克文, 江红译. 李普曼著. 公众舆论[M]. 上海:上海世纪出版社, 2007.
- [7] 谢新洲. 网络传播理论与实践[M]. 北京: 北京大学出版社, 2004.1
- [8] 戴烟登. 美国的言论自由与国家安全——最高法院处理反战言论诉讼案之研究[D]. 台北: 台湾国立政治大学政治研究所, 1988: 2.
- [9] Paul A. Dreyer Jr, Fred S.Roberts, Irreversible k-threshold processes:Graph-theoretical threshold models of the spread of disease and of opinion, Discrete Applied Mathematics(2008), doi:10.1016/j.dam.2008.09.012
- [10] Dietrich Stauffer, Muhammad Sahimi. Discrete simulation of the dynamics of spread of extreme opinions in a society. Physica A 364(2006) 537-543.
- [11] Michael Chau, Jennifer Xu. Mining communities and their relationships in blogs: A study of online hate groups. International Journal of Human-Computer Studies 65 (1) (2007) 57-70
- [12] Weitong Huang, Yu Zhao, Shiqiang Yang, Yuchang Lu. Analysis of the user behavior and opinion classification based on the BBS. Applied Mathematics and Computation 205(2008) 668-676
- [13] Feng Fu, Lianghuan Liu, Long Wang. Empirical analysis of online social networks in the age of Web 2.0 [J]. Physica A: Statistical Mechanics and its Applications 387 (2-3) (2008) 675-684
- [14] Lada A. Adamic, Eytan Adar. Friends and neighbors on the Web [J]. Social Networks 25 (3) (2003) 211-230
- [15] 陈力丹. 舆论学[M]. 北京: 中国广播电视出版社, 1999.
- [16] 韩运荣, 喻国明. 舆论学原理、方法与应用[M]. 北京: 中国传媒大学出版社, 2005.
- [17] 钟瑛. 网络传播伦理[M]. 北京: 清华大学出版社, 2005.
- [18] 刘正荣. 从非理性网络舆论看网民群体心理[J]. 现代传播(中国传媒大学学报), 2007,(03)
- [19] 刘毅. 从社会心理学视角分析网络舆论引导[J]. 当代传播, 2008,(03)
- [20] 吴风. 网络传播学[M]. 北京: 中国广播电视出版社, 2004,6
- [21] 匡文波. 网络传播理论与技术[M]. 北京: 中国人民大学出版社, 2007,11.

- [22] 喻国明, 李莹, 李彪. 基于神经科学的传播学研究——工具、路径与研究框架[J]. 国际新闻界, 2008,(07)
- [23] 王天意. 网络舆论引导与和谐论坛建设[M]. 北京: 人民出版社, 2008, 8.
- [24] 燕道成, 蔡骥. 国外网络舆论管理及启示[J]. 当代传播, 2007,(02)
- [25] 肖海林, 邓敏艺, 孔令江, 刘慕仁. 元胞自动机舆论模型中人员移动对传播的影响[J]. 系统工程学报, 20(2005), 3: 225-231.
- [26] 刘常昱, 胡晓峰, 司光亚, 罗批. 给予小世界网络的舆论传播模型研究. 系统仿真学报, 18(2006), 12: 3608-3610
- [27] 刘常昱, 胡晓峰, 罗批, 司马亚. 基于不对称人际影响的舆论涌现模型研究[J]. 系统仿真学报, 20(2008), 4: 990-996
- [28]. Stephen P. Borgatti, Martin G. Everett. Models of core/periphery structures. Social Networks, 21, 4, (2000) 375-395
- [29]. Achim Oberg, Peter Walgengach. Hierarchical structures of communication in a network organization. Scandinavian Journal of Management, 24, 3, (2008) 183-198
- [30]. Schahram Dustdar, Thomas Hoffmann. Interaction pattern detection in process oriented information system. Data & Knowledge Engineering, 62, 1, (2007) 138-155
- [31]. Xin Tong, Reid Lifset. International copper flow network: A blockmodel analysis. Ecological Economics, 61, 2-3, (2007) 345-354
- [32]. Wang Jyun-Cheng, Chiu Chui-Chen. Recommending trusted online auction sellers using social network analysis. Expert Systems with Applications, 34, 3, (2008) 1666-1679
- [33]. Cai Ya-li, Wang Wen-dong, Gong Xiang-yang, Li Yu-hong, Chen Can-feng, Ma Jian. Mobile E-commerce model based on social network analysis. The Journal of China Universities of Posts and Telecommunications, 15, (suppl 1), 2008, 19-83, 97
- [34] Garry Robins, Pip Pattison, Yuval Kalish, Dean Lusher. An introduction to exponential random graph (p^*) models for social networks. Social networks 29, 2, (2007) 173-191
- [35] Jose Luis Ortega, Isidro F. Aguillo. Mapping world-class university on the web. Information Processing and Management(2008), doi:10.1016/j.ipm.2008.10.001
- [36] Stanley Wasserman, Katherine Faust. Social network analysis: Methods and Applications[M]. Cambridge University Press, 1994.
- [37] Peter J. Carrington, John Scott, Stanley Wasserman. Models and Methods in Social Network Analysis[M]. Cambridge University Press, 2005.
- [38] Jennifer J. Xu, Hsinchun Chen. Fighting organized crimes: using shortest-path algorithms to identify associations in criminal networks. Decision Support Systems, Volume 38, Issue 3, December 2004, Pages 473-487.

- [39] Stolfo,S,J. Hershkop, S. Ke Wang, Nimeskern, O. EMT/MET: systems for modeling and detecting errant email. DARPA Information Survivability Conference and Exposition. 2003. Proceedings. 2003(2) 290-295.
- [40] Dunham, M,H. Data Mining:Introductory and Advanced Topics. Prentice Hall, 2002.
- [41] Roiger.R, Geatz.M. Data Mining: A Tutorial Based Primer. Addison-Wesley, 2002.
- [42] Agrawal.R, Imielinski.T, Swami.A. Database mining: A performance perspective. IEEE Transactions on Knowledge and Data Engineering, 1993,5:914-925,.
- [43] Agrawal.R, Imielinski.T, Swami.A. Mining association rules between sets of items in large databases. In Proc. ACM SIGMOD Intl. Conf. Management of Data, 1993, 207-216
- [44] Dunkel.B, Soparkar.D. Data Organization and Access for Efficient Data Mining. In Proc. Of the 15th Intl. Conf. on Data Engineering, 1999,522-529
- [45] Han.J, Pei.J, Yin.Y. Mining Frequent Patterns without Candidate Generation. In Proc. ACM-SIGMOD int. Conf. on Management of Data. 2000. 1-12.
- [46] MacQueen. J. Some methods for classification and analysis of multivariate observations. In proc.of the 5th Berkeley Symp. on Mathematical Statistics and Probability, 1967,281-297
- [47] Fisher.D, Iterative Optimization and Simplification of Hierarchical Clusterings. Journal of Artificial Intelligence Research, 1996, 4:147-179
- [48] Sneath.P.H.A, Sokal.R.R. Numerical Taxonomy. Freeman, San Francisco, 1971.
- [49] Ester.M, Kriegel.H.P, Sander.J, Xu.X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Knowledge Discovery and Data Mining,1996 August, 226-231
- [50] Sander.J, Ester.M, Kriegel.H.P, Xu.X.Density-based Clustering in Spatial Database: the algorithm GDBSCAN and its Applications. Data Mining and Knowledge Discovery,1998(2):169-194

附 录

A. 作者在攻读学位期间发表的论文目录:

[1] 周明, 庄经纬. 社会网络分析方法在网络论坛舆论管理中的应用. 技术经济, 2009.11:93-98.

B. 部分图表:

表 b.1 前 50 个帖子的编号

Table b.1 The number of top 50 posts				
前 50 个帖子的编号				
472239	468783	466976	474346	476812
468275	465955	474170	482229	476278
468360	465223	477820	474789	473981
466962	470170	476206	477166	476875
472434	471685	474520	477376	473594
468944	465525	474843	475294	475537
465667	469056	474816	474362	475746
472712	465201	479944	480581	477522
467477	473124	476079	478188	480294
468755	473070	478885	474879	480176

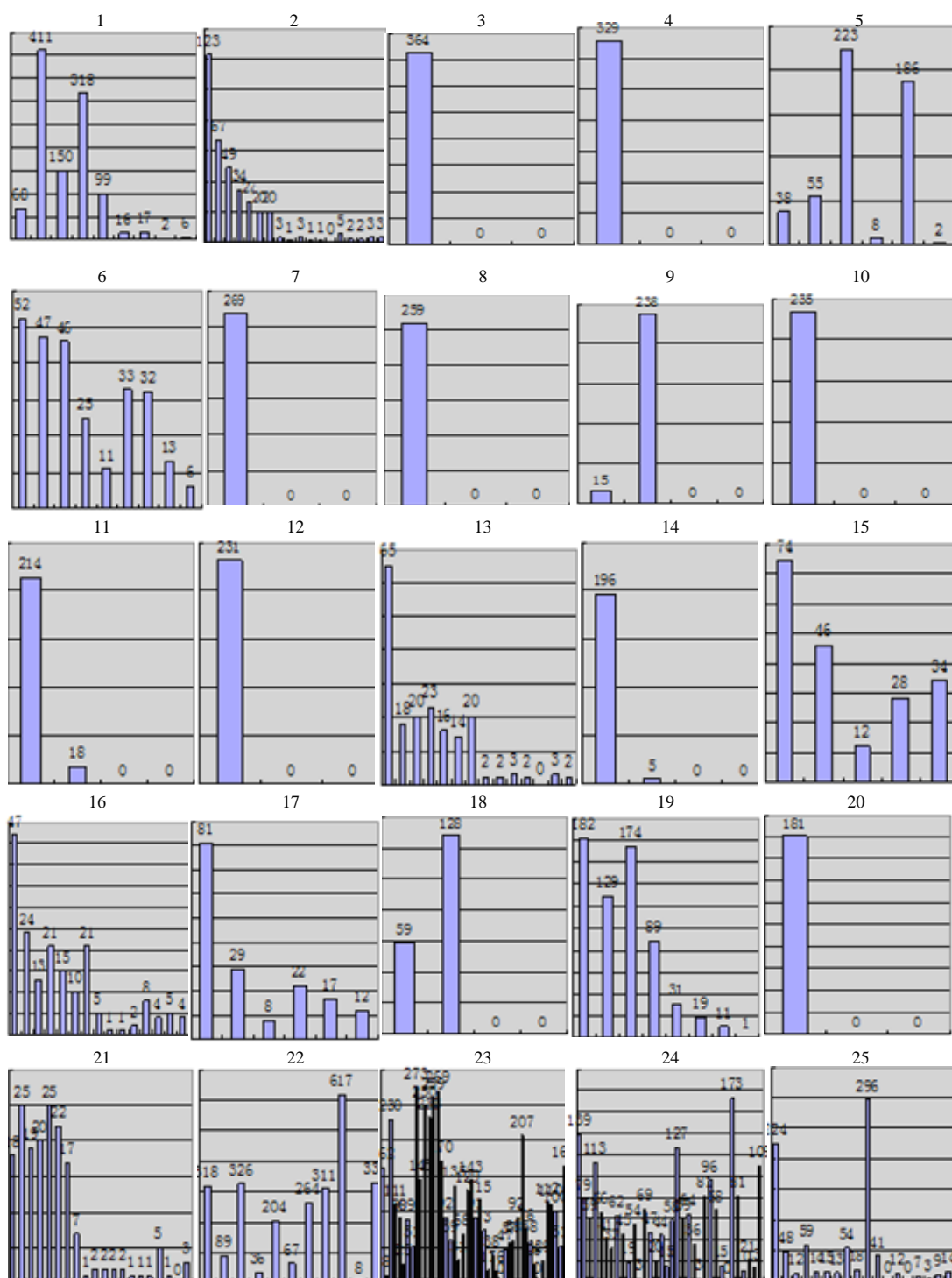


图 b.1 50 个帖子每天回帖量柱形图

Fig.b.1 Bar diagram of quantity of reply of top 50 posts

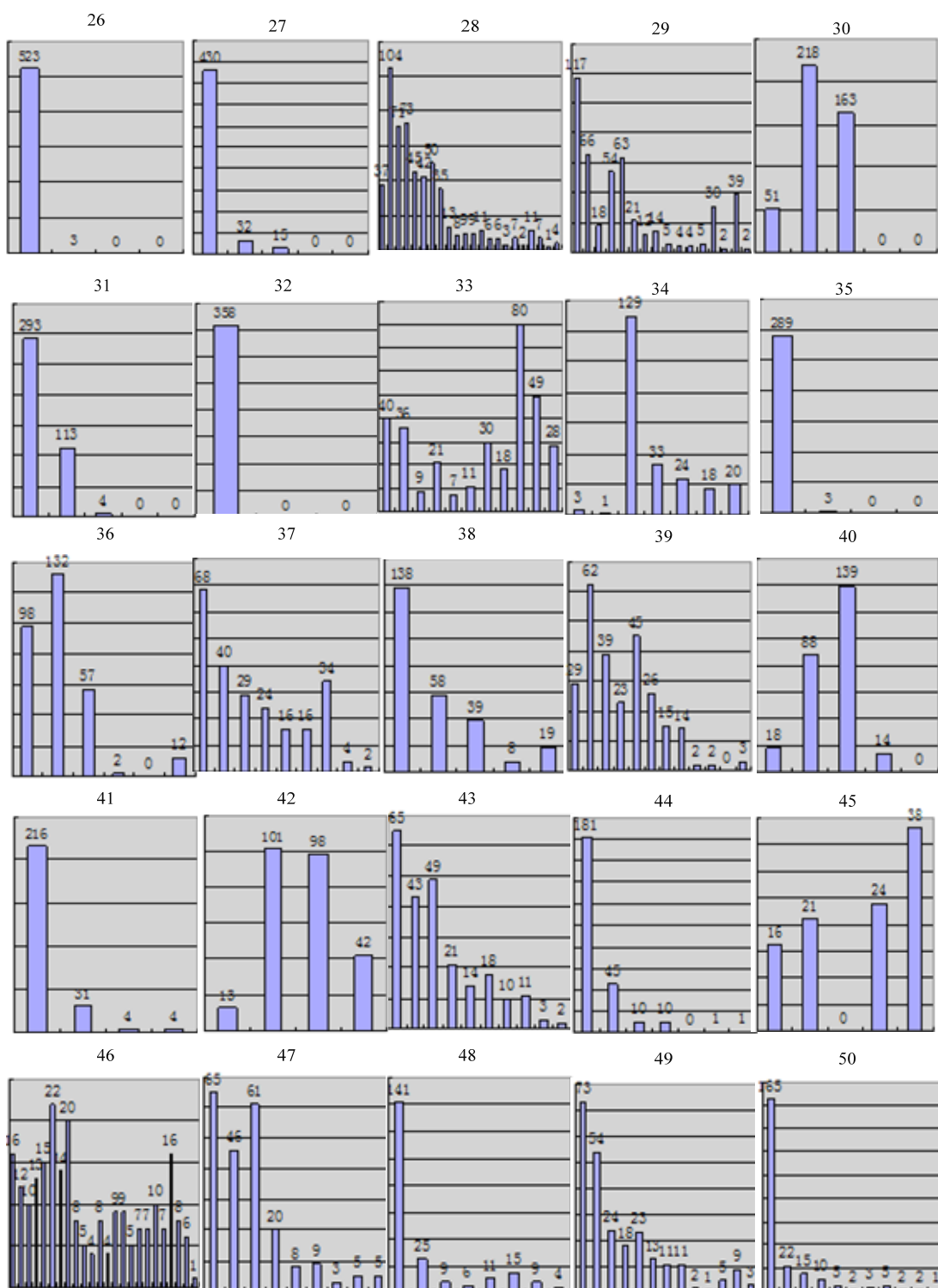


图 b.2 50 个帖子每天回帖量柱形图（续）

Fig.b.2 Bar diagram of quantity of reply of top 50 posts (extended)

表 b.2 50 个帖子特征值

Table.b.2 The eigenvalue of top 50 posts							
帖子 顺序号	持续天数	峰值所在 天数	第二峰所 在天数	帖子 顺序号	持续天数	峰值所在 天数	第二峰所 在天数
1	9	2	4	26	2	1	2
2	18	1	2	27	3	1	2
3	1	1	1	28	22	2	4
4	1	1	1	29	16	1	2
5	6	3	5	30	3	2	3
6	9	1	2	31	3	2	1
7	1	1	1	32	1	1	1
8	1	1	1	33	11	10	1
9	2	2	1	34	7	3	4
10	1	1	1	35	2	1	2
11	2	1	2	36	6	2	1
12	1	1	1	37	10	1	2
13	14	1	4	38	5	1	2
14	2	1	2	39	12	2	5
15	5	1	2	40	4	3	2
16	15	1	2	41	4	1	2
17	6	1	2	42	4	2	3
18	2	2	1	43	10	1	3
19	8	1	3	44	7	1	2
20	1	1	1	45	5	5	4
21	20	2	5	46	24	6	8
22	11	9	11	47	9	1	3
23	44	9	14	48	8	1	2
24	34	29	1	49	13	1	2
25	18	10	1	50	11	1	2

表 b.3 50 个帖子所在板块和主题内容

Table.b.3 The plate and main idea of the top 50 posts					
帖子号	板块	主题内容	帖子号	板块	主题内容
1	黄桷树下	演讲报告	26	黄桷树下	水贴
2	心语馨愿	贴图	27	黄桷树下	水贴
3	黄桷树下	水贴	28	心语馨愿	爱情
4	黄桷树下	水贴	29	黄桷树下	爱情
5	视点	演讲报告	30	黄桷树下	家乐福事件
6	心语馨愿	爱情	31	黄桷树下	水贴
7	黄桷树下	水贴	32	黄桷树下	水贴
8	黄桷树下	爱情	33	黄桷树下	爱情
9	黄桷树下	电视情感	34	激情体育	球赛贴图
10	黄桷树下	水贴	35	黄桷树下	水贴
11	黄桷树下	水贴	36	黄桷树下	水贴
12	黄桷树下	水贴	37	心语馨愿	爱情
13	爱国家	战争	38	回收站	贴图
14	黄桷树下	爱情	39	心语馨愿	爱情
15	心语馨愿	爱情	40	黄桷树下	爱情
16	心语馨愿	爱情	41	好摄之徒	运动会贴图
17	视点	食堂问题	42	黄桷树下	水贴
18	黄桷树下	水贴	43	心语馨愿	爱情
19	视点	演讲报告	44	重大影像	运动会贴图
20	黄桷树下	水贴	45	轻松一刻	贴图
21	心语馨愿	爱情	46	鱼游天下	贴图
22	黄桷树下	水贴	47	机械雄风	贴图
23	黄桷树下	水贴	48	黄桷树下	贴图
24	08 奥运	火炬传递	49	网络张贴栏	饭馆问题
25	回收站	爱情	50	鱼游天下	贴图



知网查重限时 7折 最高可优惠 120元

本科定稿，硕博定稿，查重结果与学校一致

立即检测

免费论文查重: <http://www.paperyy.com>

3亿免费文献下载: <http://www.ixueshu.com>

超值论文自动降重: http://www.paperyy.com/reduce_repetition

PPT免费模版下载: <http://ppt.ixueshu.com>
