

DOI: 10.7652/xjtuxb201406002

有向网络兴趣社区的快速挖掘算法 及其在僵尸粉检测中的应用

王晨旭¹, 秦涛¹, 管晓宏^{1,2}, 周亚东¹

(1. 西安交通大学智能网络与网络安全教育部重点实验室, 710049, 西安;

2. 清华大学自动化系智能与网络化系统研究中心, 100084, 北京)

摘要: 针对传统的无向网络社区挖掘方法无法实现大规模有向网络中社区有效发现的问题, 提出了一种新的有向图社区及其兴趣特征快速挖掘算法。采用贪心算法求解社区划分模块性最大化的优化问题, 较好地平衡了有向图社区挖掘中准确性与有效性之间的矛盾, 实现对大规模微博类有向网络社区结构的有效识别; 基于发现的社区, 采用 tf-idf 算法进一步挖掘社区用户的兴趣爱好, 实现了对微博网络中兴趣小组的精确挖掘。基于新浪微博的实验结果表明: 所提算法不仅可以快速有效地挖掘有向网络中的社区结构及其用户的兴趣特征, 还能够准确地检测出微博网络中的僵尸粉社区, 研究结果对微博系统的净化、谣言控制、网络广告的精准投放等研究具有重要的参考价值。

关键词: 微博; 有向图; 社区挖掘; 用户兴趣小组; 僵尸粉

中图分类号: TP393 **文献标志码:** A **文章编号:** 0253-987X(2014)06-0007-06

A Fast Mining Algorithm for Interest Community in Directed Networks and Its Application to Detection of Zombie Fans

WANG Chenxu¹, QIN Tao¹, GUAN Xiaohong^{1,2}, ZHOU Yadong¹

(1. MOE Key Lab for Intelligent and Network Security, Xi'an Jiaotong University, Xi'an 710049, China;

2. Center for Intelligent and Networked Systems, Department of Automation, Tsinghua University, Beijing 100084, China)

Abstract: A new fast community unfolding and interests mining algorithm is proposed to solve the problem that traditional methods cannot effectively extract communities from large-scale directed networks. A greedy algorithm is used to maximize modularity so that the tradeoff between the accuracy and efficiency in the community mining of directed networks is better balanced and its application to large scale microblog networks can be realized. The users' interests in the extracted community are then further mined using the tf-idf algorithm to score the most-occurred phrases in the community. Experimental results based on Sina Microblog show that the proposed algorithm can not only find out the community structures and their interests quickly, but also can uncover the zombie-fans community efficiently and accurately. These results exhibit great values for system purification, rumors control and accurate delivery of online advertising in microblog systems.

Keywords: microblog; directed graph; community mining; user interest groups

收稿日期: 2013-12-12。 作者简介: 王晨旭(1986—), 男, 博士生; 秦涛(通信作者), 男, 讲师。 基金项目: 国家自然科学基金资助项目(61221063, 61103240, 6113241); 国家科技支撑计划资助项目(2011BAK08B02); 中央高校基本科研业务费资助项目(2012jdhz09, xjj2011015)。

网络出版时间: 2014-05-30

网络出版地址: <http://www.cnki.net/kcms/detail/61.1069.T.20140530.1615.002.html>

微博作为一种新兴的社交媒体,自其创建以来便迅速得到了众多用户的喜爱和参与,因其对网络营销、社会舆论的极大影响力而受到越来越多学者的关注。微博通过用户之间的相互关注而构成网络,且这种关注关系是单向的,使微博关注网络成为典型的有向网络^[1]。在社会网络中,具有相似角色、拥有共同兴趣的用户往往会自觉或不自觉地紧密联系在一起,进而构成网络中的社区结构。社区发现作为复杂网络研究的核心内容,在朋友关系网络、在线社会网络以及 Web 网络等复杂网络中有着重要的应用^[2-4],对微博网络中的社区及其用户兴趣爱好进行挖掘是实现高效网络管理与控制的基础。传统方法由于忽略了边的方向性信息,导致挖掘结果有一定的误差,Leicht 等强烈建议在对有向图进行社区挖掘时应尽可能地利用边的方向性信息^[5]。Nic-
osia 等在 Newman 给出的有向图模块性定义的基础上,提出了一种可用于发现有向网络中有交叠的社区发现方法^[6]。Levorato 等提出了一种发现有向网络中 p -连接子图的方法,并将其用于有向网络中的社区发现^[7]。Wang 等研究了大规模社交网络中核心社区的检测^[8]。这些有向图社区发现算法虽然考虑了边的方向性,但计算复杂度高或定义不明确,很难应用于微博类大规模有向网络的社区发现。

针对上述问题,从有向图的模块性定义出发,本文提出一种适用于大规模有向网络的社区发现算法,该算法使用贪心策略来避免矩阵特征值的计算,在考虑了边的方向性的同时有效地提高了计算效率,能够找到与最优解相当的次优解,从而有效地平衡了算法准确性和效率之间的矛盾。随后基于所挖掘出的社区结构,对社区成员发布的微博内容进行分析,进一步挖掘社区成员的兴趣爱好。由于某些具有公共话题属性的短语并不能反映社区成员的兴趣特征,本文提出了一种基于搜索引擎 tf-idf 算法的兴趣短语评分算法,实现了对社区用户兴趣爱好关键词特征的有效提取。

基于新浪微博数据集的测试结果显示,本文所提出的算法不仅可以快速有效地挖掘出微博系统中存在的用户社区,还可以有效准确地识别社区用户的兴趣爱好。在实验中发现由正常微博用户构成的社区话题内容丰富且兴趣爱好特征明显,而僵尸粉社区则话题内容单一且用户之间的连接不符合小世界特性,因此所提社区兴趣发现算法能够应用于由恶意用户所组建的僵尸粉社区的识别。

1 微博网络社区发现算法

微博关注网络的定义如下。

定义 1 微博关注网络为 $G(V, E)$, 其中 V 为微博用户的集合, E 为用户之间关注关系的集合。边的构造原则为:如果用户 $u, v \in V$, 且用户 u 关注了用户 v , 则存在一条从 u 指向 v 的有向边, 即 $e_{uv} \in E$ 。

1.1 无向网络社区模块性定义及其增益计算

被多数研究者所接受的一个复杂网络中社区结构的定义是,社区结构内部节点连接紧密而与结构外的其他节点连接稀疏^[9-10]。Girvan 与 Newman 定义了社区结构的模块性,用来评价社区成员之间连接的紧密程度,定量地衡量社区划分的好坏^[10]。

定义 2 无向图社区结构的模块性定义为,原始网络中连接社区内部节点的边所占比例与随机网络中连接社区内部节点的边所占比例的期望的差值

$$Q = \frac{1}{2w} \sum_{i,j} \left[A_{ij} - \frac{s_i s_j}{2w} \right] \delta(c_i, c_j) \quad (1)$$

式中: A_{ij} 为边 e_{ij} 的权值; $s_i = \sum_{j \in N(i)} A_{ij}$ 为节点 i 的强度; $N(i)$ 为节点 i 的邻居节点的集合; $w = \sum_{i,j} A_{ij}$ 为无向图所有边的权值之和; c_i 为节点 i 所属社区的标签; $\delta(a, b)$ 为 Kronecker 冲击函数。如果 $a = b$, $\delta(a, b) = 1$; 否则, $\delta(a, b) = 0$ 。社区发现问题可以转化为求解模块性最大化这一优化问题,然而这一问题的精确求解已被证明为 NP-hard^[11],为了提高在大规模网络中社区发现的效率,Blondel 等提出了一种基于无向图模块性优化的社区结构快速发现方法^[12]。该方法的核心思想是计算将单个节点 i 归入其相邻社区 C 所获得的模块性增益,而该增益的计算无需考虑全局信息而仅需考虑节点 i 所在的局部结构信息,大大减少了计算所需时间。无向图模块性增益计算公式为

$$\Delta Q = \left[\frac{\sum_{in} + s_{i,in}}{2w} - \left(\frac{\sum_{tot} + s_i}{2w} \right)^2 \right] - \left[\frac{\sum_{in}}{2w} - \left(\frac{\sum_{tot}}{2w} \right)^2 - \left(\frac{s_i}{2w} \right)^2 \right] \quad (2)$$

式中: $\sum_{in}$ 为社区 C 内部所有边的权值之和; $\sum_{tot}$ 为外部节点连接社区 C 内部节点所有边的权值之和; s_i 为节点 i 的强度; $s_{i,in}$ 为节点 i 连接社区 C 内部节点的边的权值之和; w 为所有边的权值之和。

1.2 有向网络社区模块性定义及其增益计算

无向图模块性的定义并不能直接应用于有向网络,这是因为在构建随机有向网络时需要分别根据节点的出度和入度对边进行随机连接,而不是仅仅

根据节点的度(出度和入度之和)进行连接。为了计算有向图的模块性, Leicht 和 Newman 在文献[5]中对有向图的模块性做了定义。

定义 3 有向图社区结构的模块性定义为,原始有向网络中连接社区内部节点的边所占比例与有向随机网络中连接社区内部节点的边所占比例的期望的差值

$$Q = \frac{1}{\omega} \sum_{i,j} \left[A_{ij} - \frac{s_i^{\text{in}} s_j^{\text{out}}}{\omega} \right] \delta(c_i, c_j) \quad (3)$$

式中: A_{ij} 为有向边 e_{ij} 的权值; $s_i^{\text{in}} = \sum_{j \in N^{\text{in}}(i)} A_{ij}$ 为节点 i 的入边强度, $N^{\text{in}}(i)$ 为节点 i 的前驱节点的集合; $s_j^{\text{out}} = \sum_{i \in N^{\text{out}}(j)} A_{ji}$ 为节点 j 的出边强度, $N^{\text{out}}(j)$ 为节点 j 的后继节点的集合; $\omega = \sum_i \sum_j A_{ij}$ 为有向图所有边的权值之和; 其他与式(1)保持一致。

由有向图的模块性公式导出将单个节点 i 归入或移除社区 C 所获得的有向图模块性增益为

$$\Delta Q = \left[\frac{\sum_{\text{in}} + s_{i,\text{in}}}{\omega} - \frac{(\sum_{\text{tot}}^{\text{in}} + s_i^{\text{in}})(\sum_{\text{tot}}^{\text{out}} + s_i^{\text{out}})}{\omega^2} \right] - \left[\frac{\sum_{\text{in}}}{\omega} - \frac{\sum_{\text{tot}}^{\text{in}} \sum_{\text{tot}}^{\text{out}}}{\omega^2} - \frac{s_i^{\text{in}} s_i^{\text{out}}}{\omega^2} \right] \quad (4)$$

式中: $\sum_{\text{in}}^{\text{in}}$ 为社区外部节点指向社区 C 内部节点的所有边的权值之和; $\sum_{\text{tot}}^{\text{out}}$ 为社区 C 内部节点指向社区外部节点的所有边的权值之和; s_i^{in} 为连接节点 i 的入边的强度; s_i^{out} 为连接节点 i 的出边的强度; 其他与式(2)保持一致。

1.3 微博关注网络社区发现算法

在求解模块性最大化问题时引入贪心算法的思想,使得在社区发现时仅仅需要考虑单个节点的本地网络结构,即与节点相邻的社区,从而大大降低了算法的时间复杂度。算法的基本思想是,每次将节点归并到增益最大的社区,直到所划分社区的模块性不再增加为止,具体步骤如下。

步骤 1 为网络 $G(V, E)$ 中的每一个节点 $v_i \in V$ 分配一个独立的初始社区标签,记为 $c^{(l)} = \{v_i\}$, 其中 l 为社区标识号。

步骤 2 对网络中每一个节点 $v_i \in V$, 考虑其邻居节点所属社区的集合 $C^N(v_i) = \{c^{(l)} \mid v_j \in N(v_i), v_j \in c^{(l)}\}$, 根据式(4)计算将节点 v_i 从其所属社区中移除并将其归入邻居社区后的模块性增益。根据下式选择将节点 v_i 归入的社区

$$c^{(l)} = \arg \max_{c^{(k)} \in C^N(v_i)} \{\Delta Q(i, c^{(k)}) > 0\} \quad (5)$$

如果没有增益大于 0 的社区,则节点 v_i 维持原来所

属社区不变。一旦有节点被归入社区 $c^{(k)}$, 则计算将 $c^{(k)}$ 中每一个节点移除该社区所得的模块性增益,若增益大于 0,则将节点从原社区中移除。

步骤 3 重复步骤 2,直到不再有节点归并和移除为止。

步骤 4 生成社区衍生图,将步骤 2 得到的社区看作独立节点,将社区之间的边作为新的边构建新的有向图,其中边的权重为社区之间同向边的权值之和,社区内部的边用自环表示。衍生图的生成过程如图 1 所示。

步骤 5 将步骤 4 得到的衍生图应用到步骤 2、3 中,对衍生图进行社区划分,对被划分到同一社区的节点内部的子节点全部合并到同一社区。

步骤 6 重复上述步骤直到所划分社区的整体模块性不再增加为止。

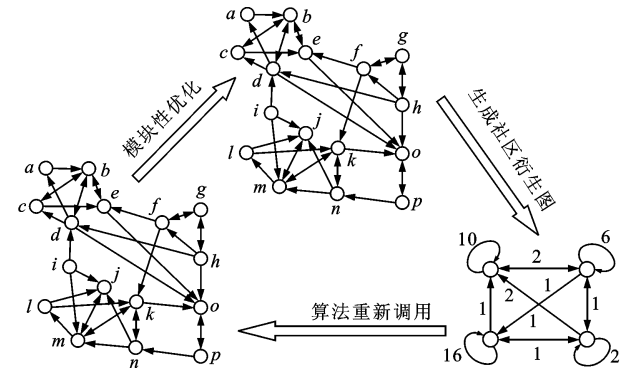


图 1 社区挖掘算法计算过程示意图

2 微博社区兴趣挖掘算法

对社区中用户发布的微博进行聚合得到社区微博集 $S_c = \{s_i^u \mid u \in c\}$, 其中 s_i^u 为用户 u 的第 i 条微博, c 为社区, 采用基于 Lucence 的 IKAnalyzer 短语抽取工具对 S_c 中的微博进行短语抽取^[13]。

对从社区中提取到的前 K ($K=100$) 条短语构建微博社区常用短语集 $T_c = \{(t, N_t^c)\}$, 其中 t 为用户微博中所含的短语, N_t^c 为短语在社区 c 中出现的次数。为了能够检测社区结构中具有代表性的短语, 采用搜索引擎中经常采用的 tf-idf 算法对常用短语集 T_c 中的短语进行评分, 首先计算话题 t 在社区 c 中出现的归一化次数

$$f_c(t) = \frac{N_t^c}{\max\{N_t^c, t \in T_c\}} \quad (6)$$

然后计算话题在所有社区中出现次数的逆向频率

$$f^i(t) = \log \frac{|C|}{| \{c \mid t \in T_c\} |} \quad (7)$$

式中： $|C|$ 为总的社区个数； $|\{c|t \in T_c\}|$ 为包含话题 t 在社区 c 中的代表数。话题 t 在社区 c 中的代表性可由 $f_c(t)$ 与 $f^i(t)$ 相乘得到

$$I_c(t) = f_c(t)f^i(t) \tag{8}$$

话题在某一社区中出现的频率越高,在其他社区中出现的频率越低, $I_c(t)$ 值越大,话题 t 在社区 c 中越具有代表性。

3 实验结果与分析

3.1 实验数据

从 2011 年 3 月到 2011 年 6 月我们爬取了新浪微博中共 70 806 个用户的用户信息及其关注列表,并利用此数据构建微博关注网络 $G(V,E)$ 。该关注网络的最大弱连接子图共有 70 806 个节点,370 748 条边。图 2 为所构建关注网络的出入度的互补累计分布函数。由图可见,微博网络中用户的关注数(出度)呈现纵尾现象,表明大多数用户仅关注较少数量的用户,而被关注数(入度)遵循幂律分布,这与其他研究结果中的结论一致^[14]。为了对所发现的微博社区进行兴趣挖掘,爬虫还爬取了所有用户最近发布的 200 条微博,并利用此数据挖掘社区用户的兴趣爱好特征。

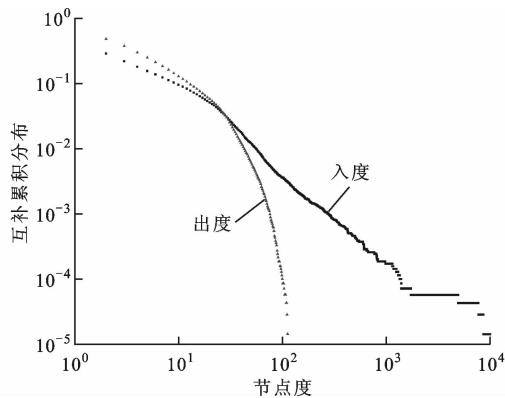


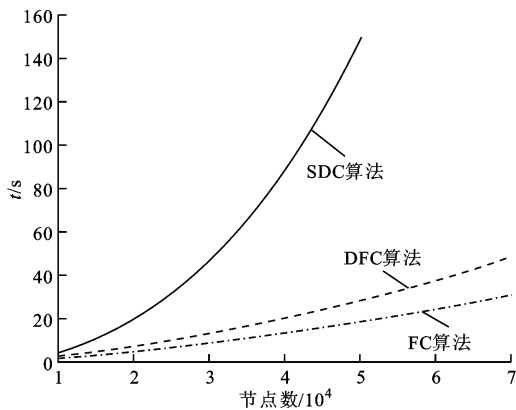
图 2 新浪微博关注网络出入度互补累积分布

3.2 社区发现算法性能评估

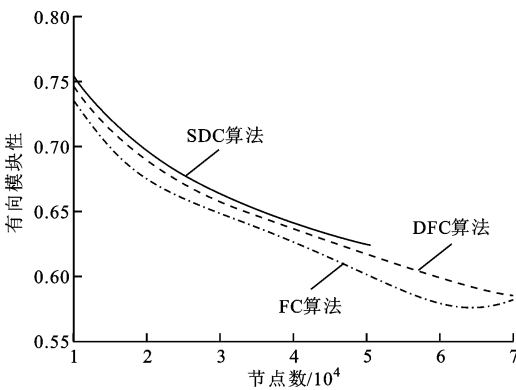
为了验证有向图社区挖掘算法(DFC)的有效性和准确性,与基于无向图的社区快速发现算法(FC)^[12]和基于谱分析的有向图社区发现算法(SDC)^[10]进行了比较。实验时对构建的关注网络进行节点采样,使得网络节点的个数按递增数列增长,最后选取采样网络的最大弱连接子图进行实验,每个网络计算 100 次,每个数据点为这 100 次实验的平均值。

图 3a 为实验中各算法所需时间与网络中节点

个数的关系。由图可知,FC 算法的计算速度最快。相比于 FC 算法,DFC 算法的计算时间有所增加,但计算时间仍几乎是随着节点个数线性增加。SDC 算法需要计算网络邻接矩阵的特征值,因此算法消耗的时间几乎随节点数量指数增加。图 3b 为算法所划分社区的有向模块性值与网络中节点个数的关系。由图可见,无论使用何种算法,所得网络社区模块性均在 0.5~0.8 之间,这表明微博网络具有很好的社区模块性,对其进行社区划分是可行的。虽然 DFC 算法所划分出的社区结构质量不及 SDC 算法,但所得结果相差很小,考虑到计算时间上的消耗,DFC 算法具有明显的优势。由以上分析可知,所提有向图社区挖掘算法有效地平衡了计算时间和准确性之间的矛盾,从而能够应用于微博类大规模有向网络的社区发现。



(a) 算法所需时间与网络中节点数的关系



(b) 社区模块性与网络中节点数的关系

图 3 社区发现算法性能评估

3.3 社区兴趣挖掘算法有效性评估

表 1 列出了由 DFC 算法所挖掘的最大的前 5 位社区的感兴趣特征:社区前 10 位的常用短语($k=10$)出现的次数以及该短语的 tf-idf 值。由表 1 可见,出现次数最多的短语不一定具有更好的代表性,

这也表明使用 tf-idf 算法对短语的有效性进行评估的必要性。各个社区之间的兴趣爱好有着很大的不同,例如社区 1 更关注“精选”“笑话”“电影”“搞笑”等内容,由此可以推断该社区中以年轻用户为主;社区 2 最常出现的短语包括“博文”“手机”“男人”“女人”“星座”等,由此可以推断该社区中的用户更关注

两性生活;社区 3 的兴趣爱好则集中于“美国”“公司”“工作”等,可以推断该社区为工作族;社区 4 更关注“路况”“孩子”“城市”等,可以推断该社区大部分用户为父母;社区 5 更关注“语录”“电子产品”以及“影评”等,可以推断其大部分社区成员为电子产品和电影爱好者。

表 1 前 5 位社区及其兴趣特征

社区 1			社区 2			社区 3			社区 4			社区 5		
短语	次数	tf-idf 值	短语	次数	tf-idf 值	短语	次数	tf-idf 值	短语	次数	tf-idf 值	短语	次数	tf-idf 值
精选	15 597	0.016 7	博文	3 925	0.006 6	美国	2 706	0.011 6	路况	1 665	0.019 0	语录	2 249	0.051 0
笑话	15 855	0.014 3	手机	3 887	0.005 5	公司	2 820	0.010 4	孩子	1 311	0.008 1	电子产品	1 736	0.039 2
排行榜	16 511	0.012 4	男人	5 188	0.004 8	北京	3 569	0.010 1	城市	1 174	0.005 6	影评	1 624	0.036 7
兔子	12 341	0.008 4	人生	4 891	0.004 7	发表	2 442	0.009 0	手机	1 072	0.005 3	经典语录	1 619	0.036 6
电影	12 195	0.007 5	世界	4 170	0.004 3	老师	2 369	0.008 7	人生	1 210	0.004 1	电影	3 078	0.035 8
奥特曼	12 040	0.007 4	活动	7 281	0.004 0	博文	3 205	0.008 6	星座	1 222	0.002 2	家居	1 551	0.035 0
搞笑	16 013	0.007 2	女人	6 181	0.003 0	孩子	2 823	0.008 1	中国	1 770	0.002 0	电子	1 048	0.023 7
投票	12 672	0.006 7	关注	5 434	0.002 5	工作	2 655	0.007 5	女人	1 102	0.001 9	汽车	1 037	0.023 4
害羞	11 582	0.006 1	可爱	3 820	0.002 2	城市	2 501	0.005 5	全球	1 399	0.001 8	娱乐	944	0.021 3
加油	11 274	0.005 5	星座	4 173	0.002 1	人生	2 855	0.004 4	爱情	1 440	0.001 7	养生	815	0.018 4

4 挖掘算法在僵尸粉检测中的应用

在社会网络中,被同一个人或组织控制的虚假帐号往往会紧密地连接在一起,Viswanath 等发现无向网络中的大多数恶意用户检测算法可以通过社区发现算法来实现^[3]。僵尸粉是微博系统中的虚假粉丝,通常是由某些应用自动产生的恶意注册用户。这些账户往往被单一的个人或组织控制,因此它们之间往往会紧密地相互连接。由僵尸粉构成的社区与兴趣小组社区有着很大的区别。首先,共同兴趣小组构成的子图具有小世界特性,而这些僵尸粉构成的网络子图则随机连接且紧密,有时甚至是全连接的。其次,僵尸粉社区从整体上看非常不活跃,即使有发布的微博,这些微博也往往非常相似,话题单一,而共同兴趣小组则非常的活跃,且所发布的微博形式多样,内容丰富。这就使得通过社区兴趣发现能够将共同兴趣小组与僵尸社区区分开来,从而达到检测的目的。

图 4 为所发现社区的衍生图,图中每个节点代表一个社区。圆圈标识出了高度可疑的社区,这些社区内部连接紧密而与外部其他社区连接较少,具备僵尸粉网络的基本特征。进一步对这些社区的常用短语挖掘发现,61 号社区的常用短语只有“试试”“UC 浏览器”“发出”“微博”等少数关键字,该社区

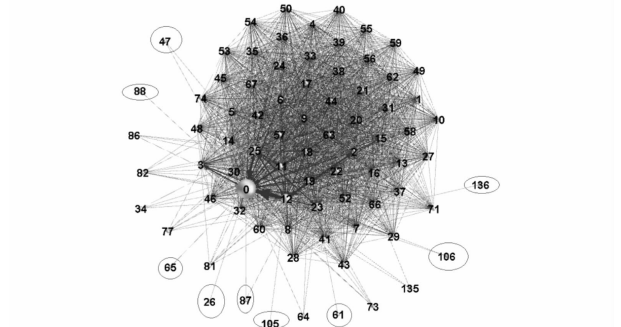


图 4 僵尸粉社区检测示意图

中共有 80 个用户,且所有用户都仅仅发布了同样的一条微博:“酷~试试一条从 UC 浏览器发出的微博”,进一步的调查显示这 80 个用户的用户名均为“U 友+数字”形式,由此可以推断该社区中的用户通过 UC 浏览器自动注册微博帐号,并自动添加其他以同样形式注册的用户为好友,从而形成典型的僵尸粉社区。这表明所提方法能够批量地发现僵尸粉用户,为微博网络的管理提供支持。

5 结 论

为了高效地挖掘微博网络中的用户兴趣小组,本文提出了一种有向图社区挖掘算法。该算法考虑了边的方向性信息,与无向图社区挖掘算法相比,较好地平衡了社区挖掘准确率和运行效率之间的矛盾。

盾,能够应用于大规模微博网络中社区的发现和提取。为了能够有效地提取社区特有的常用短语以便迅速发现社区特有的兴趣爱好,引入了 tf-idf 算法对所挖掘社区的前 100 位常用短语进行评分,实验结果表明,算法可以较好的提取社区用户的兴趣爱好特征。此外,实验发现有些兴趣小组的话题单一,社区成员之间的连接非常紧密且发布的微博内容几乎完全相同,具有僵尸粉社区的基本特征。这表明算法能够发现连接紧密的僵尸粉社区网络,从而适用于对微博网络中僵尸粉的检测。

参考文献:

- [1] JAVA A, SONG X, FININ T, et al. Why we twitter: understanding microblogging usage and communities [C]//Proceedings of the Joint 9th WebKDD and 1st SNA-KDD workshop. Berlin, Germany: Springer, 2007: 56-65.
- [2] AHN Y Y, BAGROW J P, LEHMANN S. Link communities reveal multiscale complexity in networks [J]. Nature, 2010, 466(7307): 761-764.
- [3] VISWANATH B, POST A, GUMMADI K P, et al. An analysis of social network-based Sybil defenses [J]. ACM SIGCOMM Computer Communication Review, 2010, 41(4): 363-374.
- [4] 杨楠,弓丹志,李欣,等. Web 社区发现技术综述 [J]. 计算机研究与发展, 2005, 42(3): 439-447.
YANG Nan, GONG Danzhi, LI Xian, et al. Survey of web communities identification [J]. Journal of Computer Research and Development, 2005, 42(3): 439-447.
- [5] LEICHT E A, NEWMAN M E J. Community structure in directed networks [J]. Physical Review Letters, 2008, 100(11): 118703.
- [6] NICOSIA V, MANGIONI G, CARCHIOLO V, et al. Extending the definition of modularity to directed graphs with overlapping communities [J]. Journal of Statistical Mechanics: Theory and Experiment, 2009, 2009(3): 03024.
- [7] LEVORATO V, PETERMANN C. Detection of communities in directed networks based on strongly p-connected components [C]//Proceedings of the Conference on Computational Aspects of Social Networks. Piscataway, NJ, USA: IEEE, 2011: 211-216.
- [8] WANG Liaoruo, LOU Tiancheng, TANG Jie, et al. Detecting community kernels in large social networks [C]//Proceedings of the 2011 11th IEEE International Conference on Data Mining. Piscataway, NJ, USA: IEEE, 2011: 784-793.
- [9] GIRVAN M, NEWMAN M E. Community structure in social and biological networks [J]. Proceedings of the National Academy of the Sciences of the United States of America, 2001, 99(22): 7821-7826.
- [10] NEWMAN M E J. Detecting community structure in networks [J]. The European Physical Journal B-Condensed Matter, 2004, 38(2): 321-330.
- [11] DUCH J, ARENAS A. Community detection in complex networks using extremal optimization [J]. Physical Review: E, 2005, 72(2): 027104.
- [12] BLONDEL V D, GUILLAUME J L, LAMBIOTTE R, et al. Fast unfolding of communities in large networks [J]. Journal of Statistical Mechanics: Theory and Experiment, 2008, 2008(10): 10008.
- [13] 张云. 基于开源软件的中文学术文献计量软件的开发实践 [J]. 现代图书情报技术, 2010, 4(191): 87-91.
ZHANG Yun. The Practice on the development of software on the Chinese academic bibliometrics based on the open source software [J]. New Technology of Library and Information Service, 2010, 4(191): 87-91.
- [14] 赵文兵,朱庆华,吴克文,等. 微博客用户特性及动机分析 [J]. 现代图书情报技术, 2011(2): 69-75.
ZHAO Wenbing, ZHU Qinghua, WU Kewen, et al. Analysis of micro-blogging user character and motivation [J]. New Technology of Library and Information Service, 2011(2): 69-75.

(编辑 武红江)