

# 基于公开数据的特殊人群 在线活动特征挖掘

刘楚楚<sup>1</sup>, 吕欣<sup>1,2,3</sup>

1. 国防科技大学系统工程学院, 湖南 长沙 410073; 2. 中南大学商学院, 湖南 长沙 410083;  
3. 卡罗林斯卡研究所公共卫生科学系, 瑞典 斯德哥尔摩 17177

## 摘要

通过采集百度贴吧中的相关数据,从时间、文本、网络3个维度分析不同人群的在线活动特征,探索社会特殊人群的活动规律和心理状态。研究发现,艾滋病人群在线活动更规律,在线活动的主要目的是了解病情知识、记录日记,对本类型贴吧的关注很集中;男同性恋人群活跃时间晚,大多是为了娱乐交友,对本类型贴吧的关注具有很强的偏好性。总体上,艾滋病人群对自身病情十分关心,但男同性恋人群的艾滋病防护意识还不强。

## 关键词

特殊人群;艾滋病;男同性恋;在线特征;复杂网络

中图分类号:G350,R512.91

文献标识码:A

doi: 10.11959/j.issn.2096-0271.2018047

## *Analysis of online activity characteristics of hidden populations based on public data*

LIU Chuchu<sup>1</sup>, LU Xin<sup>1,2,3\*</sup>

1. College of Systems Engineering, National University of Defense Technology, Changsha 410073, China  
2. School of Business, Central South University, Changsha 410083, China  
3. Department of Public Health Sciences, Karolinska Institute, Stockholm 17177, Sweden

## Abstract

Data from 36 Baidu Tieba was collected. The sub-groups of users from three dimensions of time (temporal), content (text) and interaction (network) were analyzed, and patterns of their online activity were tried to extract and characteristics of different populations was inferred. The result indicates that the temporal pattern of posting behavior is more regular for HIV-related users. Communities followed by these users are also HIV-related, revealing a significant clustering pattern. On the other hand, MSM groups are more active during midnight, and their motivation of online activities in the community is for entertainment and for meeting partners. There is also a strong preference of following the same type of Tiebas for MSM groups. In general, HIV-related users are very concerned about their health status. On the contrary, MSM-related users are lack of awareness for protection and prevention of AIDS.

## Key words

hidden population, AIDS, MSM, online features, complex network

\* 作者在国内外发表的论文均采用LU

1 引言

“特殊人群”一般是指那些因主客观原因导致的在社会竞争中处于不利形势的“边缘”人群<sup>[1]</sup>。在我国,艾滋病(AIDS)人群和男同性恋(men who have sex with men, MSM)人群承受了更大的社会环境压力<sup>[2-4]</sup>,拥有较高的社会距离和社交限制分值<sup>[5]</sup>,在社会竞争中处于不利地位,是我国具有代表性的两类特殊人群。由于多种原因,男同性恋人群感染艾滋病的概率比普通人群高19.3倍<sup>[6]</sup>,现存活的艾滋病患者中经男男性传播的比例大约为14.7%<sup>[7]</sup>,且近年来呈明显的上升趋势。艾滋病病毒(HIV)感染者和MSM人群属于艾滋病传染的高危人群,对这两类人群的行为特点进行研究,将对艾滋病的预防控制具有非常重要的意义。然而,由于受到社会环境压力等因素的影响,全面地、有代表性地对这两类特殊人群进行调查研究存在着许多困难<sup>[8-9]</sup>。目前对HIV感染者和MSM人群的研究还主要停留在基于实地或网络人群抽样的现场访谈和问卷调查,这种传统方法效率低、样本有限,且特殊人群的隐蔽性使研究的开展受到颇多限制<sup>[10-12]</sup>。

随着互联网技术的发展,人们的社交生活发生了巨大的改变,人们频繁地在各种虚拟社区发布、传递、共享信息,使网络虚拟社区中蕴含了海量的用户活动数据,为研究特殊人群提供了新的突破口。张建等人<sup>[13]</sup>通过对中国生殖健康网站“中国红丝带网”上直播的节目记录文件进行分析,了解网民对待艾滋病的心理、行为和信息需求。艾德华<sup>[14]</sup>通过对“同性爱”网站的观察与分析,了解“同性爱”网站的网民的信息需求和交流情况。白冠男等人<sup>[15]</sup>选取“艾滋病吧”中的1 746条回帖作为研究对

象,采用内容分析法,了解“艾滋病吧”中网民对艾滋病的态度及其主要健康问题。王国华等人<sup>[16]</sup>则采用内容分析法和社会网络分析法,以百度贴吧“HIV吧”中400条贴文作为研究对象,对艾滋病相关人群的社会支持信息和社会支持关系网进行了分析。Constantinos C K<sup>[17]</sup>对一个HIV/AIDS在线社区的帖子进行内容分析,评估艾滋病群体中社会支持交流的类型和比例。同样,Mo P K等人<sup>[18]</sup>证明了在艾滋病群体的所有社会支持类型中,信息支持和情感支持占比最大。通过分析微博中艾滋病群体发布的博文内容,Shi J等人<sup>[19]</sup>探讨了微博中艾滋病群体的社会支持内容及结构;而Guo Y 等人<sup>[20]</sup>则试图了解艾滋病群体在线社会支持的动态交互过程以及随时间的衍变情况。网络社区中特殊人群的活动数据能够真实反映出该群体的在线活动情况,通过分析这些数据可以有效地挖掘出特殊人群的行为模式。总体而言,目前国内外对特殊人群在网络社区中开源数据的研究,很大一部分集中在对特殊人群的社会支持<sup>[21-27]</sup>上,并且大多使用的是内容分析的方法,对特殊人群在线多维特征(如在线活动的时间统计特征、社交网络特征、社团效应、情感倾向等)的研究还很少。

本文聚焦百度贴吧中与HIV和MSM相关的两类贴吧,分析贴吧的用户活动数据,挖掘HIV感染人群和MSM人群的在线行为特征。从时间、内容、网络3个维度出发,分析特殊人群的在线活动时序规律、话题偏好以及由关注关系构成的共同关注者网络的特征,并以新闻时事类贴吧用户作为对照,深入了解社会特殊人群与普通人群的行为差异,从而全面了解特殊人群的在线活动模式,为社会公共卫生管理提供指导,具有重要的创新意义和现实价值。

## 2 数据来源

百度贴吧吸引了大量基于共同兴趣爱好的社会群体。自2003年12月份正式上线以来,百度贴吧的用户数目呈现出爆炸性增长,目前百度贴吧的活跃用户数已达3亿人,贴吧数目超过2 000万个,其中蕴藏的海量数据具有巨大的挖掘价值。本研究运用Python Scrappy爬虫框架采集百度贴吧中与艾滋病、男同性恋、新闻时事相关的最活跃的36个贴吧,贴吧名称及发帖数(主帖、评论贴及回复的总数)见表1。采集的数据内容包括贴吧信息、贴吧会员(贴吧关注者)信息和用户发帖信息,采集后的数据存入PostgreSQL数据库。贴吧信息包括贴吧名称、贴吧主页URL地址、贴吧话题类型、关注者人数(会员数目)、贴吧帖子总数、贴吧标语、贴吧所在目录以及该贴吧的友情贴吧名称。受百度防爬策略的约束,只能访问每个贴吧的前458页会员信息,即最新加

入的前10 992个会员,共采集了270 229条会员数据,内容包括会员昵称、会员主页URL地址、会员关注的贴吧。本研究采集了截至2016年8月1日的36个贴吧的所有发帖信息,共6 316 158条。发帖信息包括帖子的标题、帖子所在主页的URL地址、帖子创建时间、帖子被回复数、帖子所属贴吧名、发帖人昵称、发帖人主页URL地址以及发帖人等级。为了避免“水军”“僵尸”等对分析结果的影响,在数据分析过程中对贴吧中的用户进行了筛选,主要去除了发帖数少于3的不活跃用户以及帖子长度少于3个字符的文本。

## 3 研究结果

### 3.1 在线活动时间统计特征

#### 3.1.1 日分布规律

对各类型贴吧每天的发帖时间分布进行比较,可以发现,艾滋病类型和男同性

表1 贴吧名称及发帖数

| 艾滋病类     |           | 男同性恋类        |             | 新闻时事类    |            |
|----------|-----------|--------------|-------------|----------|------------|
| 吧名       | 帖子数       | 吧名           | 帖子数         | 吧名       | 帖子数        |
| hiv吧     | 3 505 723 | gay吧         | 244 202 118 | 每日新闻热点吧  | 13 705 695 |
| 艾滋病论坛吧   | 989 457   | 同志吧          | 5 939 248   | 新闻直播间吧   | 498 963    |
| 恐艾吧      | 7 295 218 | 同性恋吧         | 12 380 075  | 新闻夜航吧    | 335 937    |
| 脱恐吧      | 1 132 752 | bl吧          | 47 367 563  | 名人新闻吧    | 493 953    |
| 艾滋病患者吧   | 253 787   | 同爱吧          | 1 283 706   | 热点吧      | 364 559    |
| 艾滋病吧吧    | 104 446   | 基友吧          | 3 791 858   | 新闻吧      | 210 699    |
| hiv论坛吧   | 14 533    | 腐吧           | 2 513 171   | 时事新闻吧    | 33 540     |
| hiv爱天使吧  | 26 470    | 中国gayles反歧视吧 | 297 624     | 新闻消息吧    | 16 767     |
| hiv综合论坛吧 | 7 113     | 基友你永远不会独行吧   | 113 059     | 热点话题吧    | 14 821     |
| hiv心窗吧   | 3 804     | 基友头像吧        | 105 360     | 搜狐新闻吧    | 27 303     |
| hiv携带者吧  | 4 934     | 中国好基友吧       | 46 136      | 搜狐新闻客户端吧 | 42 232     |
|          |           | 腐男群吧         | 67 650      |          |            |
|          |           | 腐女腐男同盟会吧     | 90 258      |          |            |
|          |           | 腐女vs腐男吧      | 619 682     |          |            |

恋类型中各贴吧每天的发帖趋势基本一致，新闻时事类各贴吧的发帖时间分布差异相对较大。艾滋病人群和男同性恋人群在发帖时间上的规律性更强，其内部活动时间更一致。不同类型人群的日发帖规律也存在差异，如图1(a)所示。艾滋病人群在每晚22:00—23:00达到发帖顶峰，凌晨3:00—5:00发帖少；男同性恋人群在每晚的0:00左右达到发帖顶峰，到早上5:00才跌至波谷；新闻时事类人群（普通人群）在一天中会出现多个发帖高峰，大多位于早上7:00—8:00、中午12:00—14:00和晚上19:00—20:00，凌晨2:00—4:00发帖少，在5:00左右回升。值得注意的是，普通人群在线活跃状况从晚上20:00开始出现衰减趋势，而此时特殊人群（艾滋病人群和男同性恋人群）活跃程度正处于上升阶段，一直持续到凌晨，且男同性恋人群在凌晨2:00的发帖比普通人群中午12:00还要活跃，反映了特殊人群在线活动时间更偏好于晚上甚至深夜。这与不同人群在发帖内容上表现出来的差异吻合。艾滋病人群发帖大多围绕着病情咨询，这类人群更关注身体和作息，发帖高峰位于每晚的22:00左右，比男同性恋人群要早一些。男同性恋人群发帖大部分是关于交友、聊天、消遣，因此活跃时间在3类人群中最晚，衰减趋势也表现得更滞后。新闻时事类人群发帖的内容主要围绕政治、经济、社会等话题，发帖多位于一天中的休息时间，且与电视新闻的播放时间一致。艾滋病类贴吧和男同性恋类贴吧在一天中的发帖波动很一致，可以认为与普通人群相比，这两类人群在线活动的时间规律表现出更高的相似性。

3.1.2 季节性特征

分析各贴吧每年在各月份的发帖量，发现同类型贴吧的发帖波动基本一致，但

各个类型贴吧的发帖量没有表现出随月份明显的季节性波动，不同年份的月发帖量分布没有明显的周期性，但具有一些相似性特征。hiv吧每年的12月份是一个波峰，2月份位于波谷位置，3月份开始回升（每年的1、2月份是中国农历年末，春节到来导致发帖量下降）。gay吧近几年在各月份上的发帖波动较一致，上半年比下半年发帖量大，且随着暑期的结束，每年9月份的发帖量会比8月份有所降低（已有研究证实，在MSM虚拟社区中在校学生所占比例较大，而且MSM人群中青少年占有相当的比例<sup>[11,28-29]</sup>）。新闻直播间吧发帖跟每个月实际发生的新闻事件相关，但每年的2月份（农历年末）位于发帖的波谷位置。

3.1.3 长期特征

通过统计不同人群每年的发帖量，可以看出各类人群的发帖趋势，如图1(b)所示。艾滋病人群的年发帖量从2008年开始一路上升，2013年的年发帖量增长速度达到最大，2015年的年发帖量达到顶峰（2016年数据不完全）；男同性恋人群的年发帖量从2004年开始呈上升趋势（百度贴吧于2003年12月正式上线），2008年达到一个波峰后开始衰减，2010年又上升，至2013年达到年发帖量的最高值，随后跌落（可能由于近年来百度贴吧对内容管控逐步加强，且男同性恋交友网站大量增加）；新闻时事类普通人群的发帖在2013年呈现出爆发性增长，此前年发帖量一直很低，这可能与2013年百度贴吧用户数目大量增长有关。

2008年，男同性恋人群发帖量迎来一个小高峰，此时艾滋病人群发帖量开始增加；2013年男同性恋人群年发帖量达到最大值，此时艾滋病人群的发帖增长速度达到最大。排除重合用户影响的可能性（共同

用户数目远小于这两类贴吧各自的数目), 艾滋病人群和男同性恋人群的在线活跃情况存在着一定程度的相关性。由于多种原因, 男同性恋人群感染艾滋病的可能性远远高于普通人群, 因此随着男同性恋人群的增加, HIV在人群中快速传播, 艾滋病人群随之快速增长。已有的研究<sup>[30-34]</sup>已证明男同性恋人群是艾滋病传播的高危人群和重要桥梁, 这与本研究得出的结果相一致。

### 3.1.4 时间间隔特征

统计各贴吧发帖时间相邻的各帖子之间的时间差, 发现各贴吧前后发帖的时间间隔呈现明显的幂率分布<sup>[35]</sup>, 如图1(c)所示。绝大多数情况下后一个帖子的创建时间与前一个帖子创建时间的时间差很小, 少数帖子的创建时间与前一个帖子的创建时间间隔很大。

## 3.2 在线内容挖掘及热点分析

### 3.2.1 热词发现

本文用词项的TF-IDF (term frequency-inverse document frequency) 值<sup>[36]</sup> (即词频乘以逆向文本频率) 来定义一个词的热度。提取贴吧中帖子的标题信息, 对标题文本进行分词, 剔除常用词, 计算各个词项的TF-IDF值, 选取前100个TF-IDF值最大的词代表该贴吧的热词, 比较不同类型贴吧的热词区别, 分析不同类型人群的话题特点及差异。

通过比较各贴吧的热词可以发现, 不同类型人群发帖的热词差异较大 (如图2所示), 不同人群在线发言存在着明显的话题偏好。具体表现为, 艾滋病人群的发帖内容主要围绕着病情咨询和日记记录, 男同性恋类人群的话题大多与交友聊天和

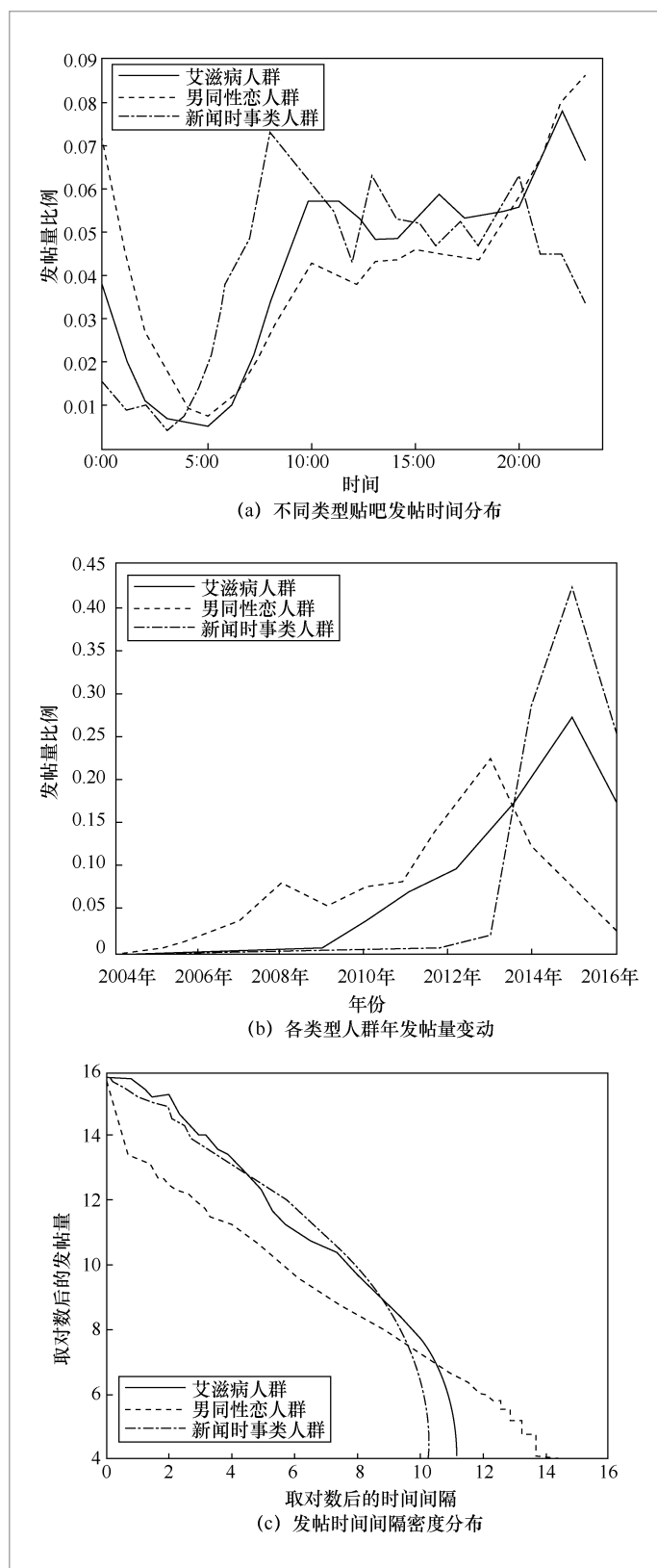


图1 各类型贴吧在线活动时间统计特征

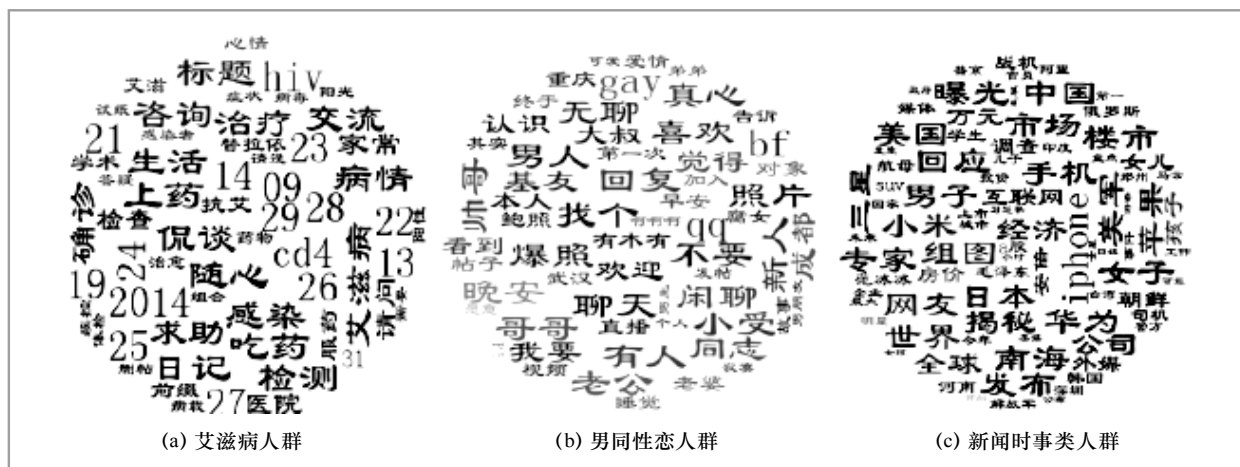


图2 不同人群发帖热词差异

感情生活有关。新闻时事类人群的讨论热点主要集中在政治、经济和社会话题。各类型人群话题分布如图3所示。

研究发现,不同类型人群之间的发帖热词相似度(热词交集比例)也存在差异。艾滋病人群与男同性恋人群发帖热词的相似度达到36%,高于与新闻时事类的热词相似度(12%),艾滋病人群与男同性恋人群在话题偏好上相比于普通人群具有更高的

相似性。在共同关注网络(见第3.3节)中,艾滋病类型贴吧与男同性恋类型贴吧之间的连接数目为1 319,大于与新闻时事类贴吧的连接数目656,说明艾滋病人群和男同性恋人群的重合人数更多,导致这两种人群的话题相似度要大于新闻时事类人群。

### 3.2.2 热词动态演变分析

研究发现,不同类型人群每年的发帖热词都会发生变化,每年都会有新增的热词(如图4所示),反映了不同人群的心理状态变迁。

艾滋病人群在2011年的发帖热词中新出现了“贱狗们”“诅咒”等表达负面情绪的词汇;2012年的热词中出现了大量艾滋病检测产品品牌;2013年关于艾滋病治疗、药品名称以及患者情绪的词汇增加;2014年艾滋病人群讨论的话题变得轻松,关于病情交流和经验分享的话题增加,出现了大量日记帖;2015年生活话题增多,目前国内最好的艾滋病治疗药物“替拉依”成为热词之一;在2016年,“工作”也成为艾滋病病友们的热点讨论话题。可以看出,艾滋病人群的话题逐步向积极健康的方向转变,艾滋病人群的心理状态从最初的紧张

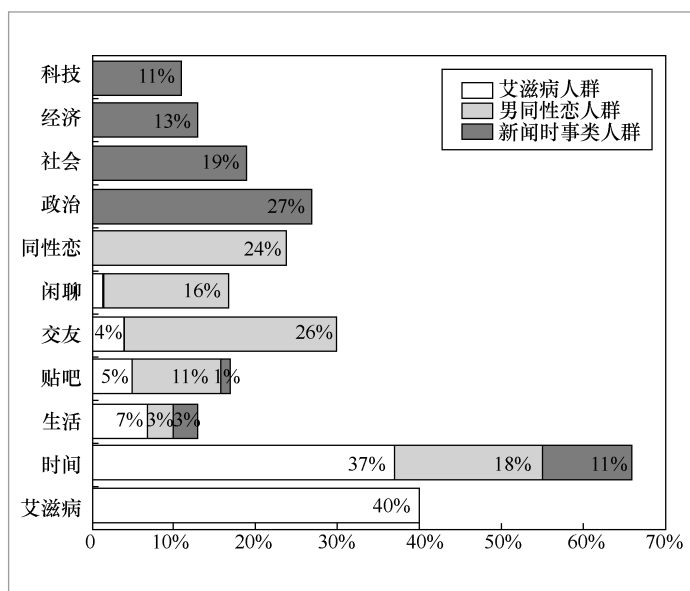


图3 不同人群话题比较

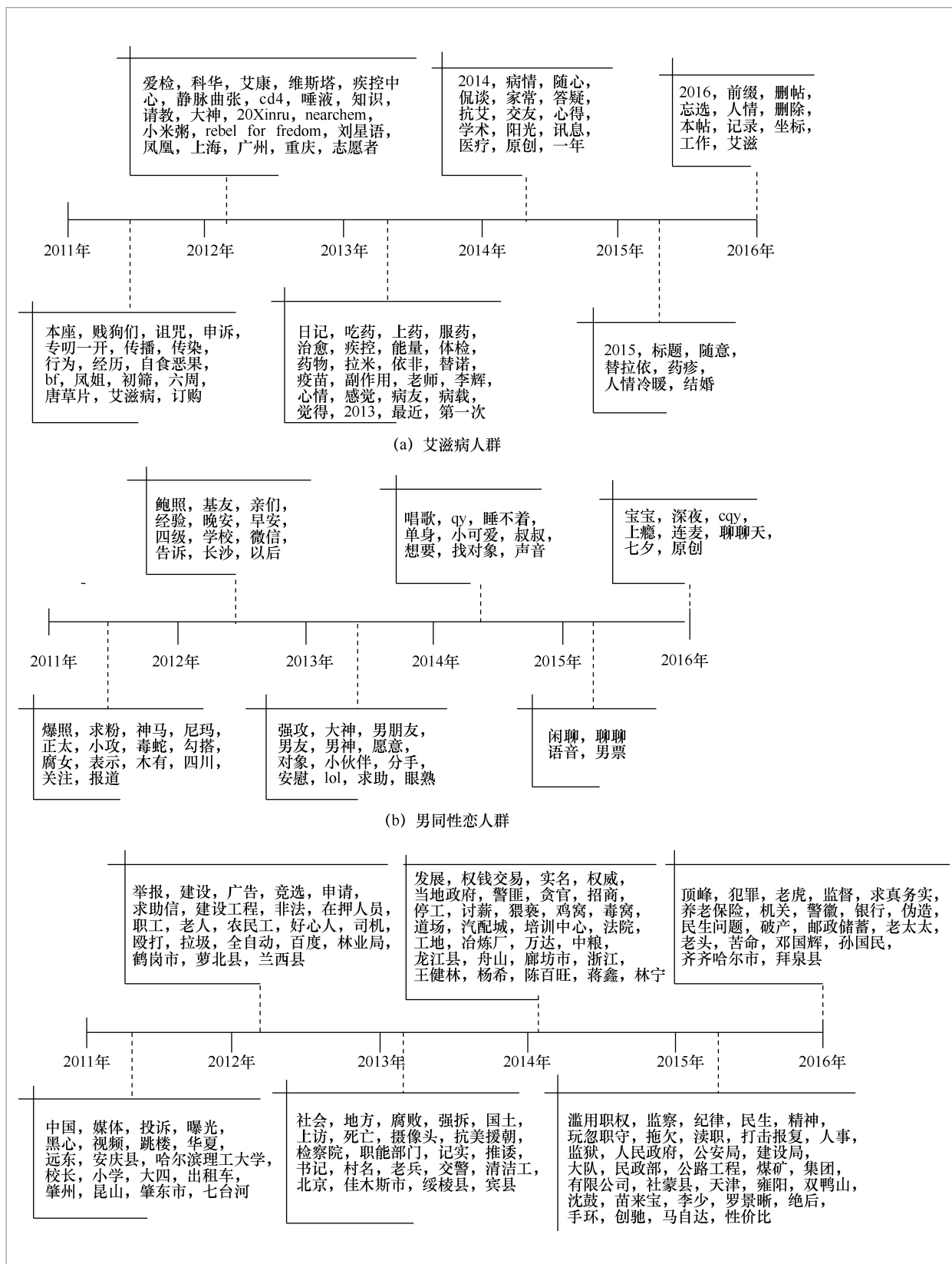


图 4 不同人群发帖热词变迁

抵触变得越来越平和轻松。笔者在后续的研究中用基于规则的情感分析方法,通过构建相应的情感词库,对艾滋病群体的情绪特点做了系统分析,发现在艾滋病社区中负面情绪仍偏多,主要表现为初期感染HIV的患者对病情的恐惧、担忧和焦虑。但是,积极情绪也占据了相当大的比例,大多关于AIDS的及时诊断与积极治疗,表达诸如信心、努力、坚强、感激等情绪。而且出于隐私保护,艾滋病人常常选择通过社交平台寻求帮助和建议<sup>[37]</sup>。

男同性恋人群每年的发帖热词变化也十分显著,每年的热词基本上与该年的网络流行词汇息息相关,如2011年出现了“神马”“正太”“腐女”等词;2012年新增了“基友”“微信”;2013年出现了“大神”;2014年出现的“叔叔”“qy”;2015年出现的“男票”“语音”;2016年出现的“宝宝”。此外,在2013年,关于同性恋的词汇大量增加,这与前文(第3.1节)中2013年男同性恋发帖数达到峰值的结论相吻合。而新闻时事类人群每年的发帖热词差异较大,话题主要与每年发生的热点事件和新闻相关。

3.3 贴吧共同关注者网络

3.3.1 共同关注者网络可视化

从网络角度挖掘特殊人群的在线活动特征,利用3种类型全部贴吧的会员数据构造共同关注者网络,各贴吧作为网络的节点,如果两个贴吧之间存在共同关注者,则对应贴吧节点之间就形成一条连接边,共同关注者数目作为边的权重。权重越大,说明两个贴吧之间的共同关注者数目越多。可视化有助于直观了解一个网络的特征。对3类贴吧的共同关注者网络进行可视化处理后的结果如图5所示,节点大小代表各贴吧度的大小,边的权重表现为边的粗细。可

以看出,该网络存在着明显的社区结构,3个类型贴吧的内部联系紧密,与外界联系相对少,不同人群对本类型贴吧更关注。3类贴吧的用户对本类型贴吧相关的主题明显更为关心,说明在艾滋病类型贴吧和男同性恋类型贴吧中活跃的用户,很大可能就是笔者要研究的特殊人群。

3.3.2 各类型贴吧的共同关注者网络差异比较

为了了解不同人群在贴吧关注上的特点,构造并比较了各类型贴吧的共同关注者网络(如图6所示),可以发现,3类人群对本类型贴吧的共同关注网络表现出很大的相似性,即都是全连接网络,且网络内部边权远远大于外部边权,各类人群对本类型贴吧更关注。但3个网络在多样性、内外连接数量比例、平均加权重等方面表现出较大的差异。

3类人群的共同关注网络社区内部连边(此处一条边代表两个贴吧之间的一个共同用户)数量与外部连边数量之比分别为14.2(艾滋病类)、8.9(男同性恋类)和8.1(新闻时事类)。各社区的内外连接数量比都远大于1,印证了这3类贴吧显著的社区结构。为了了解不同人群的贴吧关注关系特点,笔者使用基于Shannon熵的社会多样性度量<sup>[38-39]</sup>来刻画各个关注网络的多样性。

$$D_{\text{social}}(i) = \frac{-\sum_{j=1}^k p_{ij} \ln(p_{ij})}{\ln(k)} \quad (1)$$

其中, $k$ 是节点 $i$ 的连接数目, $p_{ij}$ 是邻居 $j$ 的边权与其所有邻居边权之和的比值。3个网络的多样性平均值分别为0.88(艾滋病类)、0.69(男同性恋类)和0.75(新闻时事类)。各类人群的关注网络在平均加权重和模块化指数上也存在差异。艾滋病网络的平均加权重为5 102.73,大于男同性



间具有更多的共同用户,说明艾滋病人群与男同性恋人群之间的重叠相比于普通人群更加突出,男同性恋人群相比普通人群更有可能携带或者感染了HIV,这也从侧面验证了Koblin B A等人<sup>[41]</sup>和Liu H等人<sup>[42]</sup>关于男同性恋患艾滋病的可能性要高于普通人的结论。

## 4 结束语

由于特殊人群的隐蔽性,传统方法无法有效获取这类人群真实的活动数据,难以了解特殊人群的行为特点和心理状态。本文通过分析百度贴吧特殊人群的在线活动数据,首次对艾滋病人群和男同性恋人群与普通人群(新闻时事类贴吧用户)的在线活动差异进行挖掘,以了解特殊人群在心理、社交上的特征。研究发现,我国艾滋病人群和男同性恋人群在在线活动时间、讨论话题、关注贴吧等各方面都表现出各自明显的特点。艾滋病人群在每晚22:00—23:00达到活动高峰,凌晨3:00—5:00最不活跃;男同性恋人群在每晚0:00—1:00达到活动高峰,在早上5:00跌至波谷。且艾滋病类人群在贴吧的活跃程度呈上升趋势,男同性恋人群的活跃程度呈下降趋势。艾滋病人群在线活动的主题主要是病情咨询和日记记录,男同性恋人群大多与交友聊天和感情生活有关。艾滋病人群对本类型贴吧的关注更多、更集中,一个用户同时关注多个艾滋病贴吧,且艾滋病人群在贴吧中的分布更均匀;男同性恋人群对本类型贴吧的关注较集中,但用户在贴吧中的分布很不均匀,大多集中在gay吧和bl吧。

总体上,男同性恋人群与艾滋病人群的在线活跃情况存在着一定程度的相关性。并且,相比普通人群,男同性恋人群

与艾滋病人群之间的关联性更强,男同性恋人群感染HIV的概率明显高于普通人群<sup>[41-42]</sup>。在分析结果中可以看出,艾滋病人群对自身病情比较关心,但男同性恋人群的艾滋病防护意识并不强,且网络男同性恋人群的年龄都偏低,有很大一部分是青少年<sup>[28-29]</sup>。

根据百度贴吧中男同性恋相关人群和艾滋病相关人群表现出来的特点,对公共卫生有关管理部门提出以下建议,以期能为我国特殊群体提供更好的关怀和管控以及更全面的教育和指导。

- 针对男同性恋人群艾滋病高危的特点,应加强男同性恋人群中艾滋病防护意识的培育和相关安全知识的宣传,促进男同性恋人群自我保护意识的提升,减少并尽可能杜绝高危行为的发生。
- 由于社会舆论环境,我国艾滋病群体面临着很大的社会歧视,因此表现出极强的隐蔽性。出于网络社区等虚拟平台的匿名性保证,很大一部分艾滋病患者倾向于通过互联网渠道查找资料或寻求帮助。相关管理部门可以积极利用互联网虚拟社区、论坛网站等虚拟平台,加强艾滋病基础知识的普及,多渠道扩展艾滋病咨询和检测通道,为广大艾滋病群体提供更方便、更人性化的服务。
- 目前,男同性恋人群有低龄化发展的趋势,很大一部分男同性恋者是青少年,并且近年来我国大中学生中的HIV感染者超过80%是男同性恋,因此提早对青少年开展安全性知识教育,提高其防患意识,有利于促进青少年群体对艾滋病病毒的预防。

基于开源数据的特殊人群研究,能最直接地从特殊人群在线活动产生的数据中发现问题,而已有的关于特殊人群(如艾滋病人群、男同性恋人群)的研究大多采用传统调查分析的方法。本文克服了在传

统调查中由调查对象的主观因素带来的误差,可以最真实、客观地反映特殊人群的行为特点和动机偏好,对以抽样调查为主的特殊人群研究方法是一个很大的补充。本研究对特殊人群和普通人群的在线行为模式进行多维度分析比较,对艾滋病人群和男同性恋人群的在线活动特点进行了多视角、直观的呈现,同时也为今后其他类型特殊人群的研究提供了新的思路,有着重要的创新意义和现实价值。但是受到数据的约束,此次研究只探讨了在百度贴吧中活跃的群体,对其他在线社区的相关特殊人群的特点并未做分析,今后的研究会立足在多种社交平台上的数据分析,增强研究的完整性和可靠性。

## 参考文献:

- [1] 王瑞山. 试论特殊人群的制度化排斥及其应对[J]. 上海师范大学学报(哲学教育社会科学), 2013, 45(3): 92-96.  
WANG R S. The institutional obstacles in the rehabilitation of special population and their countermeasures[J]. Journal of East China Normal University: Philosophy and Social Sciences Edition, 2013, 45(3): 92-96.
- [2] 戴丽萍, 姜宜海, 龚诚华, 等. MSM艾滋病相关知识行为及互联网的使用情况调查[J]. 中国健康教育, 2012, 28(12): 1026-1028.  
DAI L P, JIANG Y H, GONG C H, et al. Study on Internet use and AIDS related knowledge and behaviors among MSM[J]. Chinese Journal of Health Education, 2012, 28(12): 1026-1028.
- [3] 王毅, 张洪波, 徐杰, 等. 男男性行为者社会歧视经历调查[J]. 中国公共卫生, 2011, 27(1): 91-92.  
WANG Y, ZHANG H B, XU J, et al. Social discrimination experience among MSM[J]. Chinese Journal of Public Health, 2011, 27(1): 91-92.
- [4] 李现红, 刘立珍, 何国平, 等. 男男性行为HIV感染者艾滋病相关羞辱和歧视的调查分析[J]. 中华护理杂志, 2013, 48(2): 142-146.  
LI X H, LIU L Z, HE G P, et al. HIV/AIDS related-stigma and discrimination among HIV positive men who have sex with men[J]. Chinese Journal of Nursing, 2013, 48(2): 142-146.
- [5] 赵荣江, 牛雅娟, 杨少杰, 等. 中国北方城乡两地居民对10种特殊人群社会距离及社交限制的调查[J]. 中国神经精神疾病杂志, 2015(7): 395-400.  
ZHAO R J, NIU Y J, YANG S J, et al. The attitudes and social distance and restriction to ten different populations among residents in northern China[J]. Chinese Journal of Nervous and Mental Diseases, 2015(7): 395-400.
- [6] 王晓兰. 中国男性同性恋(MSM)人群艾滋病时空模型研究[D]. 烟台: 烟台大学, 2015.  
WANG X L. Spatio-temporal model of HIV/AIDS among chinese male homosexuals(MSM)[D]. Yantai: Yantai University, 2015.
- [7] 黄勤, 李巧巧, 李苑, 等. 2010-2013年中国男男性行为人群艾滋病/梅毒感染状况、性行为特征及艾滋病知识知晓情况的Meta分析[J]. 中华流行病学杂志, 2015, 36(11): 1297-1304.  
HUANG Q, LI Q Q, LI Y, et al. Prevalence of HIV infection and syphilis, sexual behaviors and awareness of HIV/AIDS related knowledge among men who have sex with men in China: a meta-analysis of data collected from 2010 to 2013[J]. Chinese Journal of Epidemiology, 2015, 36(11): 1297-1304.
- [8] LU X. Linked ego networks: improving estimate reliability and validity with respondent-driven sampling[J]. Social Networks, 2013, 35(4): 669-685.
- [9] MAGNANI R, SABIN K, SAIDEL T, et al. Review of sampling hard-to-reach and hidden populations for HIV surveillance[J]. Aids, 2005, 19(S2): 67-72.
- [10] 谭兴兰, 罗柏生, 朱丽霞. 重庆市北碚区艾滋病病毒感染者高危行为调查[J]. 预防医学情

- 报杂志, 2006, 22(5): 519-520.
- TAN X L, LUO B S, ZHU L X. Investigation on high -risk behaviors among people infected with HIV in Beibei District of Chongqing City[J]. Journal of Preventive Medicine Information, 2006, 22(5): 519-520.
- [11] 陈浩, 张艳辉, 谭红专, 等. 某城市不同活动场所男男性行为人群特征和规模估计的研究[J]. 中华流行病学杂志, 2011, 32(7): 664-668.
- CHEN H, ZHANG Y H, TAN H Z, et al. Characteristics and the estimated size of men who have sex with men in different venues of one city[J]. Chinese Journal of Epidemiology, 2011, 32(7): 664-668.
- [12] 薛黎坚, 王文明, 徐勇, 等. 卖性男男性行为者艾滋病高危行为、感染状况以及社会压力的现况研究[J]. 实用预防医学, 2010, 17(12): 2352-2355.
- XUE L J, WANG W M, XU Y, et al. Cross-sectional study on HIV high risk behaviors, HIV infection and social pressure in MSM who had sex-selling behaviors[J]. Practical Preventive Medicine, 2010, 17(12): 2352-2355.
- [13] 张建, 沈洁, 李爱萍, 等. 互联网AIDS咨询内容分析[J]. 中国艾滋病性病, 2005, 11(1): 19-22.
- ZHANG J, SHEN J, LI A P, et al. Analysis of AIDS counselling on Internet[J]. Chinese Journal of AIDS & STD, 2005, 11(1): 19-22.
- [14] 艾德华. 中文同性恋网站及艾滋病预防[C]//第一届中国艾滋病性病防治大会, 2011年11月1日, 北京, 中国. [出版地不详: 出版者不详, 2001: 353-355.
- AI D H. Chinese homosexual website and HIV prevention[C]//The 1st China AIDS and STD Prevention Conference, November 1, 2011, Beijing, China. [S.l.:s.n.], 2001: 353-355.
- [15] 白冠男, 钮文昇. 某网站贴吧“艾滋病吧”一个月帖子的内容分析[J]. 中国健康教育, 2012, 28(8): 649-652.
- BAI G N, NIU W Y. A content analysis of one-month messages posted on an online “AIDS bar”[J]. Chinese Journal of Health Education, 2012, 28(8): 649-652.
- [16] 王国华, 刘菊, 杨腾飞, 等. 网络空间中艾滋病的社会支持研究——以百度贴吧“HIV吧”为例[J]. 情报杂志, 2015(11): 105-110.
- WANG G H, LIU J, YANG T F, et al. Social support for AIDS in cyberspace: take Baidu post bar “HIV bar” for example[J]. Journal of Intelligence, 2015(11): 105-110.
- [17] COURSAIRIS C K, LIU M. An analysis of social support exchanges in online HIV/AIDS self-help groups[J]. Computers in Human Behavior, 2009, 25(4): 911-918.
- [18] MO P K, COULSON N S. Exploring the communication of social support within virtual communities: a content analysis of messages posted to an online HIV/AIDS support group[J]. Cyberpsychology & Behavior: the Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society, 2008, 11(3): 371-374.
- [19] SHI J, CHEN L. Social support on Weibo for people living with HIV/AIDS in China: a quantitative content analysis[J]. Chinese Journal of Communication, 2014, 7(3): 285-298.
- [20] GUO Y, GOH H L. “I Have AIDS”: content analysis of postings in HIV/AIDS support group on a Chinese microblog[J]. Computers in Human Behavior, 2014, 34(34): 219-226.
- [21] WINZELBERG A J, CLASSEN C, ALPERS G W, et al. Evaluation of an internet support group for women with primary breast cancer[J]. Cancer, 2003, 97(5): 1164-1173.
- [22] FOGEL J, ALBERT S M, SCHNABEL F, et al. Internet use and social support in women with breast cancer[J]. Health Psychology Official Journal of the Division of Health Psychology American Psychological Association, 2002, 21(4): 398.
- [23] WINEFIELD H R. Support provision and emotional work in an Internet support group for cancer patients[J]. Patient Education & Counseling, 2006, 62(2): 193.

- [24] COULSON N S. Receiving social support online: an analysis of a computer-mediated support group for individuals living with irritable bowel syndrome[J]. *Cyberpsychology & Behavior*, 2005, 8(6): 580-584.
- [25] FROHLICH D O, ZMYSLINSKI-SEELIG A. The presence of social support messages on youtube videos about inflammatory bowel disease and ostomies[J]. *Health Communication*, 2012, 27(5): 421.
- [26] 余歌. 虚拟社区中艾滋病患者的网络社会支持互动过程研究[D]. 武汉: 华中科技大学, 2015.
- YU G. Research on social support in virtual communities: a case study of hepatitis B bar in Baidu Tieba[D]. Wuhan: Huazhong University of Science and Technology, 2015.
- [27] 潘锐. 虚拟社区中的社会支持研究——以百度贴吧乙肝吧为例[D]. 合肥: 安徽大学, 2015.
- PAN R. Research on social support in virtual communities: a case study of hepatitis B bar in Baidu Tieba[D]. Hefei: Anhui University, 2015.
- [28] 李秀芳, 张北川, 储全胜, 等. 以互联网为最主要性交往途径的男男性行为人群人口学特征研究[J]. *中华流行病学杂志*, 2009, 30(1): 98.
- LI X F, ZHANG B C, CHU Q S, et al. Demographic profile of men who have sex with men who seek for sex mostly on internet[J]. *Chinese Journal of Epidemiology*, 2009, 30(1): 98.
- [29] EVANS A R, WIGGINS R D, MERCER C H, et al. Men who have sex with men in Great Britain: comparison of a self-selected internet sample with a national probability sample[J]. *Sexually Transmitted Infections*, 2007, 83(3): 200-205.
- [30] QIAN Z, VERMUND S, WANG N. Risk of HIV/AIDS in China: subpopulations of special importance[J]. *Sexually Transmitted Infections*, 2005, 81(6): 442.
- [31] CHOI K H, LIU H, GUO Y, et al. Emerging HIV-1 epidemic in China in men who have sex with men[J]. *Lancet*, 2003, 361(9375): 2125-2126.
- [32] WONG F Y, HUANG Z J, WANG W, et al. STIs and HIV among men having sex with men in China: a ticking time bomb[J]. *Aids Education & Prevention Official Publication of the International Society for Aids Education*, 2009, 21(5): 430.
- [33] 宋丹丹. 男性同性兼异性性行为者HIV感染状况及性行为特征[D]. 合肥: 安徽医科大学, 2013.
- SONG D D. Prevalence of HIV and sexual behaviors among men who have sex with men and women[D]. Hefei: Anhui Medical University, 2013.
- [34] 倪志敏, 李剑, 徐丹戈, 等. 浴池MSM人群艾滋病及梅毒感染率调查[J]. *浙江预防医学*, 2011, 23(5): 14-16.
- NI Z M, LI J, XU D G, et al. An investigation on the infection rate of syphilis and HIV among MSM population from public bath[J]. *Zhejiang Journal of Preventive Medicine*, 2011, 23(5): 14-16.
- [35] 周涛, 韩筱璞, 闫小勇, 等. 人类行为时空特性的统计力学[J]. *电子科技大学学报*, 2013, 42(4): 481-540.
- ZHOU T, HAN X P, YAN X Y, et al. Statistical mechanics on temporal and spatial activities of human[J]. *Journal of University of Electronic Science and Technology of China*, 2013, 42(4): 481-540.
- [36] 刘洪君. 微博网络热点话题发现技术研究[D]. 北京: 北京交通大学, 2013.
- LIU H J. Research on hot topic discovery technology of Micro-Blogging network[D]. Beijing: Beijing Jiaotong University, 2013.
- [37] LIU C, LU X. Analyzing hidden populations online: topic, emotion, and social network of HIV-related users in the largest Chinese online community[J]. *Bmc Medical Informatics & Decision Making*, 2018, 18(1): 2.
- [38] EAGLE N, MACY M, CLAXTON R. Network diversity and economic development[J]. *Science*, 2010, 328(5981): 1029.
- [39] LOU J. Entropy and diversity[J]. *Oikos*, 2006, 113(2): 363-375.
- [40] NEWMAN M E, GIRVAN M. Finding and evaluating community structure in

- networks[J]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2004, 69(2): 026113.
- [41] KOBLIN B A, CHESNEY M A, HUSNIK M J, et al. High-risk behaviors among men who have sex with men in 6 US cities: baseline data from the explore study[J]. American Journal of Public Health, 2003, 93(6): 926-932.
- [42] LIU H, YANG H, LI X, et al. Men who have sex with men and human immunodeficiency virus/sexually transmitted disease control in China[J]. Sexually Transmitted Diseases, 2006, 33(2): 68.

#### 作者简介



**刘楚楚** (1993-), 女, 国防科技大学系统工程学院博士生, 主要研究方向为大数据挖掘、复杂网络分析。



**吕欣** (1984-), 男, 国防科技大学系统工程学院副教授, 主要研究方向为大数据挖掘、人类行为动力学分析。

收稿日期: 2018-08-15

通信作者: 吕欣, xin.lu@flowminder.org

基金项目: 国家自然科学基金资助项目 (No.71522014, No.71771213, No.71790615, No.71431006)

**Foundation Items:** The National Natural Science Foundation of China(No.71522014, No.71771213, No.71790615, No.71431006)