

# 基于凝聚子群的虚拟社区社会网络结构实证分析

周明,张科

(重庆大学 经济与工商管理学院,重庆 400044)

**摘要:**网络中的虚拟社区是现代舆论传播最为自由的空间,良好的社区交流利于舆论的有序发展,反之,如果社区中出现欺诈或不文明情况时,将严重影响网络舆论的有效发展。因此对网络虚拟社区进行有效管理对国家经济政治文明的发展具有重大的影响。而要对网络虚拟社区进行有效的管理,首先必须了解网络虚拟社区中人员交流的特征。本文采用网络挖掘和社会网络分析技术相结合的方法,对虚拟社区的典型形式之一——网络论坛中的社区模式及成员间交流的特征进行了凝聚子群分析,并对划分出的子群进行凝聚性检验。得出结论:网络论坛中成员间通过交流形成不同的虚拟社区,社区内部成员具有明显的凝聚性,社区间的成员联系则较少。

**关键词:**网络舆论;网络挖掘;网络社区;凝聚子群;网络结构

**中图分类号:**TP274.2 **文献标识码:**A **文章编号:**1002-980X(2009)10-0080-07

网络的快速发展使网络虚拟社区成为现代舆论形成的主要阵地。而网络虚拟社区中传播主体的匿名性、传播的及时性等新特征使传统舆论管理的方法和手段失效,即政府无法对网络舆论实施及时、有效的管理,从而出现网络信息失真、网络诈骗及网络信息侵犯个人隐私等情况,在不利于网络舆论良好发展的同时,损坏政府的公信力,最终严重影响政府和民众之间的良好沟通,导致网络舆论突发事件产生,给社会经济和政治民主的发展及个人的发展带来了极大的负面影响。因此,要真正发挥网络舆论的力量,使其能为社会经济发展和政治民主做出更大的贡献,必须寻找有效的网络舆论管理手段填补网络舆论管理的真空。

为此,伴随着网络挖掘技术和社会网络分析方法技术的迅速发展,越来越多的学者对这一领域进行了研究和探讨。在国外,Wellman 探讨了人们生活中的私人社区<sup>[1]</sup>;Jeffrey Heer 等探讨了互联网上社会网络的构成<sup>[2]</sup>;Michael Chau 等利用社会网络分析探讨了博客空间的虚拟社区<sup>[3]</sup>;A. Arenas 等对社区的结构进行了分析和探讨<sup>[4]</sup>。在国内,郭绍忠等基于邮件联系提出了社会网络的构建方法,并借此探讨社团的挖掘算法的改进和优化<sup>[5]</sup>;梅中玲利用 Web 信息挖掘,发现网络舆情、分析网络舆情的起源及网络舆情受众<sup>[6]</sup>;郝小花提出了基于 Web 数据挖掘技术的用户社区聚类方法和相关量化技术<sup>[7]</sup>。

但是,以上学者在进行探讨时,并未将网络挖掘

和社会网络分析的技术分析最终利用到网络舆论的管理中来。以上研究主要是基于网络这一客体进行的技术上的研究,缺乏对网络虚拟社区中成员间行为的研究。鉴于此,本文借助网络挖掘技术和社会网络分析技术,将虚拟社区中最为普遍的网络公共论坛作为整个研究的平台,对网络虚拟社区中的公共论坛中的成员之间的关系进行分析,从而了解整个公共论坛中成员的行为,便于政府对网络舆论进行管理,引导网络舆论向良好方向发展。本文的主要思路是:首先从公共论坛中提取成员的基本属性数据及其相互间的交流的数据,并将其转换为适合于社会网络分析的数据形式;然后利用现有的社会网络分析技术,分析网络论坛中各个成员间的交流活动,探讨整个网络论坛中社区模式及成员间交流的特征,从而为网络公共论坛中的网络舆论管理提供依据,将网络舆论置于政府的有效管理下,并更大程度地发挥网络舆论对社会政治和经济的有利作用。

## 1 研究方法

整个研究过程主要分为两步:首先利用 Web Mining 技术对网络论坛中的数据进行提取和转换等,为社会网络分析准备相应数据;然后对转换后的数据进行社会网络分析,探讨虚拟社区中的社会网络结构。

在数据准备阶段,主要是从网络公共论坛中提取网络论坛中各个成员的属性数据及相互间交流的

收稿日期:2009-09-22

基金项目:重庆市自然科学基金项目(csct,2008BB2041);重庆市哲学社会科学规划项目(2008-zh11)

作者简介:周明(1958—),四川乐山人,重庆大学经济与工商管理学院教授,研究方向:信息管理与信息系统、决策支持系统;张科(1986—),男,云南昭通人,重庆大学经济与工商管理学院硕士研究生,研究方向:信息管理信息系统。

数据,然后对提取的数据进行处理,转换为适合于社会网络分析的结构性变量(structural variable)和组件性变量(composition variable)数据。

进行社会网络分析时,主要从以下方面进行分析:整个社会网络中各个成员威望度的计算;根据各个成员的威望度选择要进行子网探测的成员;

对构建的新社会网络进行无向化和无值化处理;对处理后的社会网络进行子网探测,主要的方法有集团化(cliques)子网探测和半径为  $n$  的集团化子( $n$ -cliques)网探测;对探测出的子网逐个进行凝聚力分析。整个研究的流程如图 1 所示。

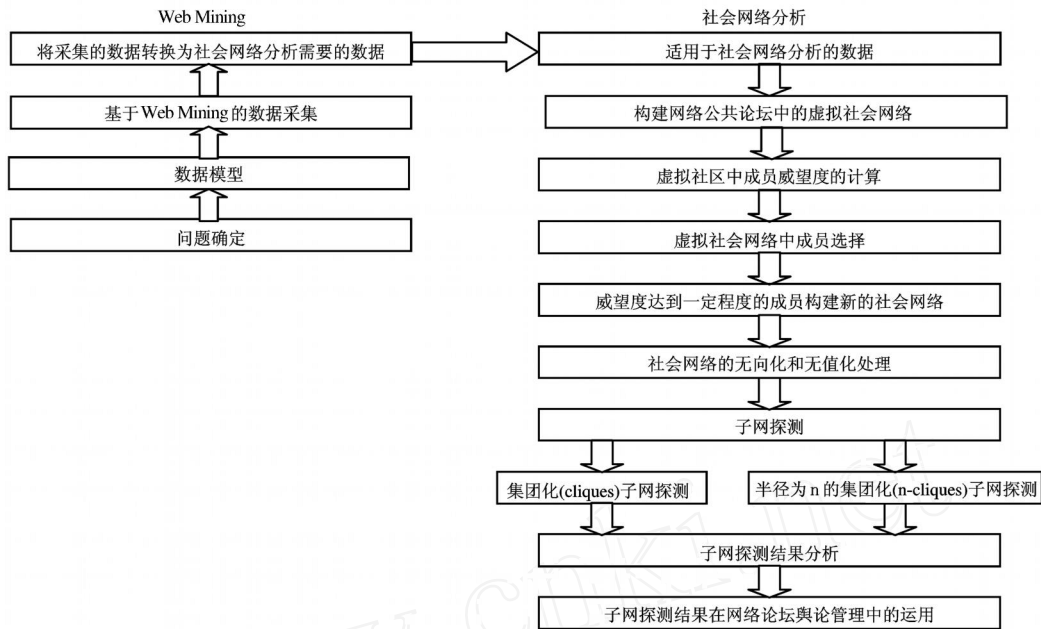


图 1 基于 Web Mining 的虚拟社区社会网络结构分析框架图

1.1 数据模型

在进行数据准备之前,首先设计好社会网络分析需要的基本数据的属性和类型,构建合适的数据模型,即首先必须选择好需要准备的数据的属性:在这里数据模型中的数据属性主要包括两类信息:成员基本信息,主要包括成员状态、性别、学院、学

历等,如表 1 所示;成员在论坛中的交流信息,主要有发帖被跟数、跟帖被跟数、共同跟帖成员、跟其帖子成员等,而此类信息又可以分为两部分,即发帖的基本情况和跟帖的基本情况,分别如表 2 和表 3 所示。

表 1 论坛成员的基本信息表

属性	类型	属性说明	属性	类型	属性说明
成员编号	Normal	论坛中成员的标识	学院	Normal	成员所在的学院
成员状态	Binary	成员是实名用户还是匿名用户	学历	Normal	成员注册时的学历
性别	Binary	成员注册时所填的性别	年级	Normal	成员交流时所处的年级

表 2 发帖基本情况表

属性	类型	属性说明	属性	类型	属性说明
发帖人编号	Normal	发帖人对应的成员编号	发帖时间	Normal	帖子发出的时间
帖子编号	Normal	为了记录而定的帖子的编号	发帖内容	Normal	帖子的主要内容

表 3 跟帖基本情况表

属性	类型	属性说明	属性	类型	属性说明
所跟帖子编号	Normal	所跟帖子的记录编号	跟帖时间	Normal	跟帖的时间
跟帖人编号	Normal	跟帖人对应的成员编号	跟帖内容	Normal	跟帖的主要内容

在以上数据模型的基础上,要进行社会网络分析,还必须将以上数据模型转换为适合社会网络分

析的数据模型。在社会网络分析中,主要包括成员编号、成员间交流的次数,如表 4 所示。

1.2 数据准备和转换

依据以上的数据模型提供的数据属性,结合网站运行的备份文件,从网站的后台数据库中提取相关属性的数据。社会网络分析接受两类数据变量,即结构性变量和组件性变量。组件性变量主要用于衡量各个成员的性质,可以直接提取得到各个成员的属性数据;社会网络分析中的结构性变量主要表示成员间交流的信息,主要通过统计得到。

表 4 各个成员间交流次数表

属性	类型	属性说明
成员编号 1	Normal	论坛中成员的标识
成员编号 2	Normal	与前一个成员有交流的成员的标号
两成员间的交流数	Normal	以上两个成员的交流数,主要通过帖子的编号识别两个成员的交流

由于 Web Mining 提取的数据是一条条的记录,而进行社会网络分析时需要的是各个成员之间联系的矩阵,所以必须将采集的数据转换为适合社会网络分析的社会关系矩阵。其具体步骤是:以帖子的编号作为主关键词将发帖基本情况表和跟帖基本情况表结合起来,得到各个成员间发帖和跟帖的情况;依据步骤的结果统计每两个成员之间联系的次数;将所有成员之间的联系统计出来,构造其相邻矩阵,矩阵中的数值即为每两个成员之间联系的次数。

1.3 数据选择

根据 1.2 节所述,通过以上 Web Mining 实现数据提取和数据转换后,可得到一个节点之间联系的矩阵 A,其值  $a_{ij}$  表示第  $i$  个成员对成员  $j$  的帖子的跟帖数量。但是在这个网络中,由于节点的数目太大,直接进行社会网络分析是不可能的,所以必须从整个社会网络中选择具有一定代表性的成员进行研究。

在一个社会网络中,扮演重要角色的节点一定与更多的其他节点发生联系,所以衡量节点的重要性时常用在整个社会网络中与其联系的节点的多少即节点的集中度来衡量。当社会网络矩阵 A 是一个有向的社会网络时,节点之间的联系存在着方向上的差别,即节点  $i$  和节点  $j$  的联系存在着进出之分,此时节点的集中度有出集中度和入集中度之分。将第  $i$  个成员的出集中度用  $C_{OD}(n_i)$  表示,入集中度用  $C_{ID}(n_i)$  表示<sup>[8]</sup>。则:

$$C_{OD}(n_i) = \frac{\sum_{j=1}^n a_{ij}}{n}; \quad (1)$$

$$C_{ID}(n_i) = \frac{\sum_{j=1}^n a_{ji}}{n} \quad (2)$$

通过以上集中度的计算,就可以对网络公共论坛中成员进行选择,将集中度小于一定程度的成员删除,同时也将这些成员的联系从整个社会网络中删除,最终形成一个由集中度达到一定程度的成员组成的虚拟社会网络。

1.4 数据预处理

在进行子网探测时,由于大部分的子网探测方法都是对无向和 0-1 的社会关系矩阵进行子网的探测,而依据以上社会网络分析方法得到的社会网络是带方向的、带值的社会关系矩阵,所以必须对社会网络进行无向化和无值化处理。

假设通过以上的办法得到一个社会关系矩阵 A,其中  $a_{ij}$  表示第  $i$  个成员对成员  $j$  的帖子跟帖的数量。以下介绍无向化处理和无值化处理。

1) 无向化处理。

矩阵中  $a_{ij}$  表示第  $i$  个成员对成员  $j$  的帖子跟帖的数量,  $a_{ji}$  表示第  $j$  个成员对成员  $i$  的帖子跟帖的数量。这两个联系都可以反映节点  $i$  和  $j$  之间的联系,为了将二者共同考虑进社会网络分析中来,并形成无向的对称矩阵,常采用式(3)进行处理。

$$b_{ij} = b_{ji} = \min(a_{ij}, a_{ji}) \quad (3)$$

2) 无值化处理

为了把矩阵 A 转换成一个只表示节点之间是否有联系的布尔矩阵,首先必须根据矩阵 A 的取值情况,确定一个节点之间联系强度的值 C,然后利用式(4)进行转换。

$$x_{ij} = x_{ji} = \begin{cases} 1 & \text{if } b_{ij} = b_{ji} \quad C \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

1.5 子网探测

通过以上社会关系矩阵的无向化和无值化的处理,将得到一个表示各个成员之间是否存在联系的社会关系矩阵。以下将依据这个矩阵探讨整个社会网络中虚拟社区的子网探测。

网络论坛中的虚拟社区是指由网络公共论坛中某些具有共同兴趣的成员构成的团体,在这个团体中,各个成员之间拥有共同的兴趣或同时被卷入某一事件。探讨整个网络论坛中的虚拟社区,就是从构建的社会网络分析中探讨具有某一相同特征或属性的子网络。为此,以下主要介绍社会网络分析中的子网探测。

1.5.1 集团化 (cliques) 和半径为 n 的集团化 (nr cliques) 子网探测

### 1) 基本思路。

在整个虚拟社会网络分析中,最核心的部分是探讨整个社会网络中的子虚拟社区,即社会网络的子网探测。社会网络分析中的子网探测的方法很多,最主要的也是最基本的是集团化(cliques)子网探测。其基本思路是寻找满足以下条件的一些子网<sup>[9]</sup>:

子网内部各个节点之间都存在着强力的联系,即节点之间都是直接联系的;

子网外部没有节点与子网内部所有节点都存在直接联系或节点间的距离不超过规定的长度。

### 2) 基本步骤。

步骤 1:建立各个节点之间距离不大于  $n$  的节点之间联系布尔矩阵  $A$ 。

步骤 2:产生初始子网。Cliques 中子网的节点个数不得小于 3,因此最先寻找节点数为 3 的子网。

寻找出各个节点之间的联系,即矩阵中值为 1 的两个节点,得到所有节点联系的集合  $C_2$ ;

在两个节点存在联系的基础上,寻找与这两个节点共同存在联系的点。主要是利用以上节点联系的集合,将集合中各个元素中的前缀即有联系的第一个节点相等的元素作为三节点最大团的第一个元素,然后看剩余的两个节点是否存在联系,如果有,则形成新的 3 节点最大团。计算完毕后,将得到一个拥有三个节点的最大团的集合  $C_3$ ,即初始子网的集合。

步骤 3:产生  $K$  节点的最大团子网的集合。在此主要是采用  $C_{k-1} \times C_{k-1}$  方法产生节点为  $K$  的最大团子网的集合。

假设  $K-1$  节点的最大团的集合  $C_{k-1}$  的各个最大团是一个事件,每一个事件含有  $K-1$  个节点,设  $A$  和  $B$  是两个事件,其中  $A \in C_{k-1}, B \in C_{k-1}$ ,当两个事件满足条件  $a_i = b_i (i = 1, 2, \dots, k)$  并且  $X_{a_{k-1}b_{k-1}} = 1$  时,将得到一个节点的  $K$  节点的最大团  $D, D = \{a_1, a_2, \dots, a_{k-1}, b_{k-1}\}$ 。

停止条件:满足以下条件之一时停止寻找最大团:

$k$  个节点的最大团子网中不存在有相同  $k-1$  个节点相同;

$k$  个节点的最大团子网中存在有相同  $k-1$  个节点,但剩余的节点之间不存在联系。

### 1.5.2 子网探测结果分析

根据以上子网探测办法将整个社会网络划分为不同的多个子网络,在社会网络分析中,主要是利用子网的凝聚力对子网的性能进行分析。即对于一个子社会网络,如果子社会网络内部的节点之间的联

系强度大于子网内部节点与其外部节点之间的联系强度,则说明这个子社会网络的凝聚力强。

基于此理念,假设整个社会网络为  $L$ ,包含有  $g$  个节点,矩阵  $X$  表示  $g$  个节点之间的联系。先存在一个探讨出的子网  $L_s$ ,包含  $g_s$  个节点,则这个子网络的凝聚力的计算公式为:

$$\frac{\sum_{i \in N_s} \sum_{j \in N_s} x_{ij} / g_s (g_s - 1)}{\sum_{i \in N_s} \sum_{j \in N_s} x_{ij} / g_s (g - g_s)} \quad (5)$$

如果这个比率大于 1,表示子网内部的联系强于子网与子网外的联系,则说明子网的凝聚力强,这样的子网划分利于网络舆论的管理;否则说明子网的凝聚力不强,这样的子网划分对网络舆论的管理没有多大的帮助。

## 2 实验展示

为了验证整个研究流程的可行性和可操作性,我们将利用网络论坛中一段时间内某一板块的成员间的交流信息数据进行试验。以下介绍整个试验的过程。

### 2.1 数据收集

根据以上数据模型描述,需要产生两方面的数据,即表示成员属性的数据和表示成员间交流的数据。而对于两种不同属性的数据采取不同的数据产生方式。

对于表示各个成员属性的数据,主要依据各个成员的注册信息得出;而成员间交流的信息主要由发帖基本情况表和跟帖基本情况表反映;成员间交流信息的数据主要是针对一个帖子进行的记录,即某一个帖子由谁发表、有哪些成员跟帖。据此,从论坛的后台备份数据库中得到需要的记录。我们从论坛的某一板块中选取 195 个成员在 2008 年 3 月至 9 月的交流信息进行实验,得到发帖和跟帖记录数据共 78956 条。

### 2.2 数据转换

通过以上方式得到的数据无法直接用于社会网络分析,必须进行处理和转换。

对于表示成员属性的数据,在社会网络分析中有一类数据变量——组件性变量与之对应,无需做很大的处理。

而表示成员间交流的数据,在产生时是用表格表示的,但是在社会网络分析中,数据信息通常是使用图形或社会关系矩阵来表示的,为此必须对原有的数据进行处理并转换。在此主要是利用 2.2 节所述的方法进行数据的转换。首先将 195 个成员之间的联系统计得到交流 553 对关系;然后将其转换为

邻接矩阵,得到一个 195 ×195 的社会关系矩阵,设为 A,则矩阵中的值  $a_{ij}$  表示第  $i$  个成员对成员  $j$  的帖子跟帖的数量。以下利用其进行社会网络分析。

2.3 数据选择

主要是选择对整个社会网络影响力大的成员,

表 5 筛选出的各个节点的出集中度和入集中度之和

节点	6	10	12	15	22	26	32	38	41	43
集中度	0.1615	0.0332	0.0245	0.0225	0.0105	0.0144	0.0137	0.1431	0.011	0.0150
节点	44	59	62	67	73	77	81	85	96	97
集中度	0.0149	0.0115	0.0348	0.0237	0.0149	0.0344	0.0104	0.0127	0.0632	0.0919
节点	102	105	113	116	120	122	134	135	147	151
集中度	0.0137	0.0123	0.0183	0.0169	0.0772	0.0143	0.0410	0.0516	0.0481	0.1054
节点	152	154	165	170	180	182	188	189		
集中度	0.0192	0.0186	0.0208	0.0317	0.153	0.0962	0.0566	0.0337		

通过以上集中度的计算,选取如表 5 所示的 38 个节点构建新的社会网络。为此,利用 1.3.2 节所述的方法将原有社会网络的其他成员及其相关的联系删除,最后形成一个新的社会网络,利用社会网络分析工具 Pejak 可以得到各个成员的交流信息图。

2.4 数据处理

通过以上成员的选择,构建了一个以网络论坛中成员为节点、成员间的发帖和跟帖为节点联系的社会网络。而这个社会网络是带值和带向的,因此必须无向化和无值化处理。在此,主要利用 1.4 节所述的方法进行无向化和无值化处理,得到节点间联系的对称矩阵。

2.5 子网探测

为了探讨整个网络公共论坛中的虚拟社区,即探讨由具有某一共同属性或共同爱好的成员所组成的社区,可利用社会网络分析中的子网探测进行。

并删除对社会网络影响小的成员,采用 1.3 节所述的成员威望度计算得出成员的出集中度和入集中度,然后将出集中度和入集中度相加,选取其和大于 0.01 的成员,共 38 个。如表 5 所示。

以下针对以上处理后的节点间联系矩阵,分别利用集团化和半径为  $n$  的集团化子网探测方法进行子网探测。

2.5.1 集团化(cliques)子网探测

利用 1.5.1 节所述的集团化子网探测的思路和方法进行子网探测。即寻找最大的子网,使得在这个子网内部的各个元素之间都存在直接联系。

依据子网探测方法的计算步骤,首先得到 419 个初始三节点的最大团,然后得到 403 个四节点的最大团、187 个五节点的最大团、39 个六节点的最大团和 3 个七节点的最大团。由于将这些所有的子网穷举出来会使子网的代表意义降低,所以必须寻找最大代表性的最大团,即利用最少的最大团代表所有的子网。综合后将得到最大团 18 个,但是节点 6 和 26 不满足任何最大团的定义,即其不属于任何最大团。集团化(Cliques)子网成员如表 6 所示。

表 6 集团化(cliques)和半径为 2 的集团化(2-cliques)子网探测结果

Clique	集团化(Cliques)子网成员							Clique	半径为 2 的集团化(2-cliques)子网成员													
	1	3	8	20	30	35	36		1	3	8	10	19	20	25	27	29	30	35	36	37	38
Clique1	1	3	8	20	30	35	36	Clique1	1	3	8	10	19	20	25	27	29	30	35	36	37	38
Clique2	8	15	16	27	36	37	38	Clique2	1	3	8	10	19	20	28	29	30	35	36	37	38	
Clique3	1	3	17	25	30	36		Clique3	1	3	14	19	20	25	27	29	30	35	36	37	38	
Clique4	3	8	20	24	30	35		Clique4	1	4	9	19	20	24	25	27	29	30	35	36		
Clique5	4	9	19	20	22	35		Clique5	1	3	8	11	19	20	24	25	27	32	36	38		
Clique6	8	15	16	27	36	38		Clique6	1	3	8	11	19	20	24	25	33	35	36	38		
Clique7	1	7	21	29	36			Clique8	1	3	18	19	20	25	27	29	30	35	36			
Clique8	1	4	19	20	34			Clique7	1	3	17	25	27	29	30	35	36	37	38			
Clique9	1	8	10	28	30			Clique9	5	8	15	16	27	29	30	35	36	37	38			
Clique10	2	13	30	35	36			Clique10	2	13	19	24	25	27	30	35	36	37	38			
Clique11	3	14	19	25	35			Clique11	1	7	25	27	29	30	35	36	37	38				
Clique12	12	19	25	27	35			Clique12	2	12	19	25	27	30	35	36	37	38				
Clique13	8	24	25	33	35			Clique13	1	3	8	10	19	23	35	36	37	38				
Clique14	8	25	27	32	36			Clique14	1	6	20	21	22	30	35	36						
Clique15	5	8	11	35				Clique15	5	8	11	31										
Clique16	1	7	23	36				Clique16	26	34	35											
Clique17	2	13	18	35																		
Clique18	1	11	31																			

### 2.5.2 半径为 $n$ 的集团化 ( $r$ -cliques) 子网探测

利用 1.5.1 节所述的半径为  $n$  的集团化 ( $r$ -cliques) 子网探测方法的思路和方法寻找子网,主要探测这样的子网:在子网内部任意两个节点之间的距离不超过  $n$ ,即任意一个子网的半径不得超过  $n$ 。在此利用  $n=2$  进行计算。

根据矩阵计算出所有节点之间距离少于 2 的布尔矩阵,然后利用最大团的基本方法和步骤计算出最大团,最终得到 1666 个 11 节点的最大团、360 个 12 节点的最大团、48 个 13 节点的最大团、3 个 14 节点的最大团等。同样,要利用最少的最大团代表所有的子网并且把所有的节点都划分到固定的最大团中,必须从现有的 2-cliques 子网中寻找最大代表性的最大团。归纳后得到 16 个最大团,半径为 2 的

集团化 (2-cliques) 子网成员如表 6 所示。

### 2.6 子网凝聚力检验

在进行子网探测后,将整个社会网络划分为不同的子网。一个子网内部的成员因为频繁的交流而拥有共同的兴趣或爱好,从而形成一个虚拟社区。但是,划分出来的社区是否好,即划分出的社区是不是真正能将社区内的成员交流与社区内部成员和其外部成员的交流划分开,这需要利用一定的方法进行检验。

子网凝聚力检验首先是利用式 (5) 计算出各个子网的凝聚力。集团化 (cliques) 子网和半径为 2 的集团化 (2-cliques) 子网的凝聚力分别如表 7 和表 8 所示。

表 7 集团化 (cliques) 子网中各子网的内部联系和其内部节点与外部节点的联系比率

Clique	Clique1	Clique2	Clique3	Clique4	Clique5	Clique6	Clique7	Clique8	Clique9
比率	2.17	2.9324	2.5946	2.2857	2.7429	3.0968	3.0556	5.8929	4.5833
Clique	Clique10	Clique11	Clique12	Clique13	Clique14	Clique15	Clique16	Clique17	Clique18
比率	1.65	2.1711	1.7935	2.1154	2.0625	2.8333	3.7778	2.9565	26.25

表 8  $r$ -cliques 中各子网的内部联系和其内部节点与外部节点的联系比率

Clique	Clique1	Clique2	Clique3	Clique4	Clique5	Clique6	Clique7	Clique8
比率	1.0633	1.1207	0.9789	1.1729	1.4444	1.3565	1.0919	0.911
Clique	Clique9	Clique10	Clique11	Clique12	Clique13	Clique14	Clique15	Clique16
比率	0.977	0.911	0.8235	0.8696	1.2389	1.0714	13.6	1.2209

然后根据计算出的凝聚力与 1 比较得出子网凝聚力的强弱。在表 7 中,所有的比率都是大于 1 的,所以这个子网探测出的子网的凝聚力都较强。但是在表 8 中,有的子网的比率小于 1,主要是这些子网内部的各个节点与子网外部其他节点之间的联系太多,导致其比率小于 1。

## 3 网络论坛舆论管理中的应用

通过以上的办法,可以将公共论坛中的任意一个成员划分到不同的社区中。对整个社会网络进行社区划分,可对网络舆论的管理提供以下方面的便利。

1) 为舆论引导提供依据。在网络舆论的引导中,最主要的是需要探讨整个网络舆论的发展方向,即要实现网络舆论的引导首先必须把握网络舆论的发展方向。而本文将整个网络划分为有限的几个子网,即探讨整个网络的虚拟社区结构。通过对整个虚拟社会网络中虚拟社区结构的探讨,将得到各个社会网络中各个虚拟社区的共同爱好或共同的行为模式,基于此可以预测各个虚拟社区在未来一定时间内的行为模式发展方向,从而了解整个虚拟社会

网络舆论的行为模式发展。利用此行为模式发展方向将可以预测网络中舆论的发展方向,从而为整个网络中舆论的引导提供依据。

2) 提高网络舆论管理的效率。目前政府无法对网络舆论实施及时、有效的管理的一个最主要的原因就是网络中的用户太多,同时鉴于人力、物力及技术方面的限制,也不可能对所有的用户进行跟踪管理。为此必须从庞大的网络用户中寻找对整个网络舆论有极大影响的用户,通过对这类用户管理实现整个网络舆论的管理。而本文主要论述从一个庞大的社会网络中筛出具有影响力的成员,然后据此构建新的社会网络,并探讨此社会网络中包含的不同社区,从而将整个虚拟的社会网络的管理落实到几个不同社区的管理,从而提高整个网络舆论的管理效率。

3) 使网络舆论管理及时、具有超前性。网络舆论管理较传统舆论管理的最大难处是网络的传播速度快、信息传播及时,因此,要实现网络舆论及时、有效地管理,必须能及时制止一些可能导致网络舆论问题的言论,而要实现这一点就必须对整个网络舆论的发展有一定的超前性认识,能提前预测网络

舆论发展的方向。本文研究思路的最核心的部分是将一个庞大的虚拟社会网络划分为几个不同的社区,而这些社区内部具有某一共同属性或某一共同特征,社区内部的成员的行为模式也具有某些程度上的相似。通过掌握不同社区的行为模式,可以对整个社会网络的行为模式有所了解,也可以对整个社会网络中网络舆论的发展进行一些预测,使得网络舆论的管理具有超前性。

## 4 结论

1) 采用基于 Web Mining 的社会网络分析方法,是适时、动态地对虚拟社区的社会网络结构进行分析的可行途径。

2) 对虚拟社区的社会网络结构进行分析,可以对网络舆论管理者提供帮助。

3) 子网探测及子网特征分析算法对模拟数据集是可行的,但是需要在后续研究中使用更大规模的实证数据集来验证算法的有效性。本项研究工作目前正在进行中,将通过实证研究验证并改进已有算法。

### 参考文献

[1] WELLMAN B. Men in network: private communication,

domestic friendships[J]. Men's Friendships, 1992, 9: 207-235.

[2] HEER J, BOYD D. Vizster: Visualizing Online Social Networks[D]. In Proceeding of the 2005 IEEE Symposium on Information Visualization (INFOVIS05), 2005: 253-258.

[3] Michael Chau, Jennifer Xu. Using Web Mining and Social Network Analysis to Study The Emergence of Cyber Communities In Blogs[M]. Integrated Series in Information Systems, Terrorism Informatics, 2008.

[4] ARENAS A, DANON L. Community analysis in social networks[J]. The European Physical Journal, 2004, 38: 373-380.

[5] 郭绍忠, 段丹. 邮件挖掘技术在社会网络分析中的研究与应用[J]. 计算机工程与设计, 2008, 29(9): 2339-2341.

[6] 梅中玲. 基于 Web 信息挖掘的网络舆情分析技术[J]. 中国人民公安大学学报: 自然科学版, 2007(4): 85-88.

[7] 郝小花, 邓小昭. 基于数据挖掘的可视化数字图书馆用户社区聚类与特征分析[J]. 情报科学, 2008(3): 396-399.

[8] WASSERMAN S, FAUST K. Social Network Analysis Methods and Applications [M]. Cambridge University Press, 2007.

[9] FREEMAN L C. The sociological concept of "group": an empirical test of two models[J]. American Journal of Sociology, 1992, 37: 3-24.

## Empirical Analysis on Social Network Structure within Virtue Community Based on Cohesive Subgroup

Zhou Ming, Zhang Ke

(College of Economics and Business Administration, Chongqing University, Chongqing 400044, China)

**Abstract:** The methods of traditional public opinion management are not fit for Web public opinion because of some characteristics of Web, for example, anonymity subject spread swiftly. However, the development of Web data mining and social network analysis make support to Web public opinion management. This paper, using the method of social network analysis, analyzes the community of members within the forum and the characteristic of members.

**Key words:** Web public opinion; Web mining; Web community; cohesive subgroup; network structure