

微博关注关系网络 K-核结构实证分析*

白林根 谌志群 王荣波 黄孝喜

(杭州电子科技大学认知与智能计算研究所 杭州 310018)

【摘要】为研究微博关注关系网络的特征,以新浪微博为例,引入复杂网络分析方法对微博关注关系网络进行实证分析。首先对微博关注关系网络进行 K-核分解,获取微博核心用户关系网络。然后计算 K-核网络基本参数、跟随比例和度相关性,发现其既具有在线社会网络的一般特征,也具有现实社会网络的一些特点。通过对 K-核网络的社区检测及节点中心性、互惠性、中间人角色的分析,发现其网络结构具有明显的社区特性。该研究能够为相关应用提供实证基础。

【关键词】微博关注关系 K-核分解 社区检测 复杂网络

【分类号】TP391

Empirical Analysis on K-core of Microblog Following Relationship Network

Bai Lin'gen Chen Zhiqun Wang Rongbo Huang Xiaoxi

(Institute of Cognitive and Intelligent Computing, Hangzhou Dianzi University, Hangzhou 310018, China)

【Abstract】In order to study the features of microblog following relationship network, the analysis method based on complex network is applied to analyze the following relationship of Sina Microblog in this paper. Firstly, the K-core decomposition operation is performed on a microblog following relationship network to obtain a core user's relationship network. Secondly, the features of online community network and those of realistic society network can be received in the K-core network by computing the basic parameters of the K-core network, following ratio and degree correlation. Finally, a conclusion can be obtained that the community characteristic of the network structure is very obvious by community detecting on the K-core network and analysis of the node centrality, reciprocity and the role of brokerage. Experimental results show that the research work of this paper can provide effectively a fundamental empirical analysis for related applications.

【Keywords】Microblog following relationship K-core decomposition Community detection Complex network

1 引言

近年来,社会媒体(Social Media)随着互联网应用的变革而迅猛发展,微博(Microblog)作为一种典型的社会媒体,目前已成为继网络新闻、网络论坛之后的又一大网络信息源^[1]。通过微博,用户可以实时生产、发布、转载和反馈信息。微博用户之间通过“关注”(Follow)操作建立联系,关注者成为被关注者的“粉丝”(Follower)。微博用户可以通过 Web、Wap 以及各种客户端组件发布信息,该用户的“粉丝”可以即时获取该信息并评论该信息,还可以通过转发让他的“粉丝”分享该信息。因此用户之间的关注关系是整个微博运行的基础。主流微博平台均拥有

收稿日期:2013-07-25

收修改稿日期:2013-10-12

* 本文系国家自然科学基金项目“引入涉身认知机制的汉语隐喻计算模型及其实现”(项目编号:61103101)、国家自然科学基金项目“基于马尔科夫树与 DRT 的汉语句群自动划分算法研究”(项目编号:61202281)和教育部人文社会科学研究项目“面向信息处理的汉语隐喻研究”(项目编号:10YJCZH052)的研究成果之一。

庞大用户群,用户之间的关注关系可以用一个规模巨大的图结构来描述,这个图结构具有与现实社会中人与人关系类似的特征,是一个“在线社会网络”(Online Social Network)^[2]。

复杂网络作为描述自然界和人类社会各种复杂系统的有力工具,目前已广泛应用于社会科学、生命科学和信息科学等领域,也是当前社会网络研究的主流方法^[3]。本文采用复杂网络分析技术对新浪微博关注关系网络的 K-核结构进行实证分析,对于揭示微博核心用户之间的关联关系特征和信息传播机制具有重要意义,研究成果有望在互联网舆情监控、信息检索、信息推荐等领域获得应用。

2 相关工作

在国外,采用复杂网络方法对社会网络进行分析已开展较多研究。Mislove 等^[4]在 2007 年首次对大规模社交网站 Flickr、YouTube 等进行研究,发现这些网站的用户关系网络具有小世界、无标度等复杂网络的典型特征。随着世界上第一个微博网站 Twitter 的出现及迅猛发展,社会网络研究的关注点转移到微博上。Java 等^[5]对早期的 Twitter 用户关系网络进行研究,发现该网络的度分布服从幂律分布,并且具有小世界特性以及较高的互惠性。Teutle^[6]研究了复杂网络及网络动力学理论在分析 Twitter 用户关系网络特性方面的可行性,并计算了 Twitter 用户关系网络的基本参数。Kwak 等^[7]对大规模 Twitter 用户的关系网络进行分析,发现其虽然符合小世界、无标度等复杂网络特征,但其节点的出度和入度并非严格遵循幂律分布,而且用户之间的连接是非对称的。

在国内,采用复杂网络方法研究社会网络也取得一定成果。胡海波等^[8]研究了一个大型在线职业人士社交网络的结构特性,发现与现实社会网络相比,它不仅表现出典型复杂网络特征,而且表现出连通子图规模分布、度分布和社团规模分布中所呈现的锯齿状以及度异配性等特性。余高辉等^[9]抽样构建了 QQ 群好友关系的复杂网络模型,并对其属性及动力机制进行分析,得出 QQ 群好友网络具有稀疏性、增长性与小世界性等特征的结论。

随着以新浪微博和腾讯微博为代表的国内微博平台的崛起,研究微博用户之间的关联关系在学术界与

工程界越来越受到重视。王晓光等^[10]以微博网站“Myspace9911”中的 54 位名人微博为研究样本,通过核心-边缘分析和聚类分析,研究了微博社区用户交流网络结构,界定了核心与边缘区域,发现了核心区域和聚类群组的联系。袁毅等^[11]随机抽取了新浪博客上的一篇博文,跟踪该博文引发用户交流而形成的各种关系网络(包括关注网络),并比较和分析其不同的结构形态及其交流特征。何黎等^[12]研究了由 102 个微博用户构成的关注关系网络,并用 UCINET 软件实现了对核心用户的挖掘操作。Kang 等^[13]研究了新浪的用户网络结构,发现其具有小世界、无标度和高聚类性,是一个典型的复杂网络。田占伟等^[14]发现新浪微博用户网络中介数中心点对信息传播效率的贡献显著,节点在信息传播中的作用差异很大,易于形成意见领袖,同时发现如果按行业属性对网络进行群落划分,各群落在微博中表现出的信息传播特性与在现实社会中相似。Fan 等^[15]研究了新浪的网络结构,给出其入度和出度分布,发现其满足幂律分布。Guo 等^[2]发现新浪微博用户关注关系网络中,大部分关联是单向的,且有一个结合紧密的核心网络,其半径相对于其他社会网络来说更短,说明微博用户之间联系更紧密。

以上研究中用户关系网络的规模都是有限的,并且没有过去噪与科学抽样处理。而主流的微博平台其用户数以亿计,随机抽取其中少量用户构造网络并进行研究,难以探究微博用户关系的普遍规律。本文以包含数百万用户的新浪微博数据集为基础,首先通过 K-核分解获取核心用户关注关系网络,然后采用复杂网络分析技术对其进行实证分析,计算获得了重要的复杂网络参数,揭示了微博核心用户之间关联关系的宏观规律和内在机制。

3 关注关系网络构建及 K-核分解

本文以新浪微博为研究对象,原始数据来自微博爬盟^[16]。微博爬盟是一个互联网公益组织,通过众包方式为高校学生和科研人员抓取所需要的互联网数据,目前以抓取微博数据为主,包括博主信息和博主的关注关系。通过微博爬盟获取了一个包含 N 个用户信息及其关注关系的数据集, $N \approx 2.2 \times 10^6$ 。以用户为节点,用户之间的关注关系为边(若用户 A 关注了用户

B 则从节点 A 到节点 B 划一条有向边) ,构建了一个用户关注关系网络,称为初始网络。微博中有大量的非活跃用户,尤其是那些“粉丝”少的用户,他们在微博中活跃度不高,对微博平台的贡献不大。为去除这些不活跃的非核心用户,采用 K-核分解对初始网络进行处理。

网络的 K-核是指网络中节点度值大于或等于 K 的节点所组成的网络,K-核分解是指把网络中节点度值小于 K 的节点去除的过程^[3]。在网络中,1-核网络就是初始网络,(K+1)-核网络是 K-核网络的子网络。针对无向图的 K-核分解过程如图 1 所示:

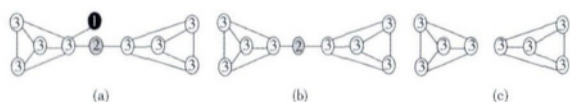


图 1 K-核分解过程

其中图 1(a) 是一个 1-核网络(初始网络),图 1(b) 是经过 2-核分解后得到的 2-核网络,图 1(c) 是经过 3-核分解后得到的 3-核网络。

有向网络有三种不同的 K-核分解形式:把有向图当作无向图处理,即不考虑边的方向;只考虑入度,称为入度分解;只考虑出度,称为出度分解。将包含 N 个节点的初始网络输入 Pajek 软件,对网络进行入度 30-核分解,得到 30-核网络,记为 K30。然后依次获得 K40、K45、K50、K55、K60、K65。本文主要考察这些 K-核网络。

4 K-核网络基本特征

4.1 K-核的基本参数

从节点数、边数、密度、平均度、连通块数等几个方面考察 K-核网络,参数计算结果如表 1 所示:

表 1 K-核基本参数

| K | 节点数 | 边数 | 密度 | 平均度 | 连通块数 |
|-----|--------|---------|-------|-------|------|
| K30 | 15 397 | 490 543 | 0.002 | 63.7 | 46 |
| K40 | 12 725 | 462 007 | 0.003 | 72.6 | 28 |
| K45 | 9 710 | 404 534 | 0.004 | 83.3 | 19 |
| K50 | 7 022 | 327 360 | 0.006 | 93.2 | 20 |
| K55 | 4 664 | 249 922 | 0.011 | 107.2 | 21 |
| K60 | 3 156 | 188 841 | 0.019 | 119.7 | 14 |
| K65 | 2 161 | 136 067 | 0.029 | 125.9 | 17 |

从表 1 可以看到,节点数和边数随着 K 值的增大而递减,因为 K 值越大,被去除的节点和边也就越多,网络规模也就越小。节点的平均度随着 K 值的增大而

递增,因为 K-核分解是把节点度值小于 K 的节点去除的过程,K 值越大,K-核分解后剩下的节点的度数必然越大。网络的密度是指网络中实际存在的边的数量与可能存在的最大边数之间的比值,实际边数越接近完全图的边数,网络的密度就越大,反之则越小。K 值越大,网络密度越大,即网络中节点之间的关系越紧密。但最大的密度值为 0.029,说明网络还是稀疏的,这也是社会网络的共同特点。连通块即网络的连通子图,连通块数的变化没有呈现与 K 值的严格相关性,这种现象可以解释为在 K-核分解中,随着 K 值的增大可能会使一个大的连通子图分解为多个小的连通子图,也可能使得一些连通子图完全消失。

4.2 跟随比例

在微博中,有的用户主要是分享博文信息,有的主要是获取博文信息,可用跟随比例^[6]来描述用户的这类行为。跟随比例是指用户的粉丝数与关注数的比例,公式如下所示:

$$FR = \frac{\text{Followers}}{\text{Following}} \quad (1)$$

当 $FR \ll 1$ 时,用户在微博中主要充当信息获取者的角色; $FR \gg 1$ 时,用户在微博中主要充当的是信息分享者; FR 约等于 1 时,用户既是信息获取者也是信息分享者。

对不同 K-核网络的用户跟随比例进行了统计,包括 FR 的最大值、均值、中位数、最小值,以及小于 0.1、小于 1、等于 1、大于 1、大于 10 的用户所占的比例,如表 2 所示:

表 2 K-核的跟随比例

| K | 最大值 | 均值 | 中位数 | 最小值 | 小于 0.1 | 小于 1 | 等于 1 | 大于 1 | 大于 10 |
|-----|--------|-------|------|-------|----------|-------|-------|-------|-------|
| K30 | 52 629 | 12.38 | 1.32 | 0.072 | 0.000 25 | 0.29 | 0.004 | 0.705 | 0.029 |
| K40 | 27 609 | 5.54 | 1.27 | 0.072 | 0.000 31 | 0.317 | 0.004 | 0.679 | 0.02 |
| K45 | 27 609 | 5.28 | 1.2 | 0.072 | 0.000 3 | 0.365 | 0.004 | 0.631 | 0.019 |
| K50 | 27 609 | 6.08 | 1.11 | 0.72 | 0.000 28 | 0.409 | 0.005 | 0.586 | 0.014 |
| K55 | 1 308 | 2.07 | 1.02 | 0.72 | 0.000 21 | 0.476 | 0.006 | 0.519 | 0.013 |
| K60 | 1 308 | 2.13 | 0.97 | 0.13 | 0 | 0.522 | 0.007 | 0.471 | 0.012 |
| K65 | 171 | 1.45 | 0.91 | 0.13 | 0 | 0.592 | 0.009 | 0.399 | 0.01 |

从表 2 可以观察到,随着 K 值的增大,跟随比例小于 1 的比例逐渐增大,等于 1 的比例也有略微增大,而大于 1 的比例却在减小。这充分说明经过 K-核分解后的网络中用户信息传输行为逐渐从信息的分享者向信息获取者过渡。从大于 10 的比例与小于 0.1 的比例可知,有一部分用户在网络中是绝对的信息分享者或者绝对的信息获取者。还可以观察到信息绝对分享

者所占的比例远远大于信息绝对获取者所占的比例,这说明总体来看,绝大多数微博用户似乎是以信息的分享者身份存在的。但其实造成这种结果另有原因,根据新浪微博平台的规则可知,在微博用户中,用户的被关注人数是没有限制的(有的名人有数千万粉丝),而关注别人一般不能超过 2 000,这就说明微博平台上用户去获取信息是有限制的。随着 K 值的增大,绝对的信息获取者和分享者都有大幅度的下降,这说明剩下的核心用户之间都存在一定程度的相互交流。

4.3 度相关性

度相关性相对于度分布来说,更能体现网络中节点之间的连接特性。网络的度相关性可以用余平均度来度量。节点的余平均度^[3]是指节点的邻居节点度的平均值。令节点 i 的 k_i 个邻居节点的平均度为 $\langle k_{nn} \rangle_i$, 则:

$$\langle k_{nn} \rangle_i = \frac{1}{k_i} \sum_{j=1}^{k_i} k_{ij} \quad (2)$$

假设网络中度为 k 的节点为 v_1, v_2, \dots, v_{i_k} , 那么度为 k 的节点的余平均度计算公式如下:

$$\langle k_{nn} \rangle(k) = \frac{1}{i_k} \sum_{i=1}^{i_k} \langle k_{nn} \rangle_{v_i} \quad (3)$$

当 $\langle k_{nn} \rangle(k) - k$ 的曲线斜率大于 0 时,该网络为同配网络,斜率小于 0 时,为异配网络。同配网络中度大的节点倾向于和度大的节点相连,异配网络中度大的节点倾向于和度小的节点相连。

以 K60 网络为代表,计算其度与余平均度之间的关系,如图 2 所示:

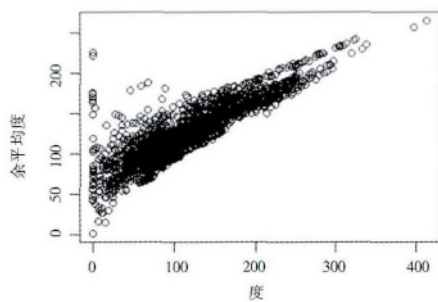


图 2 度与余平均度关系图

图 2 说明, K-核网络是同配网络。已有研究表明,大多数现实社会网络具有同配特性,即度大的节点倾向于和度大的节点相连^[17]。从社会学角度来说,这是合理的,因为在现实生活中,虽然普通人希望和名人建立联系,但名人更倾向于和自己同等地位的人交往,

从而形成同配性。然而,在线社会网络经常会表现出异配性,即度大的节点倾向于和度小的节点相连^[17]。因为在线社会网络打破了社会阶层间无形的壁垒,每个人都有机会并可以很容易跟那些人气值很高的个人建立联系。但是本文的 K-核网络并没有表现出大多数在线社会网络的异配性,而是表现出与现实社会网络类似的同配特性。产生这一结果的原因是微博关注关系网络经过 K-核分解后,保留的是联系紧密的核心用户,这些用户可能在现实生活中就是熟识的, K-核网络在很大程度上反映的就是现实生活中这些用户之间的关系(尤其是 K 值较大时),对 K-核网络的社区分析也说明了这一点。

5 K-核网络社区分析

5.1 社区检测

在线社会网络一般都表现出很强的社区特性,即网络由若干较为明显的社区构成,社区内部用户之间的联系较为紧密,而社区之间用户的联系较为稀疏。这些社区可能是网络上形成的虚拟团体,也可能是真正的社会团体,社区成员一般拥有共同的兴趣爱好或相似的背景。以 K60 为例进行社区检测和分析。K60 包括 3 156 个节点, 188 841 条边,节点的平均度为 119.67,最短平均路径为 2.9,网络密度为 0.019,聚类系数为 0.61。利用 Pons 等^[18]提出的随机游走方法在 R 语言环境下编程实现对 K60 网络的社区检测,共得到 27 个社区,其模块度为 0.947。社区的基本参数如表 3 所示:

表 3 社区相关信息

| 社区编号 | 节点数 | 边数 | 平均度 | 密度 | 社区编号 | 节点数 | 边数 | 平均度 | 密度 |
|------|-----|--------|--------|-------|------|-----|--------|--------|-------|
| 1 | 167 | 9 640 | 115.45 | 0.348 | 15 | 73 | 4 287 | 117.45 | 0.816 |
| 2 | 141 | 7 615 | 108.01 | 0.386 | 16 | 127 | 7 221 | 113.72 | 0.451 |
| 3 | 251 | 15 369 | 122.46 | 0.245 | 17 | 126 | 8 692 | 137.97 | 0.551 |
| 4 | 361 | 20 120 | 111.47 | 0.155 | 18 | 91 | 7 284 | 160.09 | 0.89 |
| 5 | 129 | 7 738 | 119.97 | 0.468 | 19 | 86 | 5 737 | 133.41 | 0.785 |
| 6 | 80 | 3 995 | 99.88 | 0.632 | 20 | 163 | 12 989 | 159.37 | 0.492 |
| 7 | 79 | 4 180 | 105.82 | 0.678 | 21 | 54 | 2 346 | 86.89 | 0.82 |
| 8 | 74 | 3 387 | 91.54 | 0.627 | 22 | 57 | 2 753 | 96.6 | 0.862 |
| 9 | 183 | 11 531 | 126.02 | 0.346 | 23 | 55 | 2 502 | 90.98 | 0.842 |
| 10 | 150 | 8 696 | 115.95 | 0.389 | 24 | 38 | 1 326 | 69.79 | 0.943 |
| 11 | 140 | 7 685 | 109.78 | 0.395 | 25 | 76 | 5 511 | 145.03 | 0.967 |
| 12 | 139 | 7 138 | 102.71 | 0.372 | 26 | 91 | 8 185 | 179.89 | 0.999 |
| 13 | 128 | 7 727 | 120.73 | 0.475 | 27 | 60 | 3 540 | 118 | 1 |
| 14 | 37 | 1 253 | 67.73 | 0.941 | | | | | |

针对社区检测的结果,分别对每个社区内的成员进行分析。在得到的 27 个社区中有 16 个社区其成员分别来自同一地区,而且在这 16 个社区中大部分都是

IT、通信行业、教育和商业服务业。IT、通信行业社区主要由电信、联通等行业的微博用户构成,教育类社区主要是由学生、教师微博用户构成,商业服务业社区主要是由商业部门的微博用户构成。27个社区的分布如表4所示,这充分说明微博用户关系的K-核网络在很大程度上反映了现实社会中的社会团体结构。

表4 社区分布

| 社区类别 | IT、通信 | 教育 | 娱乐 | 商业、服务业 | 旅游 | 育儿 | 时尚 | 公益 | 房地产业 |
|------|-------|----|----|--------|----|----|----|----|------|
| 社区个数 | 5 | 7 | 4 | 6 | 1 | 1 | 1 | 1 | 1 |

5.2 社区分析

(1) 节点中心性分析

节点中心性反映的是网络中节点的重要程度,度量节点中心性的参数有很多,如度、介数、接近度等^[3]。在27个社区中,随机抽取编号为20的社区进行分析。该社区中包括163个节点,12 989条边,分别计算每个节点的度、介数和接近度。3个参数之间的关系如图3所示:

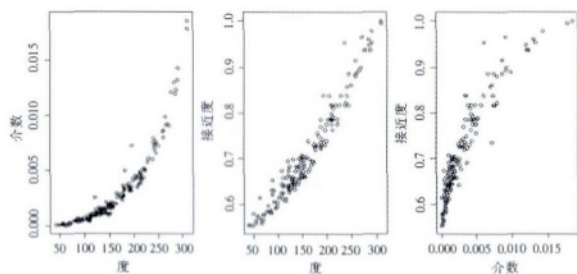


图3 度、介数和接近度关系图

从图3可以看出,3个参数基本呈现正相关关系,说明3个参数对节点中心性的描述能力是类似的。对每个节点的度、介数和接近度3个参数加权平均并进行排序,提取排名靠前的10个用户,发现他们均来自同一个技师学院,这10个用户中有9个是该学院老师,且大部分是该学院的上层领导,由此可以推定社区形成是现实关系驱动的,与现实社会网络具有相似性。

(2) 互惠性分析

在一个有向网络中,当节点A连接到节点B上,且节点B也连接到节点A上时,就称节点A与B是互惠的,一个节点的互惠数是指该节点与网络中存在互惠关系的节点总数^[19]。具体到本文的微博用户关注关系网络,相互关注的用户之间是互惠的,用户的互惠数指网络中与该用户存在相互关注关系的节点总数。

为研究社区的社区性,首先计算社区内每个用户的互惠数 d 与社区总节点数 n 的比例 d/n ,比例均值记为 bi_{in} , bi_{in} 反映了社区内部用户之间联系的紧密程度, bi_{in} 越大联系越紧密。其次,计算社区内每个用户的互惠数 d 与该用户在整个网络中的互惠数 N 的比例 d/N ,比例均值记为 bi_{out} , bi_{out} 反映的是社区内部用户与外部用户联系的紧密程度, bi_{out} 越大,社区节点越倾向于与社区内部节点建立联系,而不是与社区之外的节点建立联系。27个社区的 bi_{in} 与 bi_{out} 计算结果如表5所示:

表5 社区互惠性

| 社区编号 | bi_{in} | bi_{out} | 社区编号 | bi_{in} | bi_{out} | 社区编号 | bi_{in} | bi_{out} |
|------|-----------|------------|------|-----------|------------|------|-----------|------------|
| 1 | 0.26 | 0.36 | 10 | 0.25 | 0.17 | 19 | 0.68 | 0.54 |
| 2 | 0.33 | 0.07 | 11 | 0.3 | 0.18 | 20 | 0.4 | 0.47 |
| 3 | 0.18 | 0.18 | 12 | 0.26 | 0.28 | 21 | 0.8 | 0.27 |
| 4 | 0.13 | 0.26 | 13 | 0.36 | 0.21 | 22 | 0.85 | 0.34 |
| 5 | 0.41 | 0.21 | 14 | 0.92 | 0.14 | 23 | 0.78 | 0.23 |
| 6 | 0.55 | 0.27 | 15 | 0.83 | 0.51 | 24 | 0.92 | 0.14 |
| 7 | 0.61 | 0.15 | 16 | 0.31 | 0.21 | 25 | 0.97 | 0.7 |
| 8 | 0.59 | 0.2 | 17 | 0.53 | 0.18 | 26 | 0.99 | 0.8 |
| 9 | 0.27 | 0.11 | 18 | 0.94 | 0.61 | 27 | 0.98 | 0.92 |

一般来说,现实中的社会网络相对于在线社会网络应该具有更高的互惠指数,因为相对于面对面方式构建起来的熟人关系网,个人在线交际中对于响应用户邀请会感到较少的社会压力,从而导致了在线社会网络中接受用户邀请的比例相对偏低。而本文研究的K-核网络是经过K-核分解之后获得的只包含核心用户的关系网络,更接近现实中的社会网络。从表5中可以观察到 bi_{in} 的最大值达到0.99,最小值为0.13,平均值为0.57,说明每一社区用户平均与57%的社区内用户是相互关注的,这说明社区内部用户之间的联系较为紧密。 bi_{out} 的最大值为0.92,平均值为0.35,也在一定程度上反映了社区内部用户之间联系较为紧密,社区之间用户联系相对稀疏这一现象。

(3) 中间人角色分析

在一个社区中存在一些“中间人”,在不同的社区之间也可能存在一些“中间人”^[20]。中间人指的是居于中间位置的人。具体来讲,就是网络中存在三个节点A、B、C,它们之间的关系是B有一个指向A的有向边,A有一个指向C的有向边,而B与C是没有关系的。这样可以认为A在三者之间充当中间人的角色。根据A、B、C三者是否属于同一个社区,可以把A作为中间人的角色分为5类,分别是:协调员、顾问、代理

人、守门人、联络人,如图 4 所示:

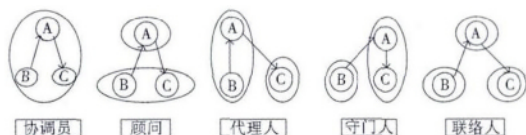


图 4 中间人角色

观察图 4 发现只有“协调员”子图中 A、B、C 处于同一社区,“联络人”子图中三个节点分别处于三个不同社区,其他三个角色子图中,三个节点分别处于两个社区。直观看,若一个网络中充当“协调员”的节点比例较高,则这个网络的社区性比较明显,若一个网络中“联络人”节点的比例较高,则这个网络的社区性就较弱。以 K60 网络及其 27 个社区为例,考察充当各类中间人角色的节点的数量情况,结果如表 6 所示:

表 6 中间人角色统计

| 等级 角色 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|----------|-------|-------|-----|----|---|---|---|---|---|----|
| 协调员 | 161 | 2 606 | 318 | 57 | 9 | 2 | 1 | 0 | 0 | 2 |
| 顾问 | 3 142 | 12 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 代理人 | 2 994 | 157 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 2 |
| 守门人 | 2 994 | 157 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 2 |
| 联络人 | 3 151 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

表 6 中每行对应一种中间人角色,记录充当该中间人角色的节点数,每行数字相加为 3 156(等于 K60 网络的节点总数)。每行分 10 级,1 级表示没有充当(或者说充当了 0 次)对应角色的节点数,2 级表示充当对应角色次数在第 2 区段的节点数。依此类推,10 级表示充当对应角色次数在最大区段的节点数。区段等级按数值均匀划分。所有 3 156 个节点中,只有 161 个没有充当过协调员,充当了协调员的节点比例为 $2\,995/3\,156=94.9\%$,比例非常高;而联络人方面,只有 5 个节点充当过,有 3 151 个节点没有充当过联络人角色,充当过联络人的节点的比例为 $5/3\,156=0.16\%$,几乎可以忽略不计。表 6 的数据表明 K60 网络的社区特性非常明显。

6 结 语

在线社会网络的量化分析是一个涉及社会学、管理学、数学、信息科学的研究分支。本文以新浪微博为例,构建并研究了微博用户关注关系网络,特别是反映微博核心用户之间关注关系的 K-核网络。通过对 K-核网络基本参数的计算和其社区结构的分析,发现

微博核心用户关系网络既具有在线社会网络的一般特征,也带有现实社会网络的一些特点。度的相关性计算显示 K-核网络没有表现出大多数在线社会网络的异配性,而是表现出与现实社会网络类似的同配特性,说明它在很大程度上反映的是现实社会中的社会团体结构。社区检测及其模块度计算显示本文网络具有明显的网络社区特性,中间人角色分析结果也证明了这一点。度中心性、介数中心性、接近度中心性计算表明,微博核心用户之间的关注关系是现实关系驱动的,相对于一般的在线社会网络其与现实社会网络更具相似性。互惠性分析从另一个角度反映了 K-核网络的这一特点。对于以上发现本文分析了原因,并给出了其合理性分析。

本文工作有助于揭示在线社会网络,尤其是微博核心用户之间的关联关系和信息传播规律,有助于理解网络虚拟社区的组织结构特征,进而实现对互联网用户的有效管理和对社会网络的有效利用。采用复杂网络理论与技术,研究微博用户关注关系网络在时间轴上的动态演化规律与动力学特征,实现对社会网络信息传播的预测与控制是今后的研究方向。

参考文献:

- [1] 刘挺,徐志明,秦兵,等.从语言计算到社会计算[J].中国计算机学会通讯,2011,7(12):31-39. (Liu Ting, Xu Zhiming, Qin Bing, et al. From Language Computing to Social Computing[J]. Communications of the China Computer Federation, 2011, 7(12): 31-39.)
- [2] Guo Z, Li Z, Tu H. Sina Microblog: An Information-driven Online Social Network[C]. In: Proceedings of the 2011 International Conference on Cyberworlds. 2011: 160-167.
- [3] 汪小帆,李翔,陈关荣.网络科学导论[M].北京:高等教育出版社,2012. (Wang Xiaofan, Li Xiang, Chen Guanrong. Network Science: An Introduction[M]. Beijing: Higher Education Press, 2012.)
- [4] Mislove A, Marcon M, Gummadi K P, et al. Measurement and Analysis of Online Social Networks[C]. In: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement. New York, NY, USA: ACM, 2007: 29-42.
- [5] Java A, Song X, Finin T, et al. Why We Twitter: Understanding Microblogging Usage and Communities[C]. In: Proceedings of the 9th WebKDD and 1st SNAKDD Workshop on Web Mining and Social Network Analysis. New York, NY, USA: ACM, 2007: 56-65.

- [6] Teutle A R M. Twitter: Network Properties Analysis[C]. In: *Proceedings of the 20th International Conference on Electronics, Communications and Computer (CONIELECOMP)*. Washington DC: IEEE Computer Society 2010: 180 - 186.
- [7] Kwak H, Lee C, Park H et al. What is Twitter, A Social Network or a News Media? [C]. In: *Proceedings of the 19th International Conference on World Wide Web*, Raleigh, USA. New York, NY, USA: ACM 2010: 591 - 600.
- [8] 胡海波, 徐玲, 王科, 等. 大型在线社会网络结构分析[J]. 上海交通大学学报, 2009, 43(4): 587 - 591. (Hu Haibo, Xu Ling, Wang Ke, et al. Structural Analysis of Large Online Social Network[J]. *Journal of Shanghai Jiaotong University*, 2009, 43(4): 587 - 591.)
- [9] 余高辉, 杨建梅, 曾敏刚. QQ 群好友关系的复杂网络研究[J]. 华南理工大学学报: 社会科学版, 2011, 13(4): 20 - 23. (Yu Gaohui, Yang Jianmei, Zeng Mingang. Research on QQ Group Based on Complex Networks[J]. *Journal of South China University of Technology: Social Science Edition*, 2011, 13(4): 20 - 23.)
- [10] 王晓光, 袁毅, 滕思琦. 微博社区交流网络结构的实证分析[J]. 情报杂志, 2011, 30(2): 199 - 203. (Wang Xiaoguang, Yuan Yi, Teng Siqi. Empirical Analysis on Communicating Structure of Micro - blog Community[J]. *Journal of Intelligence* 2011, 30(2): 199 - 203.)
- [11] 袁毅, 杨成明. 微博客用户信息交流过程中形成的不同社会网络及其关系实证研究[J]. 图书情报工作, 2011, 55(12): 31 - 35. (Yuan Yi, Yang Chengming. Empirical Analysis of All Kinds of Social Networks and Their Relationships Formed by Information Communication Among Microblog Users[J]. *Library and Information Service*, 2011, 55(12): 31 - 35.)
- [12] 何黎, 何跃, 霍叶青. 微博用户特征分析和核心用户挖掘[J]. 情报理论与实践, 2011, 34(11): 121 - 125. (He Li, He Yue, Huo Yeqing. Microblogging User Characteristics Analysis and Mining Core Users[J]. *Information Studies: Theory & Application*, 2011, 34(11): 121 - 125.)
- [13] Kang S, Zhang C, Lin Z et al. Complexity Research of Massively Microblogging Based on Human Behaviors[C]. In: *Proceedings of the 2nd International Workshop on Database Technology and Applications (DBTA)*, Wuhan, China. 2010: 1 - 4.
- [14] 田占伟, 隋场. 基于复杂网络理论的微博信息传播实证分析[J]. 图书情报工作, 2012, 56(8): 42 - 46. (Tian Zhanwei, Sui Yang. The Empirical Analysis of Micro - blog Information Flow Based on Complex Network Theory[J]. *Library and Information Service*, 2012, 56(8): 42 - 46.)
- [15] Fan P, Li P, Jiang Z et al. Measurement and Analysis of Topology and Information Propagation on Sina - Microblog[C]. In: *Proceedings of the IEEE International Conference on Intelligence and Security Informatics (ISI)*. Washington DC: IEEE Computer Society, 2011: 396 - 401.
- [16] 梁斌. 新浪微博用户信息 [DB/OL]. [2012 - 06 - 05]. <http://www.cnpmeng.com>. (Liang Bin. User's Information of Sina Weibo [DB/OL]. [2012 - 06 - 05]. <http://www.cnpmeng.com>.)
- [17] 胡海波, 王科, 徐玲, 等. 基于复杂网络理论的在线社会网络分析[J]. 复杂系统与复杂性科学, 2008, 5(2): 1 - 14. (Hu Haibo, Wang Ke, Xu Ling, et al. Analysis of Online Social Networks Based on Complex Network Theory[J]. *Complex Systems and Complexity Science*, 2008, 5(2): 1 - 14.)
- [18] Pons P, Latapy M. Computing Communities in Large Networks Using Random Walks[C]. In: *Proceedings of the 20th International Conference on Computer and Information Sciences*, Istanbul, Turkey. Berlin, Heidelberg: Springer - Verlag, 2005: 284 - 293.
- [19] 张春红, 于翠波, 朱新宁, 等. 社交网络(SNS)技术基础与开发案例[M]. 北京: 人民邮电出版社, 2012. (Zhang Chunhong, Yu Cui bo, Zhu Xinning, et al. Social Network Services Technology Base and Develop Case [M]. Beijing: Posts & Telecom Press, 2012.)
- [20] 刘军. 整体网分析讲义: UCINET 软件实用指南[M]. 上海: 格致出版社, 2009. (Liu Jun. Lectures on Whole Network Approach: A Practical Guide to UCINET [M]. Shanghai: Truth & Wisdom Press, 2009.)
- (作者 E-mail: blg501@163.com)