

**编者按** 基于互联网的社交网络正在成为人类社会关系维系及信息传播的重要渠道和载体。社交网络基础理论和关键技术研究,对互联网舆情分析引导、突发事件监测响应、在线群体社会交往等国家安全与社会发展的重大问题具有十分重要的基础性和战略性意义。本刊特推出“在线社交网络分析理论和技术”专题,从多侧面评述该领域的进展,希望其能引起读者和相关决策者的关注。该专题由方滨兴院士、许进教授指导推进。



## 社交网络分析核心科学问题、研究现状及未来展望\*

文 / 方滨兴<sup>1</sup> 贾焰<sup>2</sup> 韩毅<sup>2</sup>

1 北京邮电大学可信分布式计算与服务教育部重点实验室 北京 100876

2 国防科学技术大学计算机学院 长沙 410073



中国科学院

**【摘要】** 近年来,随着在线社交服务的快速发展,社交网络进入了我们社会经济生活的各个方面,演变为无处不在的计算平台和信息传播平台。为理解社交网络运行机制的各个方面,文章聚焦分析社交网络运行演化过程中紧密联系的一系列关键性要素:社交网络的结构属性及其演化规律、社交群体及其互动规律、网络信息及其传播方式,探讨了社交网络分析的科学问题,介绍了社交网络分析研究所面临的问题与挑战,并对社交网络的研究方向进行了展望。

**【关键词】** 社交网络,拓扑结构,网络群体,信息传播

DOI 10.16418/j.issn.1000-3045.2015.02.007

### 1 社交网络分析研究的背景

近年来,在线社交网站、微博、博客、论坛、维基等社交网络应用的出现和迅猛发展,使得人类使用互联网的方式产生了深刻变革——由简单信息搜索和网页浏览转向网上社会关系的构建与维护、基于社会关系

的信息创造、交流和共享,社交网络进入了我们社会经济生活的各个方面。

根据欧盟关于社会计算的研究报告 *Key areas in the public sector impact of social computing*,在线社交网络可分为4类:(1)即时消息类应用,即一种提供在线实时通信的

\* 基金项目:国家重点基础研究发展计划(“973”)项目(2013CB329601),国家自然科学基金项目(61372191)  
修改稿收到日期:2015年2月9日

平台,如MSN、QQ、飞信、微信等;(2)在线社交类应用,即一种提供在线社交关系的平台,如Facebook、Google+、人人网等;(3)微博类应用,即一种提供双向发布短信息的平台,如Twitter、新浪微博、腾讯微博等;(4)共享空间等其他类应用,即其他可以相互沟通但结合不紧密的Web2.0应用,如论坛、博客、视频分享、社会书签、在线购物等。

基于互联网的社交网络已经成为了人类社会中社会关系维系和信息传播的重要渠道和载体,对国家安全和社会发展产生着深远的影响:(1)社会个体通过各种连接关系在社交网络上构成“关系结构”,包括以各种复杂关系关联而成的虚拟社区;(2)基于社交网络的关系结构,大量网络个体围绕着某个事件而聚合,并相互影响、作用、依赖,从而形成具有共同行为特征的“网络群体”;(3)基于社交网络关系结构和网络群体,各类“网络信息”得以快速发布并传播扩散形成社会化媒体,进而反馈到现实社会,从而使得社交网络与现实社会间形成互动,并对现实世界产生影响。虚拟的社交网络和真实社会的交融互动对社会的直接影响巨大,所形成的谣言、暴力、欺诈、色情等不良舆论会直接影响国家安全与社会发展。

当前,社交网络应用正处在蓬勃发展期。据Adobe公司调查显示,截至2014年1月,在全球十大社交网络中,成立满10周年的社交网络Facebook领衔排行榜,注册用户约达14亿,是世界第三大“人口国”,其中美国用户最多,约为1.6亿;巴西、印度、印度尼西亚、墨西哥、土耳其和英国排在其后。目前Facebook月度活跃移动用户数为10亿。Youtube位居第2,拥有10亿多用户。中国的QQ空间和新浪微博分别以6.23亿和5.56亿用户排名第3、第4位。紧随其后的是Twitter、Google+、LinkedIn,而俄罗斯社交网络VKontakte排名第8,中国人人网、微信则分别居第9、第10位。

与传统的Web应用及信息媒体应用相比,社交网络信息的发布和接收异常简便、迅速。用户

可以通过手机等各类移动终端随时随地发布和接收信息,人人都有了网络话语权,各类涉及到国计民生的话题和观点可以随时发布,信息一旦发布就能通过“核裂变”的方式传播扩散,期间经过意见领袖的放大作用,促使具有相同观念和诉求的虚拟社区快速形成,并在线下快速组织并发动群众参与社会活动中,有可能构成社会动员力。在线社交网络在给人们生活带来便利的同时,虚拟的社交网络和真实社会的交融互动对社会的直接影响也越来越大,甚至一定程度上影响着国家安全与社会稳定,事关各国的国家战略安全。在政治方面,一些不法分子蓄意制造和传播有损国家和社会利益的谣言,影响社会稳定。

当今中国正处于经济快速发展、社会结构调整、思想文化多样、社会矛盾复杂多变的关键转型期,开展社交网络分析研究将有助于解决国家安全、社会发展等多方面存在的问题。如何有效掌控社交网络这一新型的战略资源,维护国家安全与社会稳定?如何了解民意,推广先进文化、引导网络舆论等?在社会发展层面,如何适应社交网络给人类生活带来的变化,并把握社交网络的发展方向?这些都是国家安全与社会发展所面临的巨大挑战,亟待开展社交网络分析的基础研究。

## 2 社交网络分析研究的核心问题

在线社交网络是一种在信息网络上由社会个体集合及个体之间的连接关系构成的社会性结构,包含关系结构、网络群体与网络信息3个要素。其中,社交网络的关系结构是社会个体成员之间通过社会关系结成的网络系统。个体也称为节点,可以是组织、个人、网络ID等不同含义的实体或虚拟个体;而个体间的相互关系可以是亲友、动作行为、收发消息等多种关系。其中,给定社交网络中的虚拟社区为该社交网络的一个子集,且虚拟社区内节点之间关联密切,不同虚拟社区的节点间关联稀疏。网络社会群体行为是指网络个体就某个事件在某个虚拟空间聚合或集中,相互

影响、作用、依赖,有目的地以类似方式进行的。基于社交网络的信息传播是指社交网络中的个体与个体之间、个体与群体之间、群体与群体之间的信息传递。这种传递作用可以迭代进行。因此社交网络分析主要包括3方面的研究对象:(1)社交网络本身的结构特性;(2)社交网络中的群体及其行为;(3)社交网络中的信息及其传播。

社交网络中的“关系结构”为网络群体互动行为提供了底层平台,是社交网络的载体;“网络群体”直接推动网络信息传播,并反过来影响关系结构,是社交网络的主体;“网络信息”及其传播是社交网络的出发点和归宿,也是群体行为的诱因和效果,同样影响关系结构的变化,是社交网络的客体。上述三者的关系可以归纳为3类问题。

(1)在线社交网络的结构特性与演化机理。社交网络的结构应该用什么方式来表示?什么样的表示方法既能反映社交网络的本质,又能支持计算和分析?什么样的计算方法能够准确刻画社交网络结构的演化?

(2)在线社交网络群体行为形成与互动规律。在社交网络中如何刻画群体的存在及其形成方式?群体间的交互影响如何进行表示与度量?群体间的交互过程对群体的演变所产生的影响应该如何计算?

(3)在线社交网络信息传播规律与演化机理。信息的内涵如何以可计算的形式来表达?信息在社交网络上的传播过程与态势的计算方法是什么?如何用计算的方法来刻画信息内涵与信息传播之间的相互影响?

### 3 国际学术界高度关注社交网络研究

在线社交网络分析涉及计算机科学、社会学、管理学、心理学等多个学科领域,世界各国从20世纪初就有由社会学家开始分析

研究社交网络,随着信息时代的来临,以计算技术为代表的技术手段被全面应用于社交网络分析研究。本节将针对社会网站分析的3类研究对象介绍国际学术界近年来的关注热点。

#### 3.1 社交网络的结构特性研究

在社交网络的结构特性研究方面,已有的研究工作可以概括为社交网络的结构分析与建模、虚拟社区发现、社交网络演化分析3个主要方面。

(1)社交网络的结构分析与建模。社交网络结构分析与建模是所有分析的基础。社交网络结构分析是通过统计分析方法来分析网络中节点度的分布规律、关系紧密程度、相识关系的紧密程度,某一个用户对于网络中所有其他用户对之间传递消息的重要程度等诸多统计特性。社交网络建模是针对社交网络的特性,采用结构建模的方法来研究产生这些特性的机制,以此来深刻认识社交网络的内在规律和本质特征。

在社交网络结构建模方面,图论方法得到了广泛的应用,很多学者都尝试运用图论对社交网络进行定量分析。20世纪30年代,哈佛学者莫雷诺(Moreno)首先将图论的方法引入了人类社交关系分析中<sup>[1]</sup>;1960年,密歇根大学的哈拉里(Harary)等用有向图模型刻画社交网络中的单向关系,并提出了中心性的概念<sup>[1]</sup>;为了描述个体-组织等归属关系,2008年,伊利诺伊州立大学的菲利普(Philip)等人<sup>[2]</sup>又将其扩展到二部图和多部图模型,利用二部图建模了学术网络中个体和学术会议,并给出了个体间接接近性的计算方法;针对多人合作关系,2009年,密歇根州立大学的纽曼(Newman)提出了超图建模方法<sup>[3]</sup>,认为社交网络中的一条边可以连接2个以上的节点;针对个体横跨多社交关系的情况,2005年,美国纽约州立大学布法罗





分校的裴健等人<sup>[4]</sup>在无向图的基础上,提出了基于节点唯一性标识的多图交叉(Cross-Graph)模型,在多图并集上使用可控的启发式算法发现符合特定要求的网络结构。

近年来,学者们针对社交网络的结构特性开展了大量的研究,我们对相关研究进行了归纳。在社交网络的结构特性分析方面,大量已有研究验证了在真实世界中各种不同的社交网络具有许多复杂网络所共有的结构特性,例如,“六度分隔”、小世界现象、无标度、幂律分布和结构鲁棒性等。2003年,美国哥伦比亚大学的沃茨(Watts)等人<sup>[5]</sup>在6万节点规模的邮件网络上验证了“六度分隔”和小世界模型。2005年,美国明尼苏达大学的库玛(Kumar)等人<sup>[6]</sup>研究了雅虎在线社区的路径长度,发现网络规模最大时的平均路径长度和有效路径长度分别为8和10,比通常认为的“六度分隔”要大。2007年,惠普实验室的格德尔(Golder)等人<sup>[7]</sup>研究了在线社会网络Facebook,发现好友数(度值)的中值为144,均值为179.53。2010年,IBM的查鲁(Charu)等人<sup>[8]</sup>将有向图中节点的入度考虑为节点受欢迎程度的一个因素,并发现在社交网络中,受欢迎度(Popularity)同样也服从幂率分布。

(2)虚拟社区发现。虚拟社区发现是社交网络分析的必备功能。在社会学领域,社区是一群人在网络上从事公众讨论,经过一段时间,彼此拥有足够的情感之后,所形成的人际关系的网络。社交网络中存在关系不均匀的现象,有些个体之间关系密切,有些关系生疏,从而在常规的社区之上围绕某一个焦点又形成了联系更为密切的社区形式,这可以看作社交网络中的虚拟社区结构。虚拟社区结构是在线社交网络的一种典型的拓扑结构特征。在新浪微博、Facebook等在线社交网络中,通过挖掘社区可以发现用户联系的紧密情况,获得用户之间的社交关系以及社会角色,并进一步结合社区内用户观点/行为等分析,有助于理解网络拓扑结构特点、揭示复杂系统内在功能特

性、理解社区内个体关系/行为及演化趋势。

在社交网络的虚拟社区发现分析方面,社区结构是研究热点之一。根据社区结构的定义,社区结构可分为不可重叠的社区结构发现和可重叠的社区结构发现。在不重叠的社区结构发现方面,目前被最为广泛关注的是2004年由美国密歇根大学的纽曼(Newman)等人<sup>[9]</sup>提出的通过寻找使得社区的模块度最大的网络划分来发现网络社区的算法。2007年,美国印第安纳大学的福图纳托(Fortunato)等人<sup>[10]</sup>指出模块度优化方法存在分辨率限制问题,使得基于模块度优化的方法无法识别出一些较小的社区。在可重叠的社区结构发现方面,2005年,匈牙利科学院的帕拉(Palla)等人<sup>[11]</sup>提出了一种基于K-完全子图的重叠社区发现方法,该方法的优点在于能够揭示网络社区间的重叠现象,不足之处在于其参数选择缺乏有效的理论指导。2009年,美国印第安纳大学的兰斯齐那提(Lancichinetti)等人<sup>[12]</sup>研究了网络层次化社区的发现问题。

(3)社交网络演化分析。虚拟社区具有动态演化性,需对演化机理进行分析与识别。虚拟社区结构反映了网络中个体行为的局部聚集特征,这些虚拟社区结构并不是永恒不变的,由于在线社交网络结构随着时间不断演化,虚拟社区结构也随之不断演化。在线社交网络中存在着大量各类显性或隐性的虚拟社区结构,譬如人人网中的圈子、豆瓣网上的小组等,其都在不停地动态演化。虚拟社区的演化与社交网络诸如扩散、抗毁、合作、同步等方面的功能密切关联,对社交网络自身的演化也起着基础性的作用。

在社交网络演化分析方面,学者们从社交网络演化中的统计规律展开研究,并提出了面向不同类型社交网络的演化模型。在社交网络演化规律方面,2005年,美国斯坦福大学的拉斯科维奇(Leskovec)等人<sup>[13]</sup>利用基于图论的方法,描述节点数与网络直径的动态关联和随时间演化的特性,以及边权重与网络的拓扑结构演化关联关系及社

区演化关系。2006年,美国加州伯克利大学的查布拉巴蒂(Chakrabarti)等人<sup>[14]</sup>利用数据挖掘技术给出了时间和结构相结合的虚拟社区演化模型。2007年,美国NEC实验室的迟(Chi)等人<sup>[15]</sup>扩展了相似性计算方法,用图分割(Graph cut)作为测度社区结构和社区进化的指标,第一个提出进化谱聚类算法。同年,美国伊利诺伊大学芝加哥分校的坦提帕斯安那(Tantipathananandh)等人<sup>[16]</sup>根据社会经验列举了社交网络上中观层面子结构变化的各种情况,并构造出相应变化的损耗评价体系,建立了以个体消耗、组消耗、颜色变化消耗三种因素合成的最优化模型。2008年,美国斯坦福大学的拉斯科维奇(Leskevec)等人<sup>[17]</sup>研究了社交网络的微观演化过程,发现边的生成频度与节点间的已有跳数呈反比。

### 3.2 社交网络中群体互动研究

社交网络中的群体互动方面,已有的研究工作可以概括为社交网络中群体行为建模及特征分析、群体情感建模与行为互动2个方面。

(1) 社交网络上的用户行为分析。用户个体行为是社区中的基本动作,需对其进行建模。在线社交网络上的用户行为包括展示自我、与陌生人建立关系、分享兴趣和信息、发布信息、搜索信息、浏览信息和推送信息;可以围绕各种话题与不同人群进行互动;可以构建兴趣社区、学习社区和娱乐社区,共享知识、学习交流并分享快乐。

社交网络上群体行为分析的已有研究主要集中在群体社交网络选择模型研究,以及个体行为特征分析等2个方面。在群体社交网络选择模型研究方面,2007年,美国密歇根大学的埃里森(Ellison)等人<sup>[18]</sup>将社交网络的群体行为关系分为桥接型、黏接型和维持型三种类型,并基于回归分析发现桥

接型关系对个体选择社交网络有着更重要的影响。2008年,美国密歇根大学的施泰因费尔德(Steinfeld)等人<sup>[19]</sup>以Facebook为背景,采用博弈论的方法,通过制定个体选择策略和收益量化函数对个体行为进行建模和分析,研究表明社交网络的社会资本收益、用户自尊心以及生活满意度等心理变量共同影响用户社交网络选择行为。2009年,韩国Myongji大学康永植(Kang)等人<sup>[20]</sup>研究了自我形象一致性和后悔成本等因素对网络迁移行为的影响。2009年,新加坡南洋理工大学程曾燕等人<sup>[21]</sup>的研究也表明用户对所属网络的不满意、其他网络的吸引力以及网络迁移的成本是用户网络迁移行为的主要原因。2011年,美国亚利桑那州立大学库玛(Kumar)等人<sup>[22]</sup>针对Twitter、Facebook、Youtube等7种社交网络,研究了它们的异构性及用户在其间的迁移行为,建立了个体迁移和群体迁移模式,表征用户在社交网络之间的迁移方式。在个体行为特征分析方面,2003年,希腊雅典经济与商业大学的Eirinaki等人<sup>[23]</sup>横向比较了Web服务中提供的个性化行为特征挖掘功能。2008年,加拿大西蒙菲莎大学的裴健等人<sup>[24]</sup>提出利用个体属性中多维性抽取和个体行为特征密切关联的属性取值,从而达到自动学习个体喜好和行为规律的目标。2009年,香港科技大学的黄智荣等人<sup>[25]</sup>提出了一种基于关联分析的属性独特性度量方法,用于挖掘群体中的个体行为特征。1960年,德国施拉姆(Schramm)<sup>[26]</sup>等人从传播学的角度,对个体的行为与动机进行了分析,建立了个体行为特征模式。

(2) 群体情感建模与行为互动。情感分析是针对主观性信息(“支持”、“反对”、“中立”)进行分析、处理和归纳的过程,主观性信息表达了人们的各种情感色彩和情感倾



中国科学院

向。社交网络中每个人情感状态不同,影响力也会不同。

在社交网络群体情感建模与行为互动方面,2007年,美国密歇根大学埃里森(Ellison)等人<sup>[18]</sup>发现在线交互在统计上不但不会隔离离线用户,反而能够支持用户之间的联系,为从众行为的产生提供了环境。2009年,美国斯坦福大学的凯莫勒(Camerer)等人<sup>[27]</sup>采用博弈论方法,对网络中的级联行为,个体效应与群体效应的相互影响,进行了建模和分析。2010年,英国牛津大学的格里茨(Gryc)等人<sup>[28]</sup>提出一种在博客社区上挖掘文本倾向性的方法,该方法以16 741位博主的约280万篇博文作为数据集,分析博文对奥巴马的倾向性,并将博客空间的社区划分作为用户倾向性分类的特征之一。1960年,德国施拉姆(Schramm)等人<sup>[26]</sup>利用传播学的方法,建立了群体与互动的相对完整的理论体系。2010年,美国亚利桑那州立大学的Zafarani等人<sup>[29]</sup>在LiveJournal社会网络数据集上研究倾向性的传播,并对传播的倾向性进行了度量。2011年,香港城市大学Xu等人<sup>[30]</sup>先对文本的倾向性进行分析,再结合社区发现方法发现有相同倾向性的群体。同年,美国印第安纳大学博伦(Bollen)等人<sup>[31]</sup>对Twitter进行了分析,提出用户会把自己的倾向性传播给具有连接关系的其他用户,使他们逐渐持有相同或相似的主观感受。此外,2011年日本Tokai大学的林(Aoki)等人<sup>[32]</sup>针对博客采用向量的表示方法对多元化情绪进行建模,但其基于表情符号来构建向量,并未利用内容信息进行分析研究。社交网络个体从众行为分析目前已有研究主要集中在社交网络个体从众行为的产生环境、影响因素,以及从众行为机理分析等几方面。2008年,台湾中原大学陈宜菜等人<sup>[33]</sup>通过对在线书籍评分、销售量、不同来源3个角度研究了在线购物的从众效应,结果表明对书籍的评分、销售量以及其他用户的选择能够影响用户的决策。2010年,澳大利亚悉尼大学Sharawneh等人<sup>[34]</sup>提出根据“追随领袖”模型中的领袖信用,建

立集成社交网络信息的协同过滤推荐模型,该研究表明,基于意见领袖的推荐模型能够更加准确预测消费者的选择行为。

### 3.3 社交网络中的信息传播研究

在社交网络的信息传播方面,已有的研究工作可以概括为社交网络的信息及其能量、社交网络信息传播模型和社交网络信息传播影响3个方面。

(1)社交网络的信息及其能量。信息传播是人们通过符号、信号来进行信息的传递、接收与反馈的活动,是人们彼此交换意见、思想、情感,以达到相互了解和影响的过程。在社交网络的信息及其能量方面,已有研究主要集中在信息的符号表示与意义,以及信息传播能量及演化方面。

在信息符号表示方面,1923年,德国哲学家卡西尔(Cassirer)等人<sup>[35]</sup>从人类传播的符号与意义角度,分析了传播符号与意义的关系。贝尔实验室的香农(Shannon)<sup>[36]</sup>建立了概念明确、数学表述和运算严格、自成体系且能付诸实用的信息理论,华盛顿大学的杰恩斯(Jaynes)等人<sup>[37]</sup>将其发展为统计信息理论:信息和信息熵是统计信息理论中最基本的概念和量,不仅用作信息量的度量,还用来表示自然和社会各种系统的有序度和无序度,并广泛用于通信、计算机、控制论、社会学及各种工程科学等领域。在信息传播过程中的能量及演化分析方面,2004年,法国巴黎大学丹卿(Danchin)等人<sup>[38]</sup>利用社会物理学的方法,用信息传播状态熵刻画信息在社交网络的扩散状态及势能,指出信息传播过程中,随着多种异源异构信息互为补充不断演化,增强了信息最后的能量。2011年,美国伊利诺伊大学香槟分校的韩家炜等人<sup>[39]</sup>提出了社交网络话题演化与传播路径的综合分析方法,该方法基于联合概率推理,利用高斯条件随机场模型进行文本内容、影响力与话题演化的统一分析。2011年,美国康奈尔大学Jo等人<sup>[40]</sup>提出从网络语料库中提取主题拓扑结构,对信息进化图进行建模分析。2011年,印度国际信息技术研究所



的库玛(Kumar)等人<sup>[41]</sup>基于多频封闭词集来表示信息主题提出了新的信息演化分析方法,并采用改进的矩阵迭代算法对不同的子话题进行演化计算。

(2)社交网络信息传播模型。社交网络信息传播特指以社交网络为媒介进行的信息传播过程。在线社交网络与生俱来的自由性和开放性,使其逐渐成为当代社会信息传播的重要集散地,社交网络中的信息传播活跃性达到了前所未有的程度。

在社交网络信息传播模型方面,已有研究主要集中在传染病模型、网络拓扑图模型以及基于统计推理的信息传播模型等。传染病模型最早于2000年由美国爱荷华大学赫斯科特(Hethcote)教授等人<sup>[42]</sup>提出,之后出现了很多变种,例如,SIR、SIS以及类似于SI模型的级联模型<sup>[43]</sup>。传染病模型认为,当感染个体对某个未感染个体的传播率大于某一临界值时,感染个体会将病毒传播给该未感染个体,这个过程会持续到整个网络感染个体总数处于某一稳定状态。2004年,IBM Almaden研究中心格鲁尔(Gruhl)等人<sup>[44]</sup>基于SIRS传染病模型,提出话题在博客网络中的传播模型,并提出博客阅读率与传播率的估算方法。在基于网络拓扑图的社交网络信息传播模型方面,2006年,雅虎研究中心库玛(Kumar)等人<sup>[45]</sup>研究了在线社交网络的网络拓扑结构的演化特性,包括度分布聚类系数、群聚系数、顶点度相关系数等属性。在基于统计推理的社交网络信息传播模型方面,2000年,英国剑桥大学渡尚(Doucet)等人<sup>[46]</sup>提出了BRPT算法,利用动态概率网络的结构分解来减小抽样分布的维数,利用部分样本对局部分布进行近似以提高计算效率;2001年,美国加利福尼亚大学卡斯克(Kask)等人<sup>[47]</sup>提出基于桶消元的信息传播算法;同年,美国加州大学伯克

利分校墨菲(Murphy)等人<sup>[48]</sup>提出基于联合树的静态信息传播算法,可有效计算中等规模网络的信息传播;2003年,美国加利福尼亚大学帕斯金(Paskin)等人<sup>[49]</sup>提出一种动态信息传播的条件BK(CBK)算法,通过引入局部结构的条件独立性提高动态信息传播的计算精度。在传播学领域,1957年美国麻省理工拉斯韦尔(Lasswell)等人<sup>[50]</sup>基于传播学方法,提出5W模式传播模型;1960年德国施拉姆(Schramm)等人<sup>[26]</sup>基于控制论方法,进一步建立了大众传播模式,引入了反馈等影响,更加注重信息编解码和反馈等因素在信息传播中的作用。

(3)社交网络信息传播影响。社交网络信息传播影响是社交网络信息传播领域的关键研究热点,其目的是发现社交网络中最有信息传播影响力的节点集合,从而经过信息在社会网络中的传播,最终能够最大化信息的传播范围。在实际生活的许多重要场景中,社交网络信息传播影响分析均有着广泛的应用,例如,市场营销、广告发布、舆情预警、水质监测、疫情监控、网络竞选、突发事件通知等。

国内外对社交网络信息传播学影响的研究主要有概率论方法、经济学方法和传播学方法。概率论方法是基于概率图模型对研究信息传播的敏感度和影响力的方法,可以对概率图参数、结构的重要性进行量化分析。2006年,加拿大西蒙菲莎大学的马肯德诗(McCandless)等人<sup>[51]</sup>利用概率论,提出了基于蒙特卡罗的信息传播计算方法,但是这种类型的影响力分析方法不能有效揭示非线性关系。同时,研究者探讨了概率结构模型的影响力分析方法,如美国匹斯堡大学的王海琴等人<sup>[52]</sup>研究发现,贝叶斯网络对于参数概率的变化非常敏感,并论证了敏感度分析对贝叶斯网络参数分析非常有效,但这



中国科学院

些贝叶斯网络敏感度分析方法还仅涉及到单个参数;荷兰乌特勒支大学的Renooij等人<sup>[53]</sup>将贝叶斯网络灵敏性分析延伸到多个参数,利用动态概率结构模型进行影响力的敏感度分析等。经济学中的效用理论是度量信息传播影响力的另一种有效理论。基于效用函数的信息传播影响力分析模型是利用函数评估信息传播对于系统所产生的影响,且易于实现信息传播影响力的度量。1992年,纽约大学的Shenoy等人<sup>[54]</sup>利用经济学方法,提出了基于效用评价的信息传播影响力求解方法,利用一种可替换公式,基于效用评价系统,来对信息传播影响力问题的表示和求解。1999年,丹麦奥尔堡大学的麦德森(Madsen)等人<sup>[55]</sup>提出了一种基于熵的网络信息传播的影响力评估函数方法,通过在强连接树中进行传播信息的影响力推理估算。2006年,法国巴黎大学Sam Maes等人<sup>[56]</sup>基于多Agent因果概率图模型给出了一种利用部分信息的影响力分析方法。在传播学领域,20世纪70年代以后,出现了传播效果研究的高潮,产生了一系列理论。1974年,德国纽曼(Neumann)等人<sup>[57]</sup>基于传播学方法,提出了“沉默螺旋”等理论,指出个人在表达自己观点时受环境影响,当发现自己属于优势意见时,倾向于表明自己观点;反之,则转向沉默或附和。于是少数派的声音越来越小,多数派的声音越来越大,形成一种螺旋式上升的模式。1972年,美国传播学家麦克博斯(McCombs)等人<sup>[58]</sup>提出的议程设置功能理论是指大众传媒作为“大事”加以报道的问题,同样也作为“大事”反映在公共意识中,传统媒体给予的强调越多,公众对该问题的重视程度就越高。1927年,奥地利媒介理论家弗洛伊德形成一种新的宣传理论,即魔弹理论<sup>[59]</sup>,把媒介对人的刺激看作是魔弹打入大脑,能迅速地被受众所接受,即:人们在接受外部信息的时候多是被动的状态,接收到信息便认为是什么信息,这个信息将造成某种事先设计好的行为。1970年,美国传播学家Tichenor等人<sup>[60]</sup>提出“知识沟理论”,是指由于社会经济地位

高者通常能比社会经济地位低者更快地获得信息,大众媒介传送的信息越多,这二者之间的知识鸿沟也就越有扩大的趋势。

## 4 我国在社交网络分析方面的研究进展

在我国,社交网络分析受到了学术界和工业界的高度关注,研究与国际上同步开展,国家自然科学基金委、科技部等都部署了系列相关项目。本文的研究就是在国家重点基础研究发展计划(“973”)项目“社交网络分析与网络信息传播的基础研究”项目支持下开展的。

在科学研究方面,国防科技大学、上海交通大学、合肥工业大学、北京邮电大学、中科院计算所、北京大学、清华大学、北京科技大学、浙江大学、哈尔滨工业大学、中科院信工所等单位都开展了系统性的研究。在结构分析方面,2009年,中科院计算所的程学旗等人<sup>[61]</sup>分析了社区结构的特性,改进了网络层次化重叠社区的发现方法,能够同时揭示网络的层次化和重叠社区结构。群体特性方面,2009年,合肥工业大学杨善林等人<sup>[62]</sup>利用元胞自动机仿真从众行为,发现当群体出现完全从众行为时,从众行为的结果对初始状态相当敏感,不同的初始状态就可能有不同的演化结果。信息传播方面,2004年,北京理工大学邢修三<sup>[63]</sup>提出了以表述信息演化规律的信息熵演化方程为核心的非平衡统计信息理论,从定量的角度对信息演化机理进行了有益的探索。方滨兴等人<sup>[64]</sup>也在系统梳理国内外相关理论和技术的基礎上,编著了《在线社交网络分析》一书,从上述3个角度入手为相关研究者提供理论性、系统性、工具性的研究指导。

社交网络一直是各类互联网技术、数据挖掘技术方向国际学术会议的热门主题。在国际影响力方面,我国多次与国际知名学术机构联合,先后承办包括WWW 2008(世界万维网大会)、SIGKDD 2012(ACM数据挖掘及知识发现大会)、ASONAM 2014(IEEE/ACM社交网络分析大会)等在内的多个国际权威学术会议,提升了我国学



术机构在国际上的影响力。

## 5 社交网络研究尚有诸多问题亟待突破

国际上人们对于大型社交网络的本质特征和网络信息传播的基本规律的研究仍处在相对初级的阶段,尚未提出完整的社交网络分析的基础理论和方法,仍然值得我们进一步进行研究和突破。

(1)在线社交网络的结构具有节点海量性、结构复杂性和多维演化性等特点,拓扑结构随着时间不断演变,对社交网络结构演化规律还需要进一步有效的表达和计算加以解决。

(2)在线社交网络的信息传播具有信息的多源并发性,其相互影响形成了路径多变和内容演化的特点。网络群体方面,已有研究对其产生、发展、消亡规律的内部交互作用机理知之不深;传统的群体建模及其互动方法无法准确刻画大规模在线社交网络中的强互动演变、公众情绪漂移等特征,不能真实分析出舆情的倾向性。在个体行为特征分析方面,传统的研究主要局限于用户个性化模型及其对社交网络的选择分析上,未涉及个体向群体演化过程中的个体行为表征等问题。

(3)在线社交网络的群体互动具有强互动演变、公众情绪漂移等特征点,公众立场不断变化,兴趣点不断演化。已有的信息传播模型多基于传染病模型、网络拓扑图以及统计推理等方法,在描述社交网络传播模式以及计算效率和精度方面仍存在不足。研究视角上,尚缺乏从信息传播的时间、空间特征以及信息传播的双向性3个维度,对信息传播的内在机制进行深度分析。传统理论和方法局限在“还原论”的角度解决问题,不能准确描述在线社交网络中信息的多源

并发性所带来的相互影响等特性,因此需要研究新理论与新方法,以便在信息传播的相互作用中形成对舆情的驾驭能力。

## 6 对社交网络研究问题的思考

社交网络分析的3个要素“结构”、“群体”和“传播”是相互作用的关系。社交网络结构建模和群体互动规律分析的目标都是支持信息传播形态研究,因此是否能够将结构建模与群体互动者两个独立的要素结合起来,研究结构与互动相结合的动态建模方式?已有的信息传播过程分析中,核爆炸式、烟花式、星球式、水母式等信息爆发式传播形态都能客观表示,如何能够让拓扑结构建模对这些信息传播的模型有所贡献?如何从结构建模的角度支持这些外显形态的存在?群体事件、群体极化、群体演化等社交网络中的群体形态是否也能被模型化?信息传播的影响力是否能够形成一个可计算的量纲?如何将碎片化、时变的、演化的话题分析技术和信息传播、溯源技术结合起来,形成信息传播的可计算模型。

作为一种新型媒介,在线社交网络的根本在于信息流动和传播,而信息的传播根植于社会群体的互动,但又受限于社交结构与社区的规模。信息传播过程可以被视作社交网络中一个时间函数的求解变量,其中的群体正是函数因子,而社交网络的拓扑结构是控制其发展的边界条件。

21世纪,是人类高度依赖数据、深度融入信息社会的世纪。在信息社会中,在线社交网络构建了庞大的功能平台。人们在社交网络上表达观点、交友互动,每天都产生数亿计的信息,在线社交正在改变着人们的行为模式和社会形态,而在线社交网络数据也正在成为最成熟的大数据,通过研究和分析这一技术,人们有望对在线社交网络大数



中国科学院

据背后的用户行为、社会现象的理解达到空前的深度。

### 参考文献

- 1 History of Social Network Analysis. [2012-6-7]. <http://www.analytictech.com/networks/history.htm>.
- 2 Tong H, Papadimitriou S, Philip Y et al. Fast monitoring proximity and centrality on time-evolving bipartite graphs. *Statistic Analysis on Data Mining*, 2008, 1: 142-156.
- 3 Ghoshal G, Zlatic V, Caldarelli G et al. Random hypergraphs and their applications. *Phys Rev E*, 2009, 79: 066118.
- 4 Pei J, Jiang D, Zhang A. Mining cross-graph quasi-cliques in gene expression and protein interaction data. *Proceedings of the 21st International Conference on Data Engineering*. National Center of Sciences, 2005, 353-356.
- 5 Dodds P, Watts D, Sabel C. Information exchange and robustness in organizational networks. *Proceedings of the National Academy of Sciences*, 2003, 100: 12516-12521.
- 6 Liben D, Novak J, Kumar R et al. Geographic routing in social networks. *Proceedings of the National Academy of Sciences*, 2005, 102: 11623-11628.
- 7 Golder S, Wilkinson D, Huberman B. Rhythms of social interaction: messaging within a massive online network. *Proceedings of the 3rd Communication Technology Conference (CT2007)*. East Lansing: Springer, 2007, 41-66.
- 8 Charu A, Wang H. *Managing and mining graph data*. New York: Springer-Verlag New York Inc, 2010.
- 9 Newman M, Moore C. Finding Community Structure in Very Large Networks, Aaron Clauset. *Physical Review Letters*, 2004, 70:066-111.
- 10 Fortunato S, Barthélemy M. Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 2007, 104: 36-41.
- 11 Palla G, Dernyi I, Farkas I. Uncovering the overlapping community structure of complex network in nature and society. *Nature*, 2005, 435: 814-818.
- 12 Lancichinetti A, Fortunato S, Kertesz J. Detecting the overlapping and hierarchical community structure in complex networks. *New J Phys*, 2009, 11: 033015.
- 13 Leskovec J, Kleinberg J, Faloutsos C. Graphs over time: densification laws, shrinking diameters and possible explanations. *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, 2005 Aug 21-24, Chicago, Illinois, USA. 2005, 177-187.
- 14 Chakrabarti D, Kumar R, Tomkins A. Evolutionary clustering. *Proceedings of the 12th International Conference on Knowledge Discovery and Data Mining*, 2006 Aug 20-23, Philadelphia, USA. 2006, 554-560.
- 15 Chi Y, Song X, Zhou D et al. Evolutionary spectral clustering by incorporating temporal smoothness. *Proceedings of the 13th International Conference on Knowledge Discovery and Data Mining*, 2007 Aug 24-27, San Jose, California, USA. 2007, 153-162.
- 16 Tantipathananandh C, Berger-Wolf T, Kempe D. A framework for community identification in dynamic social networks. *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, San Jose, California, USA. 2007, 717-726.
- 17 Leskovec J, Backstrom L, Kumar R et al. Microscopic evolution of social networks. *Proceedings of the 14th International Conference on Knowledge Discovery and Data Mining*, 2008 Aug 24-27, Las Vegas, Nevada, USA. 2008, 462-470.
- 18 Ellison, N, Steinfield C, Lampe C. The benefits of facebook "friends": social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, 2007, 12: 1143-1168.
- 19 Steinfield C, Ellison N, Lampe C. Social capital, self-esteem, and use of online social network sites: a longitudinal analysis. *Journal of Applied Developmental Psychology*, 2008, 29: 434-445.
- 20 Kang YS, Hong S, Lee H. Exploring continued online service usage behavior: the roles of self-image congruity and regret. *Computers in Human Behavior*, 2009, 25: 111-122.
- 21 Chen Z Y, Yang Y, John L. Cyber migration: an empirical investigation on factors that affect users' switch intentions in social networking sites. *Proceedings of the 42nd Hawaii International Conference on System Sciences*, 2009 Jan 5-8, Big Island, HI, USA.

- 2009,1-11.
- 22 Kumar S, Zafarani R, Liu H. Understanding user migration patterns in social media. Proceedings of the 25th AAAI Conference on Artificial Intelligence, 2011 Aug 7-11, San Francisco, California, USA. 2011. 1-6.
- 23 Eirinaki M, Vazirgiannis M. Web mining for web personalization. ACM Transactions on Internet Technology, 2003, 3: 1-27.
- 24 Jiang B, Pei J, Lin X et al. Mining preferences from superior and inferior examples. Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008 Aug 24-27, Las Vegas, Nevada, USA. 2008, 390-398.
- 25 Wong RC, Pei J, Fu A et al. Online skyline analysis with dynamic preferences on nominal attributes. IEEE Transactions on Knowledge and Data Engineering, 2009, 21: 35-49.
- 26 Schramm W. Mass communications (2nd ed.). Urbana, IL: University of Illinois Press, 1960.
- 27 Camerer C. Behavioral game theory: experiments in strategic interaction. Princeton: Princeton University Press, 2009.
- 28 Gryc W, Moilanen K. Leveraging textual sentiment analysis with social network modeling: sentiment analysis of political blogs in the 2008 U.S. presidential election. From Text to Political Positions. Text analysis across disciplines, 2010, 11: 47-70.
- 29 Zafarani R, Cole W, Liu H. Sentiment Propagation in Social Networks: A Case Study in LiveJournal. Advances in Social Computing, 2010.
- 30 Xu K, Li J, Liao S. Sentiment community detection in social networks. Proceedings of the 2011 Conference, 2011 Mar 4-7, Berlin, Germany. 2011, 804-805.
- 31 Bollen J, Goncalves B, Ruan G et al. Happiness is assortative in online social networks. Artificial Life, 2011, 132-136.
- 32 Aoki S, Uchida O. A method for automatically generating the emotional vectors of emoticons using blog articles. International Journal of Computers, 2011, 5: 346-353.
- 33 Chen Y F. Herd behavior in purchasing books online. Computers in Human Behavior, 2008, 24: 1977-1992.
- 34 Al-Sharawneh J, Williams M. Credibility-based social network recommendation: follow the leader. Proceedings of Australia Conference on Information System, 2010 Dec 1-3, Brisbane, Australia. 2010, 24.
- 35 Cassirer E. The Philosophy of Symbolic Forms. Volume One: Language. New Haven: Yale University Press, 1923.
- 36 Shannon C E. A mathematical theory of communication. Bell System Technical Journal, 1948, 27: 379-423.
- 37 Jaynes E T. Information Theory and Statistical Mechanics. Physical Review, 1957, (4): 620-630.
- 38 Danchin E, Giraldeau L, Valone T J et al. Public information: from nosy neighbors to cultural evolution. Science, 2004, 5683: 487-491.
- 39 Lin C X, Mei Q, Jiang Y et al. Inferring the diffusion and evolution of topics in social communities. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, 2010 Jul 25-28, Washington, DC, USA. 2010, 1019-1028.
- 40 Jo Y, Hopcroft J E, Lagoze C. The web of topics: discovering the topology of topic evolution in a corpus. Proceedings of the 20th International Conference on World Wide Web, 2011 Mar 28-Apr 1, Hyderabad, India. 2011, 257-266.
- 41 Kumar S, Zafarani R, Liu H. Understanding user migration patterns in social media. Proceedings of the 25th AAAI Conference on Artificial Intelligence, 2011 Aug 7-11, San Francisco, California, USA. 2011, 1-6.
- 42 Hethcote H. The mathematics of infectious diseases. SIAM Rev, 2000, 42: 599-653.
- 43 Kempe D, Kleinberg J, Tardos E. Maximizing the spread of influence through a social network. Proceedings of





- the 9th International Conference on Knowledge Discovery and Data Mining, 2003 Aug 24-27, Washington, DC, USA. 2003, 137-146.
- 44 Gruhl D, Guha R, Liben-Nowell D et al. Information diffusion through blogspace. ACM SIGKDD Explorations Newsletter, 2004, 6: 43-52.
- 45 Kumar R, Novak J, Tomkins A. Structure and evolution of online social networks. Proceedings of the 12th International Conference on Knowledge Discovery and Data Mining, 2006 Aug 20-23, Philadelphia, USA. 2006, 611-617.
- 46 Doucet A, Freitas N, Murphy K. Rao-blackwellised Particle filtering for dynamic bayesian networks. Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence, 2000 Jun 30-Jul 3, Stanford, California, USA. 2000, 176-183.
- 47 Kask K, Dechter R, Larrosal J et al. Buchet-tree elimination for automated reasoning. Artificial Intelligence. 2001, 125: 91-131.
- 48 Murphy K, Weiss Y. The factored frontier algorithm for approximate inference in DBNs. Technical Report, Computer Science Department, University of California, Berkeley. 2001, 176-183.
- 49 Paskin M. Thin junction tree filters for simultaneous localization and mapping. Technical Report, Electronics and Computer Science, University of California, Berkeley. 2003.
- 50 Lasswell H D. The structure and function of communication in society. The communication of Ideas, 1948, 117-118.
- 51 McCandless LC, Gustafson P, Levy A. Bayesian sensitivity analysis for unmeasured confounding in observational studies. Statistics in Medicine, 2006, 26:2331-2347.
- 52 Wang HQ. Building Bayesian networks: elicitation, evaluation, and learning. Doctor Dissertation. Pittsburgh: University of Pittsburgh, 2004.
- 53 Renooij S. Efficient sensitivity analysis in hidden markov models. Int J Approx Reason, 2012, 53: 1397-1414.
- 54 Shenoy p p. Valuationbased systems for Bayesian decision analysis. Operations Research, 1992,40: 63-84.
- 55 Madsen A L. Lazy propagation:A junction tree inference algorithm based on lazy evaluation. Artificial Intelligence, 1999, 113: 203-245.
- 56 Maes S, Philippe L. Multi-agent causal models for dependability analysis. Proceedings of the 1st International Conference on Availability, Reliability and Security, 2006 Apr 20-22, Vienna, Austria. 2006, 794-798.
- 57 Noelle-Neumann E. The Spiral of Silence A Theory of Public Opinion. Journal of Communication, 1974, 24: 43-51.
- 58 McCombs M E, Shaw D. The Agenda-Setting Function of Mass Media. POQ, 1972, 36: 176-187.
- 59 Sigmund Freud. The Ego and the Id (German: Das Ich und das Es). 1927.
- 60 Tichenor J, George A, Clarice O. Mass media flow and differential growth in knowledge. Public Opin Q, 1970, 34: 159-170.
- 61 Shen H W, Cheng X Q, Cai K et al. Detect overlapping and hierarchical community structure in networks. Physical A-Statistical Mechanics and Its Applications, 2009, 388: 1706-1712.
- 62 杨善林, 朱克毓, 付超等. 基于元胞自动机的群决策从众行为仿真, 系统工程理论与实践, 2009, 29(9): 115-124.
- 63 邢修三. 非平衡统计信息理论. 物理学报, 2004, 53(9): 2852-2863.
- 64 方滨兴, 许进, 李建华等. 在线社交网络分析. 北京: 电子工业出版社, 2014.

## Social Network Analysis—Key Research Problems, Related Work, and Future Prospects

Fang Binxing<sup>1,2</sup> Jia Yan<sup>2</sup> Han Yi<sup>2</sup>

(1 Beijing University of Posts and Telecommunications, Beijing 100876, China;

2 National University of Defense Technology, Changsha 410073, China)

**Abstract** Recently, with the rapid development of online social network services, such as Weibo, Twitter, and Facebook, social networks pervade nearly every aspect of our daily life. Social networks connect all aspects of our social and economic life. The role of social networks has been rapidly becoming ubiquitous platforms of communication and making connections. It plays an

important role as indispensable tools for professional networking, social recommendations, or online advertisement. Internet-based social networks consist with the most important virtual society of maintaining social relationships. Meanwhile, social networks also have far-reaching effects for national security and social development. With billions of different connections, individuals constitute a “relational structure” on social networks, which includes a large number of complex relationships, such as social communities, social ties, or linkage farms, etc. Based on the relationship of social network structure, the connected individuals gather with a large number of ongoing events. They influence each other by interactions. Thus, the individuals form a variety of networking crowds with common behavioral characteristics. Based on relational structure and social networking crowds, various kinds of information has been quickly published and disseminated, which leads to the formation of the social media. Virtual world gives feedback to the reality societies. Therefore, the virtual and reality keep interacting and influencing to each other. To fully understand the various aspects of social networking operating mechanism, this paper focuses on the analysis of a series of key and tightly knitted elements in the evolution of social networks: (1) the structural properties of social networks and their evolution; (2) social groups and their interaction law; and (3) social networking information and its dissemination. In this paper, we first introduce the scientific connotation and related research progress of these three issues, and the prospects for future research.

**Keywords** social network, topological structure, crowds, information dissemination

**方滨兴** 工程院院士, 北京邮电大学教授, 博士生导师。国家信息化专家咨询委员会网络与信息安全专委会副主任, 国家互联网信息办公室专家委员会委员; 第十一届全国人大代表, 北京市第十四届人大常委会委员; 国家“863”计划“十二五”专家委员会委员, 中国通信学会副理事长。国家重点基础研究发展计划(“973”)项目“社交网络分析与网络信息传播的基础研究”首席科学家。E-mail:fangbx@bupt.edu.cn

**Fang Binxing**, an academican of Chinese Academy of Engineering, former president of Beijing University of Posts and Telecommunications, principal representative of the 11<sup>th</sup> National People's Congress. His main research area covers network security, information content analysis, parallel processing, and internet technology and so on. He is conducting research on the computability of the social network properties and attributes. Prof. Fang is the Chief Scientist of National “973” Program on social network and information dissemination analysis. E-mail: fangbx@bupt.edu.cn



中国科学院