

在线论坛中的意见领袖分析

高俊波¹, 杨静²

(1. 上海海事大学信息工程学院 上海 浦东新区 200135; 2. 马鞍山市电力公司 安徽 马鞍山 232000)

【摘要】将论坛中帖子间的回复关系映射为发帖者之间的关联关系, 构造出一个社群网络, 并采用了统计物理学中的方法对这个复杂网络进行了特征分析, 验证了其具有小世界网络的特征, 并对基于小世界网络在线论坛中意见领袖的发现进行了研究, 找出了论坛中的意见领袖并验证了其正确性, 为进一步研究复杂网络中的人物行为奠定了基础。

关键词 聚类; 信息传播; 在线论坛; 意见领袖; 小世界网络
中图分类号 TP311.1 **文献标识码** A

Analysis of Opinion Leader in On-Line Communities

GAO Jun-bo¹, YANG Jing²

(1. Faculty of Information Engineering, Shanghai Maritime University Pudong Shanghai 200135;
2. Maanshan Electronic Power Company Maanshan Anhui 232000)

Abstract In this paper, we investigate an online forum and construct a social network with the replying relations between comments mapped to the comment authors' relations. We analyze the network's feature and apply the statistical physics methods to this complex network research. Then we validate its characteristics of small world networks and successfully find out the opinion leaders in on-line community based on small-world network.

Key words cluster; information communication; on-line community; opinion leader; small-world network

20世纪40年代, 美国哥伦比亚大学的传播学者保罗·拉扎斯费尔德提出了“意见领袖”的概念。他认为, 对于媒介所传播的信息和观点, 有部分受众会积极接受, 并加以再传播, 这些人即为“意见领袖”; 而另一部分人则主要依靠与这些“意见领袖”的接触来指导自己的行动。随着信息技术的高速发展, Internet已成为信息发布的第四媒体, 信息的分析与处理成为人们关注的焦点之一^[1]。基于上述事实, 本文以在线论坛为研究对象, 通过分析帖子间的回复关系, 获得论坛中发帖者之间的相互关系, 进而构建出论坛中的社群网络。对所构建的社群网络进行特征分析, 发现其具备小世界网络的特征, 最后, 结合实例利用小世界网络找出了在线论坛中的意见领袖。

1 小世界网络特征

文献[2-3]提出了单参数的小世界网络(Small World Network)模型。该网络模型介于规则网络与随

机网络之间, 在它们之间架起了桥梁。原始的WS模型描述如下:

(1) 初始化。考虑一个具有 N 个节点的邻近节点耦合的环状网络, 其中节点 i 连结到它的 K 个邻近的节点 $i\pm 1, i\pm 2, \dots, i\pm \frac{K}{2}$, K 是一个偶整数。假定 $N \gg K \gg \ln(N) \gg 1$ 。保证整个网络是相互连结的, 但又是稀疏的。

(2) 随机化。以概率 p 随机地改写网络的每一条边, 即以概率 p 将一条现成的边重新连结到另一个顶点上, 同时避免将自己连结到自己或者与已有的边相重合的情形。该过程引入了 $\frac{pNK}{2}$ 条边, 它们连接到新的节点上, 重新连结的边通常称为捷径。当调节参数 p 从0(有序)到1(随机)时, 可以密切监视整个变换过程。

小世界网络介于规则网络和随机网络之间, 它实现了从规则网络到完全随机网络之间的连续演变, 如图1所示。

收稿时间: 2007-09-07

基金项目: 上海市重点学科资助建设项目(70602); 上海市教育科研资助项目(06FZ006)

作者简介: 高俊波(1972-)男, 博士, 主要从事数据挖掘、知识发现方面的研究; 杨静(1974-), 女, 工程师, 主要从事信息检索方面的研究。

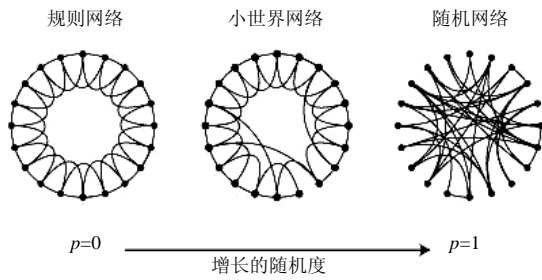


图1 复杂网络的演变过程

经过研究,文献[3]改进了原始的WS模型,在NW模型里,代替改写节点之间的连结,随机地增加一些新的边,即所谓的捷径,且不移走已经存在的边。显然,若 $p=0$,则NW模型变成原始的邻近节点耦合的环状网络;若 $p=1$,则NW模型变成全局耦合的网络。然而,对于充分小的概率 p 和足够大的 N ,NW模型等价于WS模型。随着节点数的增加,WS和NW模型展示了从“大世界”(平均路径长度线性增长)到“小世界”(平均路径长度对数增长)的变换。可以看出小世界网络有如下特征:(1)不考虑网络拓扑结构,节点间有很小的平均距离(平均路径长度);(2)有较大的聚类系数,某节点 a 在其所有邻居中,邻居间的连边数占包括 a 在内的子网连边总数比例,称为聚类系数。它反映了某节点邻居间的聚类程度^[4-5],与网络规模无关且恒小于1。即如果一个网络的聚类系数远远大于随机网络的聚类系数,而且这个网络的平均路径长度近似于随机网络的平均路径长度,就认为该网络具有小世界特性。

2 社群网络的构建与分析

2.1 社群网络的构建

在研究BBS中成员的行为时发现,通过简单地利用成员发表言论中的词语传播概率^[6]来确定人物行为是不完善的。因为这种方法没有考虑到主题之间的关联关系、人物之间的互动关系及主题的时序变化等,不能够准确地分析人物的行为。既然这种从单个个体出发来研究个体行为的方法有缺陷,就应考虑是否能够从整个群体出发,揭示群体组织结构特征,进而研究群体间的人际关系,从而达到研究人物行为的目的^[7]。本文主要采用社会网络分析法构建与分析社群网络,该方法起源于美国社会心理学家莫雷诺创立的社会计量法。社会计量法为社会网络分析奠定了计量分析基础,莫雷诺在分析人际关系的时候所使用的社会计量法的基本思路至今依然影响着社会网络分析的定量分析发展。根据BBS的特性,对BBS中的成员所构成的社群网络给

出如下定义。

(1) 网络中的节点。BBS中的成员(包括主帖者和回帖者);

(2) 网络中的边。成员之间的回复关系(有回复关系的两个节点之间加一条边)。

这种由人及其互动关系所组成的网络是一个复杂网络。为了能简明扼要地说明问题,本文定义该复杂网络图是一个无向无权图。

2.2 小世界特征验证实验

为验证所构建的社群网络满足小世界网络聚类系数高且平均路径短的特征,选择中华军事网论坛(<http://bbs.china.com/military/bbs.jsp>)为实验数据源,针对该论坛数据量大的特点,以单日数据做为一个实验数据集,并以此构建一个无向无权图 G 。一共取了2007-05-20~2007-06-30共计41天数据,做了几十组实验,部分实验结果如表1所示。

由于 G 往往不是一个连通图,因此考虑其连通子图(其最大的连通子图 g_M 中的顶点数达到了 G 中的90%以上),实验数据如表1所示。

表1 最大连通子图的顶点数

数据源日期	$ V(G) $	$ V(g_M) $	$ V(g_M) / V(G) *(100\%)$
2007-06-01	1 878	1 779	94.73
2007-06-02	1 849	1 740	94.10
2007-06-03	2 146	2 041	95.12
2007-06-04	1 496	1 367	91.38
2007-06-05	1 511	1 441	95.37

因此,将验证 g_M 是否具有小世界网络的特征,以JUNG(Java Universal Network/Graph Framework)为基本实验平台,获得的结果如表2所示。

表2 最大连通子图的网络特征

数据日期	$ V $	$ E $	C	C_{rand}	L	L_{rand}
2007-06-01	1 779	3 132	0.558 16	0.001 98	6.277 8	5.797 1
2007-06-02	1 740	3 217	0.546 08	0.002 48	6.278 1	5.773 3
2007-06-03	2 041	3 529	0.598 23	0.001 24	6.200 0	5.908 0
2007-06-04	1 367	2 145	0.589 37	0.001 06	6.629 0	5.555 6
2007-06-05	1 441	2 367	0.571 91	0.00359	6.482 8	5.608 2
2007-06-06	1 414	2 146	0.63825	0.003 74	6.653 2	5.605 1
2007-06-07	1 394	2 237	0.575 98	0.004 03	6.590 1	5.604 2
2007-06-08	1 772	3 002	0.562 89	0.001 26	6.476 9	5.794 4
2007-06-09	1 512	2 337	0.597 13	0.001 08	6.644 7	5.649 7
2007-06-10	1 439	2 433	0.578 11	0.002 00	6.294 2	5.610 0

表中, C 、 L 分别表示图 g_M 的聚类系数和平均路径长度; C_{rand} 、 L_{rand} 分别表示相应的随机图的聚类系数和平均路径长度。

从表2中不难看出图 g_M 满足如下特征:

$$C \gg C_{rand}, L \approx L_{rand}$$

根据小世界网络的定义, 可发现图 g_M 具有小世界网络的特征, 因而验证了本文构建的社群网络具有小世界网络的特征。

3 意见领袖分析与实验

3.1 意见领袖的特点

文献[8]在归纳观念领导人即意见领袖的特点时指出, 观念领导人必须具有一定的外部示范作用, 知识面广, 社会资源丰富, 有一定数量的追随者, 综合能力强, 与众不同的魅力等特征。同样在BBS群体中, 能够处于核心位置的必定是那些有着更开阔的视野, 积极参与讨论, 并能不断提出引人注目的个性观点的群体成员。在客观条件上, 这部分成员往往上网方便且都有大量时间上网, 主观方面则积累了大量的网络经验, 并通过发表有一定见地的言论在论坛中建立威信。他们凭借自己优秀的个人品质、广泛的信息来源渠道, 在论坛中发起广受瞩目的话题, 吸引大量的成员参与讨论, 对周围其他成员产生强烈的影响力, 使得他们在论坛中拥有大批的拥护者, 迅速获得资源优势, 成为论坛的中心人物。这些成员善于交往, 能频繁进行思想情感沟通, 发出和接受大量信息, 通过沟通表达自己的观点。某个论坛成员成为意见领袖, 是凭借其内在的号召力达成的。这种影响力虽然表面上没有明显的约束力, 但在实际上常能发挥强制性影响力所不能发挥的约束作用。在很多时候, 论坛管理者的强制作用(如删帖、封ID)还不及意见领袖的影响作用有效。通常, 自然性影响力主要包括品格因素、能力因素、知识因素、感情因素等。但在论坛中意见领袖的影响力也糅合了强制性影响力中一些因素的影响, 比如资历因素。在生活中, 资历反映了一个人的生活阅历与经验。同样, 在论坛中, 登录和发帖数量会累积成该成员的积分, 积分越多意味着成员的上网时间越长, 参与度越高, 代表着成员网络经验的丰富。在一定条件下资历因素也会影响领导的有效性, 论坛成员普遍对资深领导人的信任度更高一些。

3.2 意见领袖获取

基于上文所构建的社群网络, 可以得到一个以论坛作者为节点, 他们之间的回复关系为边的网络。意见领袖的获取过程就是在网络中找寻关键结点的过程, 通过对平均路径长度的分析来找到这些关键

结点。基于上文中构建的社群网络, 本文给出如下定义。

定义 1 CN是所构建的社群网络, L 是网络CN的平均路径长度。

定义 2 CN_i 是没有第 i 个节点的社群网络, 即在这个网络中, 所有与第 i 个节点所连接的边也被去除了。 L_i 是网络 CN_i 的平均路径长度。

由以上两个定义, 可以给出平均路径长度的变化值 $\Delta L_i = L - L_i$ 。

那么, 通过 L 和 L_i 的变化就可以判断这个节点的重要性了。也就是说, ΔL 高的节点就是网络中比较关键的节点, 这个节点对于整个网络的连接起到了相当重要的作用^[9]。这个具有高 ΔL 值的节点可以将表面上看起来无关的节点甚至是群(cluster)连接起来, 那么在整个网络中, 这种节点就可以被认为是关键结点^[10-11], 也就是所寻找的意见领袖。

3.3 实验

在上文中, 以论坛的单日数据作为一个实验数据集, 构建的是一个无向无权图 G 。由于论坛数据量较大, 为了能够更直观地说明问题, 只取了部分数据, 所构建的社群网络如图2所示。意见领袖在图2中显示为实心节点。

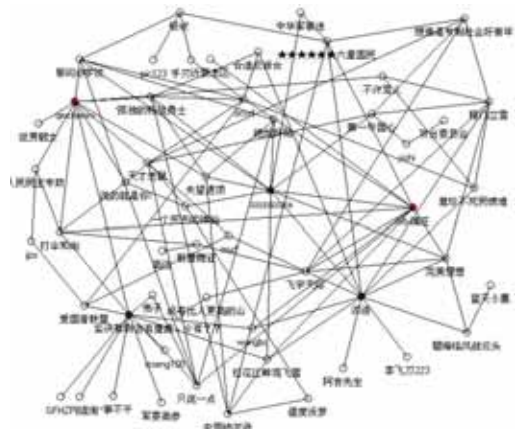


图2 社群网络图

同样, 首先验证了这个社群网络的小世界特征, 如表3所示。

表3 社群网络的网络特征

$ V $	$ E $	C	C_{rand}	L	L_{rand}
54	110	0.395 0	0.080 3	2.913 9	2.953 9

然后, 通过计算每个结点的 ΔL 值, 找出论坛中的意见领袖, 如表4所示。点出度是某成员对他人帖子的回复总量, 他人对某成员的回帖总数被称为点入度。

表4 论坛意见领袖

序号	节点	ΔL	发帖数	点出度	点入度
1	洪浩	0.904 584	3	16	33
2	坚决抵制	0.903 564	2	10	23
3	Zzzzzzzza	0.902 865	2	14	18
4	Qinchenxv	0.902 166	1	8	16
5	Hrb—浦汪	0.902 136	1	6	16

从表4中可以看到,选取出的意见领袖都是网络群体中的核心人物,以成员1来说,他无疑是网络中最重要的人物,他的发帖引起了网络中大多数成员的特别关注,同时他也积极回复其他成员的帖子,与多名成员建立了交流关系,地位在其他意见领袖之上;其他几位意见领袖的发帖也都引起了群体成员的关注。另外,了解到这几位意见领袖都是论坛的资深会员,在过去几年中都发表了大量的帖子,引起论坛成员的广泛关注,知名度都很高,从上文中意见领袖特点来看,他们也完全符合意见领袖的特点。所以,这种基于小世界网络寻找意见领袖的方法是正确的,找出来的作者的确是论坛中的意见领袖。

4 总结

网络的发展迅速地改变着人类这个真实的世界,而虚拟社区就是网络为人类提供的一个崭新空间和交往环境。基于虚拟社区的网络交往特征,人际互动关系是国内外学者关注的热点。本文对在线论坛进行了研究,构建出论坛中的社群网络,并对这个网络进行了特征分析,发现其具备小世界网络

的特征,基于小世界网络本文找出了在线论坛中的意见领袖。后继的工作将分析网络成员在网络中的动力学行为,找出他们的网络特征,对复杂网络中人物行为分析研究领域做进一步的研究。

参考文献

- [1] OHSAWA Y. Chance discoveries for making decisions in complex real world[J]. *New Generation Computing*, 2002, 20(2): 143-164.
- [2] WATTS D J, STROGATZ S H. Collective dynamics of 'small-world' networks[J]. *Nature*, 1998, 393: 440-442.
- [3] WATTS D J. *Small World*[M]. Princeton: Princeton University Press, 1990.
- [4] NEWMAN M E J. Scientific collaboration networks: II. Shortest paths, weighted networks and centrality[J]. *Phys Rev E*, 2001, 64(016132): 1-7.
- [5] NEWMAN M E J. The structure and function of complex networks[J]. *SIAM Review*, 2003, 45(2): 167-256.
- [6] HOBBS J R. Information extraction from biomedical text[J]. *Journal of Biomedical Informatics*, 2002, 35(4): 260-264.
- [7] 宫辉, 徐渝. 高校BBS社群结构与信息传播的影响因素[J]. *西安交通大学学报(社会科学版)*, 2007, 27(81): 93-96.
- [8] ROGERS E M. *Diffusion of innovations*[M]. [S.l.]: The Free Press, 1962.
- [9] KONGACHANDRA R, KIMPANT C, SUWANAPONQ T, et al. Newly-born keyword extraction under limited knowledge resources based on sentence similarity verification[C]// *Communications and Information Technology (ISCIT 2004)*. Sapporo, Japan: IEEE, 2004: 1183-1187.
- [10] 高俊波, 张敏, 王煦法. 一种新的征兆发现算法研究[J]. *小型微型计算机系统*, 2006, 27(4): 687-690.
- [11] 邹刚, 刘洋, 刘群, 等. 面向Internet的中文新词语检测[J]. *中文信息学报*, 2004, 18(6): 1-9.

编辑 税红

(上接第1197页)

参考文献

- [1] Cisco Systems. Cisco NetFlow introduction[EB/OL]. <http://www.cisco.com/warp/public/732/Tech/nmp/netflow/index.shtml>, 2007-07-28.
- [2] Cisco System. Netflow services solutions guide[EB/OL]. http://www.cisco.com/en/US/products/sw/netmgtsw/ps_1964/products_implementation_design_guide09186a00800d6a11.html, 2007-07-30.
- [3] 刘思峰, 党耀国, 张岐山. 灰色系统理论及其应用[M]. 第3版. 北京: 科学出版社, 2004.
- [4] 肖新平, 宋中民, 李峰. 灰技术基础及其应用[M]. 北京: 科学出版社, 2005.
- [5] 杨嵘, 张国清, 韦卫, 等. 基于NetFlow流量分析的网络攻击行为发现[J]. *计算机工程*, 2005, 31(13): 137-139.
- [6] 熊齐邦, 黄明哲. 基于Netflow和异步服务的网络流量监测系统[J]. *计算机工程*, 2006, 32(13): 144-146.
- [7] INMON W H. *数据仓库*[M]. 第4版. 王志海, 译. 北京: 机械工业出版社, 2006.
- [8] MORAN J, GRANADA E, MIGUEZ J L, et al. Use of grey relational analysis to assess and optimize small biomass boilers[J]. *Fuel Processing Technology*, 2006, 87(2): 123-127.
- [9] LIN S J, LU I J, LEWIS C. Grey relation performance correlations among economics, energy use and carbon dioxide emission in Taiwan[J]. *Energy Policy*, 2007, 35(3): 1948-1955.

编辑 熊思亮