

## 符号网络研究综述\*

程苏琦<sup>1,2</sup>, 沈华伟<sup>1</sup>, 张国清<sup>1</sup>, 程学旗<sup>1</sup>

<sup>1</sup>(中国科学院 计算技术研究所, 北京 100190)

<sup>2</sup>(中国科学院大学, 北京 100190)

通讯作者: 沈华伟, E-mail: shenhuawei@ict.ac.cn

**摘要:** 符号网络是指边具有正或负符号属性的网络, 其中, 正边和负边分别表示积极的关系和消极的关系. 真实世界的许多复杂网络中都存在对立的关系, 尤其是在信息、生物和社会领域. 利用边的符号属性去分析、理解和预测这些复杂网络的拓扑结构、功能、动力学行为具有十分重要的理论意义, 并且对个性化推荐、态度预测、用户特征分析与聚类等都具有重要的应用价值. 然而, 当前人们对网络的符号属性关注较少. 综述了符号网络的研究背景及意义、国内外研究现状和最新进展, 并讨论了目前存在的主要问题, 试图让人们人们对符号网络这一研究方向能有清晰而全面的认识, 为网络数据挖掘、复杂网络分析、社会学、生物信息学等相关领域的研究者提供有益的参考.

**关键词:** 符号网络; 网络科学; 网络拓扑; 拓扑分析; 在线社会网络; 复杂性科学

**中图法分类号:** TP301      **文献标识码:** A

中文引用格式: 程苏琦, 沈华伟, 张国清, 程学旗. 符号网络研究综述. 软件学报, 2014, 25(1): 1-15. <http://www.jos.org.cn/1000-9825/4503.htm>

英文引用格式: Cheng SQ, Shen HW, Zhang GQ, Cheng XQ. Survey of signed network research. Ruan Jian Xue Bao/Journal of Software, 2014, 25(1): 1-15 (in Chinese). <http://www.jos.org.cn/1000-9825/4503.htm>

## Survey of Signed Network Research

CHENG Su-Qi<sup>1,2</sup>, SHEN Hua-Wei<sup>1</sup>, ZHANG Guo-Qing<sup>1</sup>, CHENG Xue-Qi<sup>1</sup>

<sup>1</sup>(Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100190, China)

<sup>2</sup>(University of Chinese Academy of Sciences, Beijing 100190, China)

Corresponding author: SHEN Hua-Wei, E-mail: shenhuawei@ict.ac.cn

**Abstract:** Signed network is a kind of network including edges with the property of positive or negative sign. The positive and negative sign represent positive relationship and negative relationship, respectively. Many real complex networks have opposite relationships, especially in social, biological and information fields. Using the sign properties of edges to analysis, understand and predict the topological structures, functions and dynamic behaviors of these real networks has important theoretical significance and practical applications, such as personalized recommendation, prediction of attitudes, user analysis and clustering and so on. However, academic research has devoted little attention to signed network. This paper reviews the background, significance, research situation and recent progress of signed networks, and discusses the main problems of existing works. This study is to provide a clear and comprehensive understanding to this meaningful research area, and to benefit to the researchers from the fields of network data mining, complex network analysis, sociology and biological information.

**Key words:** signed network; network science; network topology; topology analysis; online social network, complexity science

符号网络是指边具有正或负符号属性的网络, 其中正边和负边分别表示积极的关系和消极的关系. 具体而

\* 基金项目: 国家重点基础研究发展计划(973)(2012CB316303); 国家科技支撑计划(2012BAH39B04); 国家自然科学基金(61232010, 61202215, 61174152); 北京市自然科学基金(4122077)

收稿时间: 2012-07-26; 修改时间: 2013-06-09; 定稿时间: 2013-09-09; jos 在线出版时间: 2013-11-01

CNKI 网络优先出版: 2013-11-01 13:49, <http://www.cnki.net/kcms/detail/11.2560.TP.20131101.1349.002.html>

言,符号网络中的正边可以表示朋友、信任、喜欢、支持等积极关系,使用正号“+”标识,而负边通常用于表示敌人、不信任、讨厌、反对等消极关系,使用负号“-”标识.在社会、生物和信息领域,很多复杂系统中都存在对立关系,例如:社会领域中,人与人之间存在朋友与敌人关系<sup>[1]</sup>;国际关系中,有合作与敌对关系<sup>[2]</sup>;生物领域里,神经元之间存在促进和抑制关系<sup>[3]</sup>;信息领域中,用户在社交网站或社交媒体上可以对其他用户表达信任或不信任态度<sup>[4]</sup>、标注朋友或敌人关系<sup>[5]</sup>,还可以投票赞成或反对一个用户的管理员提名<sup>[6,7]</sup>.另外,在线社区中,用户观点也存在相近与对立关系<sup>[8]</sup>.除上述具有明确的正、负关系标识的情形以外,还有一些复杂系统的正负关系是隐含的.如万维网中一个页面包含的超链接既可能表示对链接目标页面的赞同,也可能是反对,这需要通过分析两个页面的语义来判定<sup>[9]</sup>.这些复杂系统都可以抽象为符号网络而加以描述.

对符号网络的研究最初集中于社会学领域.早在20世纪40年代,Heider就基于社会心理学理论探讨了人作为认知主体的三角形关系中消极关系与积极关系的相互作用模式<sup>[1]</sup>,后由Cartwright和Harary等人<sup>[10]</sup>用图论语言进行描述并推广至整个网络.此后,人们主要关注于研究带符号的社会网络的结构和演化问题,致力于理解和揭示社会群体中的派系结构与发展规律,取得了不少研究成果.但是由于研究对象绝大部分为小规模社会网络,并且对网络的观察时间也有限,研究具有局限性.

信息技术领域的快速发展和复杂网络研究的兴起,为符号网络的研究和应用带来了新的机遇和挑战.一方面,随着信息技术领域的发展,无处不在的网络应用和海量的数字化网络数据给研究者提供了丰富的研究对象.以时下流行的在线社会网络为例,部分在线社会网络具有明确的符号标识,并且全程记录了整个网络的演化过程,比如消费者评论网站 Epinions、技术新闻评论网站 Slashdot、协同编辑百科全书 Wikipeda 的投票网络,还有包含多种对立关系的在线游戏网络等等,它们为符号网络的研究提供了良好的研究案例.在这些网络上的研究已表明,真实网络的演化既与传统社会学中的符号网络演化理论有较高的一致性,同时还受到其他机制的作用<sup>[6]</sup>.然而,由于在线社会网络的规模都很巨大,这对传统的符号网络分析方法提出了挑战.另一方面,随着复杂网络研究的发展,将复杂系统抽象为网络进行研究的思路和方法得到了学术界和工业界的普遍认可.除传统社会网络外,其他还有许多真实复杂系统中包含着对立关系,比如神经网络、万维网、信任网络等等,都可以抽象为符号网络进行研究.人们越来越意识到,基于符号属性去研究这些网络对准确认识这类复杂系统和其上的应用设计具有重要意义.例如,结合负边信息能在语义网络上更准确地识别话题<sup>[11]</sup>,能在社交网站上更有效地进行推荐等<sup>[7]</sup>.总体而言,近期符号网络研究在社会、生物,尤其是信息技术领域呈现增长的趋势.这些都说明,符号网络在现阶段具有重要的研究意义和应用价值,并已引起不同领域研究人员的关注.

符号网络研究的关键在于综合利用正、负边信息.现有的针对无符号简单二值网络的研究思路和方法并不适用于符号网络,采用任何方式将符号网络转为简单二值网络,要么丢失传统测度意义的延续性,要么丢失部分或全部的负边信息.一方面,负边对网络的形成、结构演化、动力学等都具有重要的影响,不可忽略;另一方面,负边的引入在提高研究准确性的同时,对研究思路和方法也提出了挑战.如何正确定位负边的作用与意义,合理处理正边和负边之间的作用关系,是符号网络研究的关键和难点.

本文第1节介绍符号网络的基础理论,包括结构平衡理论和地位理论.第2节介绍符号网络静态拓扑分析的思路和方法.第3节介绍符号网络演化模型和动力学研究进展.第4节介绍符号网络研究在信息领域的应用.第5节总结符号网络的研究现状,讨论目前面临的主要困难并进行展望.目前,符号网络领域的研究文献集中在国外,国内较为欠缺.鉴于符号网络的重要研究意义和实用价值,我们有必要总结和学习该领域现阶段的研究成果,并深入分析和预测其发展趋势,期望能够更好地指导未来的研究工作.因此,本文撰写的目的不在于对比分析符号网络研究方法的优劣,而在于概述该领域的研究问题、研究方法和最新进展,以期为社会学、生物学、信息学等不同学科领域的研究人员提供一些有益的参考.

## 1 符号网络的基础理论

符号网络领域中最基础的理论是结构平衡理论(structural balance theory),最早在1946年由Heider提出<sup>[1]</sup>.该理论的提出既拉开了符号网络研究的序幕,也为之打下了坚实的理论基础.该理论在提出之后被不断完善与

发展,并一直作为符号网络研究的主导理论.而近期,Leskovec 和 Kleinberg 等人根据在多个在线社会网络上的实证分析,提出了一个针对有向网络的地位理论(status theory)<sup>[6]</sup>,为符号网络研究带来了新的理论补充.

### 1.1 结构平衡理论

结构平衡理论的基础是社会心理学理论,起源于 Heider 提出的人们对事物态度的平衡模型<sup>[1]</sup>.该模型把人与人或人与物之间的关系分为积极和消极两种类型,并实证分析关系类型的演化规律.Cartwright 和 Harary<sup>[10]</sup>将 Heider 的理论进一步加以推广,用数学语言将包含积极和消极两类关系的系统形式化地描述为符号网络,网络中用边的正、负符号分别表示积极关系和消极关系.

结构平衡理论围绕三角形的平衡性分析开始.在符号网络中,考虑符号性后三角形的关系具有如图 1 所示的 4 种模式.结构平衡理论认为,图 1 中,图 1(a)和图 1(b)对应的两种三角形是结构平衡的,而图 1(c)和图 1(d)对应的两种三角形是结构不平衡的.上述三角形的结构平衡性判定是从社会和心理角度出发,综合而言,该判定可以简单概括为以下 4 个直观认识:1) 朋友的朋友是我的朋友;2) 朋友的敌人是我的敌人;3) 敌人的朋友是我的敌人;4) 敌人的敌人是我的朋友.其中,第 1 个直观认识解释了图 1(a)中三角形的结构平衡性问题,其余 3 个直观认识解释了图 1(b)中三角形的结构平衡性问题.而图 1(c)中的三角形不满足上述 4 个直观认识的前 3 个,图 1(d)中的三角形不满足第 4 个直观认识,因此,这两种三角形均被认为是结构不平衡的,存在向前两种三角形转化的倾向.

三角形的结构平衡性可以通过三角形 3 条边的符号之积来判定:若为正,则该三角形是结构平衡的;否则是结构不平衡的.基于该认识, Cartwright 和 Harary 进一步将针对三角形的结构平衡分析扩展到分析环的结构平衡性,利用环中所有边的符号之积来判定环的结构平衡性.具体而言,当且仅当环上所有边的符号之积为正时该环是结构平衡的,也即一个结构平衡的环中含有偶数条负边;相应地,具有奇数条负边的环是结构不平衡的.如果一个符号网络中所有的环都是结构平衡的,则称该网络是平衡网络. Cartwright 和 Harary 证明,一个符号网络是平衡网络的充分必要条件是:网络的节点集合能够被分割为两个子集,每个子集内的所有边均是正边,子集间的边均为负边.因此,尽管结构平衡理论的定义是基于三角形等微观结构,但该理论同样能够判定网络全局的结构平衡性.

真实的符号网络往往存在两个以上的对立子群,而 Heider 结构平衡约束下仅可能出现两个子群.为揭示该现象, Davis 放宽了 Heider 结构平衡理论的约束条件,提出了“弱结构平衡理论”<sup>[13]</sup>.该理论中放弃了 Heider 结构平衡理论中的第 4 个直观认识“敌人的敌人是我的朋友”,将 3 条边的符号均为负号的三角形视为结构平衡的.这样扩展得到的“弱结构平衡理论”带来了新的概念“ $k$ -平衡网络”,是指满足下述条件的网络:网络中节点可以分成  $k$  个子集,使得子集内部节点间的边全为正边,连接不同子集节点的边全为负边.根据  $k$ -平衡网络的定义,2-平衡网络是满足 Heider 结构平衡理论的平衡网络. Davis 进一步证明,一个网络是  $k$ -平衡网络当且仅当网络中不存在只含有一条负边的环.

近几年,对真实网络的实证分析证实了 Heider 结构平衡理论和扩展后的弱结构平衡理论<sup>[6,14]</sup>. Leskovec 等人<sup>[6]</sup>分析了多个大规模带符号在线社会网络.统计显示,当忽略边的方向性时,满足结构平衡的三角形(三边全正、两负一正)数量显著多于不满足结构平衡的三角形数量. Szell 等人<sup>[14]</sup>用符号网络描述了一个在线游戏中玩家间的关系.在该网络上,满足结构平衡的三角形所占的比例随时间日益增加,而不满足结构平衡的三角形所占的比例随时间降低.这些实证分析验证了结构平衡理论的重要性.但是,绝大部分的真实符号网络甚至不能严格满足弱结构平衡的条件.因此,结构平衡理论在符号网络中仅具有统计意义上的正确性.

近几年,对真实网络的实证分析证实了 Heider 结构平衡理论和扩展后的弱结构平衡理论<sup>[6,14]</sup>. Leskovec 等人<sup>[6]</sup>分析了多个大规模带符号在线社会网络.统计显示,当忽略边的方向性时,满足结构平衡的三角形(三边全正、两负一正)数量显著多于不满足结构平衡的三角形数量. Szell 等人<sup>[14]</sup>用符号网络描述了一个在线游戏中玩家间的关系.在该网络上,满足结构平衡的三角形所占的比例随时间日益增加,而不满足结构平衡的三角形所占的比例随时间降低.这些实证分析验证了结构平衡理论的重要性.但是,绝大部分的真实符号网络甚至不能严格满足弱结构平衡的条件.因此,结构平衡理论在符号网络中仅具有统计意义上的正确性.

### 1.2 地位理论

结构平衡理论为分析无向符号网络提供了基础,但在刻画具有方向性的符号网络时存在较大偏差.为此,

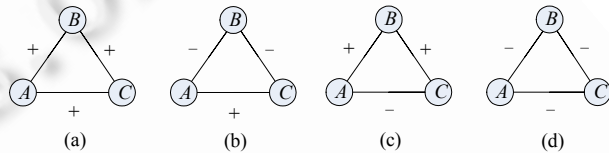


Fig.1 Relationship patterns of triads in signed network<sup>[12]</sup>

图 1 符号网络中三角形的关系组合<sup>[12]</sup>

Leskovec 和 Kleinberg 等人<sup>[6]</sup>通过实证分析大规模在线社会网络中用户建立连边的意图以及网络中三角形的关系模式,提出了适用于有向符号网络的新理论——地位理论.该理论认为,边的符号取决于节点地位的差异,具体表述如下:一条由  $A$  到  $B$  的正边表示  $A$  认为“ $B$  的地位比  $A$  高”,一条由  $A$  到  $B$  的负边表示  $A$  认为“ $B$  的地位低于  $A$ ”.这种地位高低关系具有传递性.

同样地,地位理论最初也用于分析 3 个节点构成的三角形.图 2 中给出了地位理论认为合理的三角形关系组成模式,共有 8 种.可以看到,地位理论中认可的关系模式有一部分与结构平衡理论一致,如图 2(a)、图 2(d)、图 2(e)和图 2(h)所示,但也有不一致的情形,如图 2(b)、图 2(c)、图 2(f)和图 2(g)所示.图中所示的 8 种关系模式的统计显著性已在真实网络中得到确认.Leskovec 等人还进一步验证了该理论在全局上的显著性<sup>[7]</sup>.在有向符号网络中,如果所有的三角形都满足地位理论,那么该网络中的节点可以形成一个特殊的节点序列,使得网络中所有的正边均是由排名低的节点指向排名高的节点,负边均由排名高的节点指向排名低的节点.若将网络中的负边全部转换为逆向的正边,寻找这种特殊全局节点序列的问题即转换为在全正边网络上寻找最大无环子图问题.通过在多个在线社会网络上的分析发现,能找到的最大无环子图中包含的边数高达 80%~85%,显著高于关系随机组合的情况.这在一定程度上从全局角度证实了地位理论的存在性.

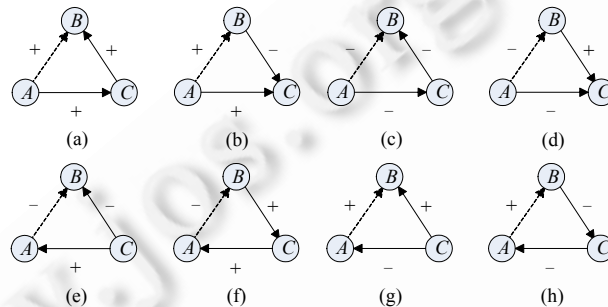


Fig.2 In status theory, only eight relationship patterns of triads shown here are reasonable<sup>[6]</sup>

图 2 在地位理论中认可的 8 种关系组合模式<sup>[6]</sup>

比较而言,结构平衡理论可视为对喜好的模拟,即喜欢和讨厌;而地位理论则是基于个人身份地位的判定,与喜好无关.这两种理论对于三角形结构动力学的演化既有相同之处,又有不同之处.再者,结构平衡理论对连边的方向性并无要求,而身份理论只适用于有向网络.研究表明,地位理论和结构平衡理论具有互补性,相结合能够有效地预测符号网络中边的符号<sup>[7,15]</sup>.然而,这两种理论在符号网络演化中的角色至今还未有明确的认识.

## 2 符号网络静态拓扑分析

对网络进行静态拓扑分析,往往从小尺度、大尺度和中尺度这 3 个层次展开,符号网络也不例外.对于符号网络,在小尺度上,核心问题是如何对网络中节点或边的特征进行度量和分析;大尺度上主要关注平衡网络的判定以及网络不平衡程度的度量;中尺度上的研究工作主要围绕社区结构的识别与分析而展开.下面分别针对 3 个尺度在符号网络的现有研究成果进行介绍.

### 2.1 小尺度结构分析

在小尺度层面,即单个节点或边的层面,现有研究主要关注于如何把无符号网络上的节点或边的度量扩展到符号网络上.节点的中心性度量是该层面研究的主要关注点.与传统的基于简单二值网络的度中心性相对应,符号网络中节点具有正度值和负度值两个基本属性<sup>[7]</sup>.FMF(fans minus freaks)<sup>[5]</sup>被定义为节点正、负度值的绝对值之差,可用于衡量网络中节点的受欢迎程度.度中心性的其他变种,如特征向量中心性<sup>[16]</sup>、 $c(\beta)$ 中心性<sup>[17,18]</sup>、PageRank<sup>[19,20]</sup>、谱排序<sup>[19]</sup>等值也被扩展到符号网络中用于度量节点的受欢迎程度或信任程度.其中,特征向量中心性还能够指示网络的聚集结构<sup>[16]</sup>.在符号网络中,负边对于一个节点的中心性或地位等的作用既可能是积极的,也可能是消极的.以特征向量中心性为例,负边的意义表现为:与对立的节点建立负边会提升一个节点的



地位,而与高地位的人建立负边则会降低该节点的地位.图 3 中给出了负边对节点特征向量中心性影响的一个简单示例,其中,图 3(a)的网络 *A* 和图 3(b)的网络 *B* 的区别仅在于节点 3 与节点 5 之间边的正负性.从图 3(c)列表中给出的归一化后节点的特征向量中心性可以看到:由于节点 5 与具有最高中心性的节点 3 之间存在负边,从而网络 *B* 中节点 5 的特征向量中心性变为负值;与此同时,与节点 5 之间存在正连接的节点 4 的特征向量中心性值也大幅降低了.由此可见,负边对节点或边的重要性或地位的评价具有重要影响.总体而言,如何引入更多具有针对性的符号网络测度,正确而合理地利用负边的信息还需要深入研究;与此同时,保持传统的那些网络测度在符号网络上物理意义的延续性,也是一个值得探讨和关注的研究问题.

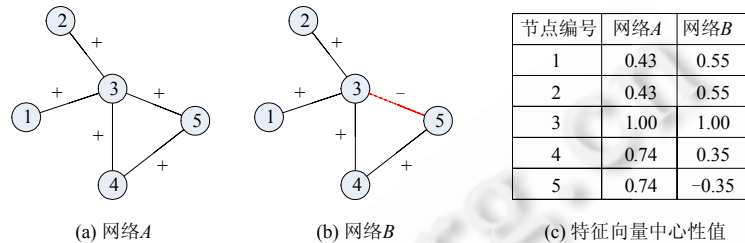


Fig. 3 Impact of negative edge on the eigenvector centrality<sup>[16]</sup>

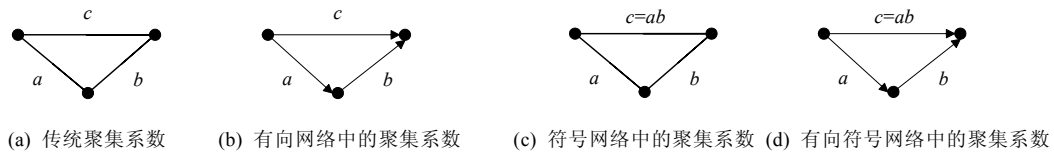
图 3 负边对节点特征向量中心性值的影响<sup>[16]</sup>

## 2.2 大尺度结构分析

判断一个符号网络是否是平衡网络,是符号网络大尺度分析的一个基本问题.现已有线性时间的算法能够解决这一问题<sup>[21-23]</sup>.文献[21]中给出的一种方法为:先获取给定符号网络的一棵最小生成树,然后从任意选择一个源点出发给各节点赋 1 或 -1 的符号值.除源节点外,各节点的符号值赋为其父节点的符号值与相连边上符号的乘积.在给所有节点赋完符号值之后,再遍历原图中的各边.如果存在一条边的符号不等于两端节点的符号值之积,该网络是非平衡的;否则为平衡网络.对弱平衡网络的判定更为简单,寻找所有正边组成的连通片,如果每个连通片内部的节点间不存在负边,那么该网络即满足弱平衡结构的条件.并且,对于 *k*-平衡网络,将所有负边移走后所得到的连通片分布即为 *k*-平衡状态下的网络划分.

然而,现实中的符号网络绝大多数不是平衡网络,甚至不满足 *k*-平衡网络要求,因而,度量网络的不平衡程度成为人们关注的重点.网络不平衡程度是指网络当前的结构状态与完全平衡状态的差距.最初,心理学和社会学理论假定了网络向平衡结构发展的趋势,因此,判断一个符号网络是否在向平衡状态演变以及目前已演化到何种程度具有重要意义.尽管判断网络是否平衡存在有效算法可解,但精确度量当前网络与平衡网络的距离却往往很困难.Harary 总结了两大类测度:第 1 类测度是采用所有平衡的环路占全网所有环路的相对比重进行度量;另一类测度是针对那些对非平衡状态有贡献的边的取缩量.

第 1 类测度的复杂度取决于要计算的环路类型及平衡的判定方式.目前,仅统计网络中平衡三角形数量占全网三角形数量的比例的方式被广泛使用<sup>[5,6,14]</sup>.Kunegis 等人<sup>[5]</sup>将该方式数学化,称为符号聚集系数(signed clustering coefficient),并通过增加对方向性的约束,将该概念进一步扩展到有向无权符号网络中.图 4 中给出了符号网络中聚集系数的示意图:图 4(a)所示为最基本的聚集系数定义下认可的三角形,即只要 3 个节点间各自有边相连即可;图 4(b)所示为有向网络下的聚集系数;图 4(c)所示为无向符号网络中的聚集系数,即要求第 3 边的符号等于另外两边符号之积,可知这正好是结构平衡理论中认可的平衡三角形;图 4(d)所示为有向符号网络下的聚集系数,也即在传统的有向图上同样增加了对符号模式的要求.该指标计算时的最坏情况为:对任意 3 个节点均检查它们是否构成三角形并判断其平衡属性,时间复杂度为  $O(n^3)$ ,其中, *n* 为网络中节点总数.然而在稀疏网络中,实际上只需要通过对每个节点分别检查其邻居之间是否存在连边来计算三角形个数,因而其时间复杂度会远低于上述量级.除这些仅考虑局部三角形平衡性的指标外,人们还提出了从全局考察网络中所有不同长度的环平衡状态的指标,通过加权累积的方式计算整个网络的不平衡程度<sup>[24,25]</sup>.这些指标需要统计所有可能长度的环,一般利用网络邻接矩阵的连乘与求逆操作进行运算,整体时间复杂度为  $O(n^3)$ .

Fig.4 Illustration of signed clustering coefficient<sup>[5]</sup>图4 符号聚集系数示例<sup>[5]</sup>

第2类测度的计算更为复杂,度量的是可使当前网络转变为平衡网络需要移除或修改符号的最少边数,其中对应的边集合被称作“negation-minimal”<sup>[10,26,27]</sup>,该定义可以扩展到  $k$ -平衡网络上.这个问题是 NP 完全问题,与最大割问题等价,可以直接使用现有的一些多项式时间的优化算法,目前能够保证的最佳近似比  $\alpha \approx 0.878$ .

谱分析的方法也被运用于符号网络不平衡程度的刻画.网络的拉普拉斯矩阵的最小特征值即是这样一个指标<sup>[23]</sup>,是立足于刻画网络环路中的不平衡信息.

然而,以上这些测度各自具有局限性:基于网络中环路的测度仅关注网络组成单元的平衡信息,而忽略了网络更大范围乃至全局角度的平衡信息;谱分析虽然从全局角度出发,但又无法揭示网络中的不平衡区域.因此,Facchetti 等人<sup>[28]</sup>提出利用伊辛自旋玻璃(Ising Spin glasses)的模型<sup>[29-31]</sup>来计算网络的基态以表示网络的平衡状态.他们提出了一种高效的启发式方法对该平衡状态进行求解,在3个大型在线社会网络上的研究表明,这些真实网络目前处于极端平衡状态,这与以往的认识大为不同.

### 2.3 中尺度结构分析

在中尺度层面,符号网络的社区结构发现是重要内容.社会学和复杂网络领域对此问题都有一部分研究工作,但两个领域的关注点有所不同:社会学领域主要关注于结构平衡,希望找到满足结构平衡的节点组划分;复杂网络领域则更多地从社区内的连边密度、社区间的连边稀疏程度等角度来进行社区发现<sup>[32]</sup>.

#### 2.3.1 社会学角度的中尺度结构分析

社会学领域中,针对符号网络的中尺度结构分析主要关注于符号网络的分割问题,即把符号网络按照某种标准分割成几个分块.研究方法大多以结构平衡理论作为符号网络分割的依据,寻找尽可能满足分块内部全为正边、分块间全为负边的理想分割方式,也即与  $k$ -平衡网络最接近的分割.基于该认识,Doreian 等人<sup>[33]</sup>提出了一种误差函数用于度量当前划分与理想划分之间的差异,该误差函数定义为分块内部的负边数和分块之间的正边数的加权和,通过最小化该误差函数,能够找到尽可能接近理想划分的网络分割方式.随后,该思路被扩展到一般性的块模型<sup>[34]</sup>,并通过引入多重指标来解决一般性的块模型中存在的过拟合问题和对噪音数据的敏感性<sup>[35]</sup>.Brusco 等人<sup>[36]</sup>提出了一种基于分枝界定思想的分割算法,能够得到符号网络划分的最优解.基于分枝界定的算法能够得到精确解,但其运行效率与网络的性质关联很大.基于上述原因,Brusco 等人提倡尽可能同时使用传统的启发式算法与这种分枝界定算法,而不是只依靠其中一种.前述这些方法中,对于目标函数中正、负边权值的设定对最终获得的分块结果的影响目前还欠缺讨论.除上述基于平衡理论的分割以外,还有学者指出可以依据特征向量中心性识别网络中的聚类结构,利用二分法和递归进行符号图分割<sup>[16]</sup>.具体地,对一个给定的符号网络,计算所有节点的特征向量中心性,按该值的正和负将所有节点划分成两个子块,之后再针对这两个子块重复上述划分.该算法的优势在于,能够直接获得最终分块结果以及网络的层次结构,然而由于该方法的出发点与  $k$ -平衡网络无直接关联,因此无法保证最终所得分块与  $k$ -平衡网络之间的接近程度.另外,还有部分工作研究了带符号的二模网络(signed two-mode network)<sup>[37,38]</sup>和带自环的符号网络<sup>[39,40]</sup>上的分割问题.

#### 2.3.2 复杂网络角度的中尺度结构分析

复杂网络领域针对符号网络的中尺度结构分析以社区发现为主,与社会学领域中的符号网络分割有些不同:社会学研究的目标是使网络分割与网络平衡状态尽可能地接近;而社区发现更侧重于社区内边的稠密性和社区间的稀疏性,对于符号网络还注重社区内和社区间正负边的情况.另外,社会学里符号网络分割算法不适用于全正边的网络,而社区发现算法则希望能够兼容无符号网络.目前,符号网络中的社区发现有两种思路.

第 1 种思路是把社区看成网络的一种划分。

该思路首先设计刻画网络划分优劣的度量,然后通过优化该度量求解最好的划分从而发现社区。Gómez 等人<sup>[41]</sup>和 Traag 等人<sup>[42]</sup>改进了传统针对全正网络的社区划分评价函数——模块度(modularity)——以适用于符号网络。这两个工作的主要思路都是将正、负边构成的子图拆开来单独计算模块度,然后将这两个模块度值加权求和得到整个网络的模块度。由于模块度函数中加入了连边密度的考虑,从而得到的划分符合传统意义上对社区的定义。Gómez 等人分析了一个零售店网络, Traag 等人则研究了一个具有敌对和联盟关系的国际关系网。实证分析结果表明,所提出的优化目标对应的划分与传统方法所得结果有较大差异,能够更精确地聚类。Li 等人<sup>[43]</sup>提出了一种结合 GN(Girvan-Newman)算法和层次聚类算法的 GN-H 协同训练算法。该算法的第 1 阶段利用 GN 算法对正边构成的子图进行划分,后一阶段利用结合负边信息的评价指标确定整个网络的最终层次聚类。由于 GN 算法计算复杂度较高,为  $O(m^2n)$ ,其中,  $m$  为边数,  $n$  为节点数,从而该算法难以适用于大规模网络。还有一部分工作由于应用场景需求,在社区划分中往往仅注重于社区内和社区间正、负边的分布。Bansal 等人<sup>[44]</sup>受文档聚类和不可知论学习的启发,在文档网络中引入了关联聚类的概念,并定义了两种聚类目标:一是最大化聚类内正边与聚类间负边之和,二是最小化聚类内负边与聚类间负边之和。这里的聚类即等价于社区。他们证明:在一个完全符号网络下,这两个优化问题均是 NP-hard 问题。Demaine 等人<sup>[45]</sup>进一步在一般性符号网络上研究了这两个问题的性质,证明它们均是 APX-hard 的,与最小割集问题等价,并基于线性规划舍入和区域增长技术提出了一种  $O(\log n)$  的近似算法。Larusso 等人<sup>[8]</sup>依据最大化社区内部正边权重和社区间负边权重的原则提出了一个系统能量函数,并利用模拟退火的方法进行求解。随后,他们将该能量函数修改为最小化组内的负边权重以及组间的正边权重,同样使用模拟退火的方式求得近似解。他们的工作体现在文献[44]和文献[45]中,其不同在于目标函数中引入了边上的权重,但同样未考虑连边密度。因此,后面这几项工作所得到的社区划分与社会学角度所得到的划分本质上相同。另外,还有学者利用谱聚类的方法进行社区发现<sup>[23]</sup>。

第 2 种思路是进行抽取式社区发现。

这种思路强调从局部出发进行社区发现。Yang 等人<sup>[46]</sup>提出一种通过随机游走来寻找符号网络中社区的算法 FEC(finding and extracting community),该算法首先随机地从一个还未标识其社区的节点出发进行随机游走,考察游走一定步数所能到达的节点集和相应概率,在游走过程中忽略掉所有负边,然后再根据一个截断函数确定哪些节点与随机游走的初始节点属于同一个社区。这个截断函数中充分考虑了分区自身内部以及与网络剩余部分之间的正、负边分布情况。FEC 在时间和识别精度方面表现出良好的性能,时间复杂度为  $O(Kl(n+m))$ ,其中,  $K$  为发现的社区数量,  $l$  为随机游走的步长,  $n$  和  $m$  分别为节点数和边数。然而,该算法中存在一些不确定因素会造成其性能不够稳定,如初始种子节点的选择、随机游走步长的设定等。针对这些不足,孔令旗等人<sup>[47]</sup>通过添加选择目标节点功能,改用自动步数探测等方法对 FEC 算法进行了改进。Shama 等人<sup>[48]</sup>提出了 CRA(clustering re-clustering algorithm)算法,由两阶段组成:第 1 阶段是基于广度优先聚类(breadth first clustering)算法,在全正网络上进行分割;第 2 阶段利用上一阶段的输出作为输入,依据一个指示节点参与程度的测度对拥有负边的节点重新定位其所属社区。在此,节点的参与程度定义为节点在某社区中的正边数与在该社区的总边数之比。节点会被划分到参与程度值最高的社区。该方法的优势在于无需设置参数。Huang 等人<sup>[11]</sup>既讨论了无重叠社区的情形,也讨论了重叠社区的情形,并提出相应的启发式算法求解。他们利用 CPM(clique percolation method)等传统的社区发现算法对社区进行二次分割。前述这些抽取式算法与划分式社区发现算法相比,其优点在于能够处理重叠社区,并能更好地胜任只需要发现特定局部社区的场景。

总的来说,目前符号网络上的社区发现方法要么是直接改进传统社区划分度量函数,要么是采用如下的两阶段式处理方式:首先针对全正边网络利用传统方法进行社区发现,然后利用符号网络特定的社区划分质量评价函数来调整分区。后一种方式很灵活,能够有效地利用现有的社区发现算法并结合各种特定需求的符号网络分区质量评价函数。然而,由于前期分区结果往往对第 2 阶段的调整造成较大影响,从而这种方式不能保障充分利用了负边信息。从这个角度将这两个阶段进行合理融合,形成一个统一的过程是一个值得研究的问题。另外,目前,在这两种方式中,对符号网络社区的质量评价函数中正、负边权重的讨论还较为欠缺,在实际应用中仍有

待进一步明确。

### 3 符号网络的演化动力学

符号网络的演化动力学绝大部分是基于结构平衡理论。虽然结构平衡理论表面上是一个静态理论,仅仅关注网络局部和全局是否结构平衡,但其本质思想是动态的——非结构平衡的三角形会向结构平衡的三角形演变<sup>[1]</sup>。基于结构平衡理论,人们对符号网络的演化动力学进行了广泛而深入的研究。

符号网络演化动力学早期的研究大多基于离散时间步的模拟。为验证基于结构平衡理论的符号网络演化过程是否总能到达全局结构平衡状态,Wang 等人<sup>[49]</sup>基于蒙特卡罗模拟方法提出一个离散时间的符号网络演化模型。该模型如下设计:初始时,网络中边的符号随机配置,每个时间步依次检查每个三角形是否是结构平衡的,对于非结构平衡的三角形,随机选择其包含的一条边改变符号。使用该模型,Wang 等人分析了不同规模的符号网络,发现这些符号网络最终均演化到结构平衡状态。当符号网络处于平衡态时,网络节点分处于两个完全对立的节点集合。当初始正边比例大时,到达平衡时,规模较大的节点集合的规模远大于规模小的节点集合;而当初始负边比例大时,达到平衡时,两个节点集合的规模近似。随后,Antal 等人<sup>[30,50]</sup>提出了类似的离散时间的符号网络演化模型,该模型将三角形按照包含的负边数分为 4 类: $\Delta_0, \Delta_1, \Delta_2$  和  $\Delta_3$ ,即具有  $j$  条负边的三角形记为  $\Delta_j$ 。按照结构平衡理论, $\Delta_0$  和  $\Delta_2$  是结构平衡的, $\Delta_1$  和  $\Delta_3$  是非结构平衡的。初始时,网络中边的符号随机配置。每个时间步随机选择的一个三角形,若是  $\Delta_3$ ,其中的一条边符号将会被改变,从而得到  $\Delta_2$ ;如果是  $\Delta_1$ ,则以概率  $p(0 \leq p \leq 1)$  转变为  $\Delta_0$ ,或以概率  $1-p$  转变为  $\Delta_2$ 。研究表明,网络获得平衡状态的时间依赖于网络规模,且与  $p$  值相关。另外,模型在模拟时存在相变点:当正边比例小于该相变点时,网络最终达到的平衡状态由两个完全对立的节点集合构成,正边比例越小,两个节点集合的大小越接近;正边比例越大,则两个节点集合的大小差异越大。最后,当正边的比例大于相变点时,整个网络在达到平衡态时仅包含一个节点集合,即所有边均为正边。

在上述符号网络的演化模型中,节点间的关系被描绘成离散的变量,使得网络演化模型模拟结果依赖于不平衡三角形进行边符号调整的顺序。为解决该问题,Kulakowski 等人提出了一个连续时间模型<sup>[51]</sup>。他们使用一个  $n \times n$  的实对称矩阵  $X$  来表示一个全连通社会网络,实体  $x_{ij}$  的绝对值表示节点  $i$  和  $j$  之间的正关系和负关系的强度, $x_{ij}$  的符号即对应边上的符号,从而,符号网络的演化模型可用如下边权重变化的微分方程来刻画:

$$\frac{dx_{ij}}{dt} = g_{ij} \sum_k x_{ij} x_{kj} \quad (1)$$

其中, $g_{ij} = 1 - (x_{ij}/R)^2$  用以确保  $x_{ij}$  的绝对值限制在  $[-R, R]$  内。初始  $x_{ij}$  的值是在范围  $(x_m - \varepsilon, x_m + \varepsilon)$  中随机选择,默认情况下, $x_m$  被设置为 0。当前状态与平衡状态的偏离程度用非平衡三元组的个数来衡量。仿真实验结果表明:对于任意的初始条件和网络,在有限时间步内,系统都会达到 Heider 平衡状态,其中,大部分测试网络演变为两个完全对立的分派。Gawronski 等人<sup>[52,53]</sup>将该模型应用于多个网络中,包括经典的 BA 无标度模型、Natchez 的妇女网络和 Zachary 空手道俱乐部网络,验证了模型的有效性。Marvel 等人<sup>[54]</sup>也深入研究了上述连续时间模型,给出了一个闭合表达式,其中,派系成员作为初始条件的函数。他们的研究表明:在大规模社会网络中,初始的正边数量决定了网络最终的演化结果——两个相互对立的派系并存和仅存在一个派系。Gawroński 等人<sup>[55]</sup>将该模型扩展到关系不对称的情形,发现网络达到平衡所需要的时间呈现重尾分布,并且关系的不对称性不利于网络走向平衡。

除上述这些模型外,还有一些其他的符号网络动力学建模思路。比如,有学者使用统计物理的建模方法。文献<sup>[56,57]</sup>探讨了非平衡网络和不稳态自旋玻璃模型之间的相近性。结果表明:在符号网络中,借鉴能量谱(landscape)、自旋玻璃和不稳定性(frustration)的思想是可行的。在自旋玻璃模型的基础上,可以研究符号网络中单节点的状态演化行为。此外,Szell 等人<sup>[14]</sup>提出了 STC(signed triadic closure)模型,能够模拟一个大规模在线游戏网络中不同类型三角形的演化过程。Malekzadeh 等人<sup>[58]</sup>基于博弈论提出了一个动态模型,其中,各节点调整关系的策略是使自身参与的平衡三角形尽可能地多。研究发现,该模型下任意初始状态的网络最终都会达到平衡。

有关符号网络演化动力学目前还仅限于学术性研究,主要目的在于探讨具有符号属性的真实系统可能的演化模式和相应机制的作用分析,加深人们对这类系统的理解与认识。虽然这部分研究成果目前还无具体的应



用,但存在潜在的应用场景,比如舆论引导,可以通过定位一些关键节点或边,对当前符号网络加入少量的干扰,使其最终演化收敛到所希望的状态,如所有用户观点的统一状态.

#### 4 符号网络研究在信息领域的应用

前面我们介绍了符号网络的基础理论、符号网络的静态结构、符号网络的动态演化等方面的研究,本节将介绍上述研究成果在信息领域的应用.在此,我们主要关注于信息领域中的在线社交网站与社会媒体.根据应用场景的不同,符号网络在这两类平台上的应用主要包括个性化推荐、态度预测、用户特征分析和聚类等.除此之外,符号网络在信息领域的其他方面也有一些具体应用,如垃圾站点识别、女巫攻击(sybil attack)防御等.本节将依次介绍符号网络在这些应用场景下的具体应用及一些代表性方法.

##### 4.1 个性化推荐

符号网络研究在推荐系统中已有较多应用.研究者们致力于结合正负关系以更有效地对用户进行项目(item)推荐.这里,根据应用场景的不同,项目既可以是各种物品,如评论、音乐、电影、商品等等,也可以是网络中的其他用户.后一种情形的推荐即为朋友关系推荐.

Victor 等人<sup>[59,60]</sup>提出了一个适用于推荐系统的信任模型,该模型中使用一个由信任评分与不信任评分构成的二元组来表示节点的信任程度.如果一个用户接收到的入边中存在负边,那么他的信任程度会更低.Victor 等人定义了一系列的操作,用于计算用户信任程度的传递性.Ma 等人<sup>[61]</sup>提出了一个基于因子分析的通用框架,用于整合正、负边信息,进行用户-项目的个性化推荐.他们的研究工作依旧基于信任网络,是将用户间的信任与不信任关系直接建模为用户观点间的相似与不相似属性.正边和负边信息直接作为正则项加入到用户-项目的矩阵分解对应的目标函数中.在真实数据上的测试结果表明:这种方式相比以往协同过滤和仅基于正边的推荐算法,能够有效提高推荐的准确度.Chen 等人<sup>[62]</sup>综合了正、负边信息,解决推荐系统中新用户的冷启动推荐(new user cold start recommendation)问题.他们提出的方法包括两个阶段:在第 1 阶段,使用一般性聚类算法对用户按偏好进行聚类,然后用 PageRank 算法识别出各个类别中的专家,同时,利用负边信息剔除掉不可靠的候选用户;在第 2 阶段,识别出与新用户联系紧密的类别,并依据该类别中专家的建议进行项目推荐.除上述针对事物类的推荐以外,Symeonidis 等人<sup>[63]</sup>将符号网络应用于朋友关系推荐.他们扩展了传统的 Jaccard 系数,依据地位理论加入节点接收和发送的所有边中正、负符号的分布信息,得到相邻节点的相似性测度.在该测度的基础上,他们又进一步定义了基于传递性的非相邻节点间的相似性指标.该指标是两个节点间所有可达路径的相似度贡献之和,这里,每条路径的贡献即为该路径上各边的局部相似度之积.这种节点相似性测度中融合了网络的局部信息与全局信息,并且能够处理存在负边的情况.

总体而言,研究人员开始注意到负边在个性化推荐中的重要作用,并提出了一些启发式方法来利用负边信息,但大部分都很简单.如何更加合理和充分地利用负边信息,仍然是一个值得研究的问题.

##### 4.2 态度预测

态度预测即为推断用户对某个项目存在的潜在态度,该方向的研究用于为个性化服务提供支撑.可以设想,若能有效预测用户已存在的潜在态度,将对系统进行推荐或做其他决策起重要的指导作用.由于符号正代表了态度,因此,用户态度预测也就对应到符号网络的符号预测问题<sup>[7]</sup>.

具体地,符号预测问题可以形式化描述如下:给定一个符号网络,利用除边 $(u,v)$ 以外其他边的符号信息预测边 $(u,v)$ 的符号  $s(u,v)$ .图 5 中给出了符号预测问题的一个简单示例(在图中所示网络中,除节点  $u$  和  $v$  形成的边上的符号  $s(u,v)$ 未知外,其他各边的符号均为已知,要预测  $s(u,v)$ ).

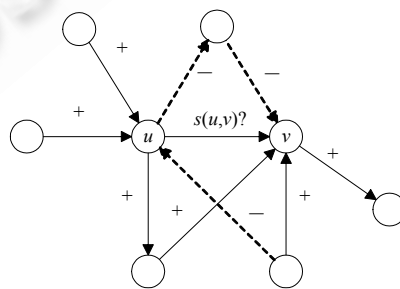


Fig.5 An example of sign prediction problem<sup>[7]</sup>

图 5 符号预测问题示意图<sup>[7]</sup>

符号预测的主流思路主要基于社会学理论(如结构平衡理论和地位理论)和机器学习.Guha 等人<sup>[4]</sup>最早研究了信任网络中边的符号预测问题,他们利用数学形式描述信任关系和不信任关系的传播行为,将信任传播描述为一系列反复的矩阵操作,并考虑了两种不信任的传播机制:一种是假定不信任关系仅会传播一次;另一种是不信任关系会连续传播.实验结果表明,前种机制更为简洁、有效.Leskovec 等人<sup>[7]</sup>首先对该问题进行了形式化,并运用机器学习的方法求解.他们考察了两类结构特征:第 1 类特征为传统的节点局部特征测度,如出度、入度等;第 2 类结构特征借鉴于结构平衡理论,是待预测边所处的所有三角形的各种关系模式.之后,他们运用逻辑回归模型对边进行分类.实证发现:符合结构平衡理论和地位理论的三元组对于待预测边的符号预测起主要作用;并且,不同真实网络训练出的预测模型间具有通用性,这表明符号网络中三元组的符号组成模式是相似的.Chiang 等人<sup>[15]</sup>提出了一个框架,用于将任意度量网络不平衡程度的指标应用于边符号预测算法的设计.根据结构平衡理论,符号网络中的有序环(包括长度大于 3 的环)都反映了符号网络的不平衡程度,从而,Chiang 等人基于 Katz 提出的一个不平衡测度指标<sup>[25]</sup>提取了特征集——长度为  $k$  的环的平衡程度集合,然后,利用分类算法进行边符号预测.实验结果表明:当  $k$  大于 5 后,增加的信息量对预测准确度影响甚微.他们的工作是从全局的角度出发来利用网络的不平衡程度信息,而 Leskovec 等人<sup>[7]</sup>则仅是基于局部信息.Kunegis 等人<sup>[23]</sup>从谱分析的角度出发设计分类方法进行边符号预测,指出,许多基于谱分析的方法可以扩展到带符号的情况.他们研究了多种不同的矩阵,如邻接矩阵、拉谱拉斯矩阵等.文献[64]利用 Spring 嵌入算法进行符号预测,该算法中假定正边两端的节点相互吸引,负边两端的节点相互排斥.实验结果表明,该方法能够达到 80%~90%的精确度.

除单纯利用网络结构信息作为特征外,节点内容属性、上下文特征等也被用于符号预测问题.文献[65]提出了利用用户的属性特征和用户间的交互特征两者相结合的分类方法,并发现,用户交互行为的特征在符号预测中的作用大于用户属性特征.Zolfaghar 等人<sup>[66]</sup>考虑了社会信任包括的 4 个关键因素:名誉、知识、相似性和基于个人的信任,并将这些因素对应到具体的特征集合.这些特征均来源于从用户行为中抽取的上下文信息和信任网络拓扑中已有的结构数据信息.基于上述特征,他们提出了一个分类系统,用于预测信任和不信任关系.该分类系统由多个分类模型组合而成,包括 SVM、RB 网络、决策树、逻辑回归模型等.这种组合模型的预测能力强于单个的传统模型.同时,他们发现,基于结构的特征比基于上下文的特征对符号预测更为重要.Borzemek 等人<sup>[67]</sup>将基于信任网络结构的特征和基于评论的特征相结合,以进一步提高边符号预测的准确度,这有助于在结构信息不足的情况下进行信任关系的预测.Maniu 等人<sup>[68]</sup>利用用户间的交互行为来推断用户间的态度,以构建符号网络,并使用机器学习方法在该网络中进行边符号预测,以证明所构建符号网络的正确性.

### 4.3 用户特征分析和聚类

符号网络研究在用户特征分析上具有重要的应用.

一方面,结合符号属性,能更准确地识别重要用户.

符号网络中认为,用户接收和发出的负边均会影响该用户的权威性与可信性.Bonacich 等人<sup>[16]</sup>利用符号网络的特征向量中心性分析了一个僧侣网络的组成结构及重要节点.他们利用该测度的正、负对网络进行划分,并依据该测度的绝对值去度量节点在对立派系中的重要性.Mishra 等人在信任网络上<sup>[69]</sup>设计了指标用于识别网络中用户的偏好性和权威性,这里,偏好性是指一个用户信任或不信任其他用户的偏好.如果一个用户总是信任或不信任其他用户,其态度则被认为是有偏的,他做出的信任评价可信性差.他们在识别出这类节点的基础上,降低这类节点给出的信任分值所占的权重,以更真实地评价其他节点的权威性.Li 等人<sup>[70]</sup>也做了类似的工作,在权威性评价中增加了无偏节点给出的信任分值所占的权重值.

另一方面,利用符号网络还能够识别一些特殊用户.

Kunegis 等人<sup>[5]</sup>提出一个测度,用于识别网络中的煽动性用户(troll),这类用户的特点是频繁发表具有攻击性、煽动性的言论.在 Slashdot 网站上的实证研究表明:对于这一类用户,考虑符号属性计算得到的 PageRank 值往往远小于忽略符号下得到的 PageRank 值;而对于普通用户,这两个值基本一致.因此,利用这两种 PageRank 值之差,定义为 Negative Rank 测度,可以有效定位此类用户.

符号网络还有一个很重要的应用即用于进行用户的聚类.根据场景的不同,聚类的依据可能是用户立场、

观点或兴趣的相近性。Traag 等人<sup>[42]</sup>将基于模块度优化的符号网络划分算法应用于一个具有敌对和联盟关系的国际关系网络。通过调整算法中的参数,可以识别出不同层次的政治集团结构。Li 等人<sup>[43]</sup>从用户产生内容中提取用户交互行为及相应的情感属性(用于标识正、负),构建了一个带用户情感倾向的交互网络,这正是符号网络。然后,他们利用两阶段式符号网络社区发现方法进行用户聚类:先采用传统社区发现算法处理仅含有正边的网络,之后结合负边信息调整前一阶段的分区。实证分析证明:引入情感信息(负边)后,所识别的社区更为精准,社区内话题更为集中。另外,所识别的社区中的意见领袖虽然从全局来看度中心性排序可能不高,但在自身所处的局部社区中影响力很大。

#### 4.4 其他应用

符号网络的研究也被应用于信息领域的其他方面,例如:网页排序中垃圾站点或页面的识别,仅考虑正边时,垃圾页面可以通过互指大大提升自身的排名。传统的 PageRank 网页排序算法无法应对该种作弊行为。为此, Kerchov 等人<sup>[20]</sup>对 PageRank 进行了改进,提出了一种 PageTrust 算法。该算法中,一个节点接收到的负边越多,对其排名越不利。同时,该算法对于恶意诋毁竞争对手的负边具有较好的鲁棒性;再比如,符号网络还被应用于女巫攻击防御<sup>[71]</sup>。女巫攻击是指恶意用户伪造多个身份在社会媒体上进行恶意内容评价、打分和投票等。假定社交网络上在正常用户与 Sybil 用户之间存在小部分不信任关系(负边)和虚假的信任关系(正边),除此外都是真实可信任的关系(正边),那么,可以通过定位所有由正边组成的、连通负边两端节点的路径集中的最小边割集,以准确定位 Sybil 区域,从而过滤掉来自该区域的大部分恶意评价或打分,实现对该类攻击的防御。

可以看到,符号网络研究的应用目前已较多。具体应用方法与前述的符号网络基础理论以及静态拓扑结构分析联系紧密。比如,个性化推荐和态度预测重点在于结构平衡理论和地位理论的使用,用户聚类和女巫攻击防御可对应到社区划分问题,用户特征分析和垃圾站点识别对应到基于静态拓扑的节点特征分析。由于信息领域中许多复杂系统,如万维网、部分在线社会网络等都具有符号属性,因而将它们抽象为符号网络分析更为精确,这也就直接吸引了越来越多的学者将符号网络应用到相应的实际问题中。

然而目前,符号网络的应用还不够成熟。这一方面是由于符号网络的研究本身还有待充实。以个性化推荐为例,虽然在一些在线社会网络的符号预测上已达到很高的精确度,而对传统链路预测准确度的提升效果则显得一般化<sup>[7]</sup>。如何进一步发现简单、准确的特征集或是其他辅助信息来进行符号网络的结构预测,特别是链路预测,还有待深入探讨;另一方面,由于应用背景各不相同,在某些上下文环境中,符号网络的构建及负边的使用等需要结合特定背景进行设计。比如在女巫攻击防御算法设计中,需要考虑在对抗型且动态变化的网络上恶意用户群的发现,而目前已有的符号网络社区划分方法仅适用于静态网络。这些都给符号网络的应用提出了新的挑战。

## 5 展 望

目前,在社会学领域,符号网络的研究已经开展 60 多年,研究成果颇丰。然而在信息领域及其他领域中,符号网络的研究仅有 10 余年,虽有一定的研究成果但不够成熟。本节我们对符号网络研究的未来方向加以展望。

符号网络的形成和演化机制是符号网络研究的基础问题,无论是对符号网络建模还是进行边及边符号的预测,都避不开此问题。但是迄今为止,成熟且被广泛认可的理论仅有结构平衡理论。对这个问题的突破,可以从对真实网络的实证分析着手。在线社会网络的涌现提供给我们许多详细的符号网络案例,分析这些真实网络的拓扑特征和演化过程,不仅能对传统社会学理论进行验证,同时还有可能发现新的规律。比如近几年有学者提出的针对有向网络的“地位理论”,这是对传统符号网络研究工作有益的补充和新视角的引入。目前,还有许多对数据更细致的分析工作值得去做,比如引入时间因素、节点与边的生灭变化来分析符号网络演化规律等。

从复杂网络角度来看,符号网络中的社区发现近些年才受到人们的关注。尽管传统社会学中对符号网络分割问题的认识已基本稳定,但分割问题与复杂网络视角的社区发现区别很大:一方面,符号网络分割并不考虑边的密度问题,并且基于全局划分的方式,这样做灵活性差,不能单独运用于局部,并且无法识别重叠社区;另一方面,也无法兼顾边密度、社区的重叠性及多尺度特性等。符号网络中社区的定义目前还未达成一个共识,并且已

有的方法中由于正、负边权重分配问题往往会引入多个参数进行控制,这使得参数选择又成为一个难题.总之,在符号网络中社区的定义、高效社区发现算法、重叠社区与多尺度社区发现上还有很多研究的空间.

目前,已有的符号网络演化模型均假定节点集与边集固定不变,从而演化中仅需考虑边上符号的改变.该类模型忽略了真实符号的一些特点,比如节点的新增和消亡、边的建立和断开等,这些均未在现有的符号网络演化模型中有所体现.再者,目前的演化机制基本上是根据结构平衡理论来进行设计的,但研究中表明,还存在其他各种因素,比如地位理论、时间因素的影响等.一项针对 eBay 网络上的正反馈和负反馈行为的研究表明,用户在某时刻收到的反馈类型很大程度是受到该用户近期收到的反馈类型的影响<sup>[72]</sup>.如何结合新发现的机制以及时间、空间等因素设计更合理的演化模型,具有重要意义.

符号网络上的动力学行为,如博弈<sup>[73]</sup>和同步<sup>[74]</sup>等相关的研究较少,但这个方向具有重要的应用价值.比如,研究信任网络上的信息扩散、合作演化等对网络舆情和舆论引导具有重要意义,而符号网络上同步行为的研究对网络拓扑优化具有指导意义.符号网络动力学行为研究的难点在于正确定位正、负边对动力学行为的影响,构建合适的动力学模型,这将是未来符号网络研究的一个热点和挑战.

## 6 结束语

在前面的章节中,我们首先介绍了符号网络的基础理论,并对符号网络领域研究的两个主要方向——静态拓扑分析、演化动力学的研究进展进行综述,之后介绍了信息领域中符号网络研究的具体应用场景和代表性工作.实际上,除上述综述的研究问题外,符号网络领域还有其他一些被关注的问题,如符号网络的构建<sup>[75]</sup>、嵌入问题<sup>[76]</sup>.总体来看,在网络科学和信息领域快速发展的背景下,符号网络的研究面临着新的机遇和挑战,以前的许多问题还有待深入研究,比如符号网络中社区发现、演化动力学、符号预测问题等.同时,如何把符号网络的研究成果应用到实际的场景和具体的复杂系统中,也是一个重要的关注点.本文的撰写希望对不同领域,特别是信息领域的人员提供有益的参考.

**致谢** 在此,感谢论文评阅专家的宝贵意见和建议,也感谢 NASC 小组(<http://www.nascgroup.org>)全体成员对本文工作的支持与建议.

## References:

- [1] Heider F. Attitudes and cognitive organization. *Journal of Psychology*, 1946,21(1):107–112. [doi: 10.1080/00223980.1946.9917275]
- [2] Ghosn F, Palmer G. The MID3 data set, 1993–2001: Procedures, coding rules, and description. *Conflict Management and Peace Science*, 2004,21(2):133–154. [doi: 10.1080/07388940490463861]
- [3] Parisien C, Anderson CH, Eliasmith C. Solving the problem of negative synaptic weights in cortical models. *Neural Computation*, 2008,20(6):1473–1494. [doi: 10.1162/neco.2008.07-06-295]
- [4] Guha R, Kumar R, Raghavan P, Tomkins A. Propagation of trust and distrust. In: *Proc. of the 13th Int'l Conf. on World Wide Web*. New York: ACM Press, 2004. 403–412. [doi: 10.1145/988672.988727]
- [5] Kunegis J, Lommatzsch A, Bauckhage C. The slashdot zoo: Mining a social network with negative edges. In: *Proc. of the 18th Int'l Conf. on World Wide Web*. New York: ACM Press, 2009. 741–750. [doi: 10.1145/1526709.1526809]
- [6] Leskovec J, Huttenlocher D, Kleinberg J. Signed networks in social media. In: *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*. New York: ACM Press, 2010. 1361–1370. [doi: 10.1145/1753326.1753532]
- [7] Leskovec J, Huttenlocher D, Kleinberg J. Predicting positive and negative links in online social networks. In: *Proc. of the 19th Int'l Conf. on World Wide Web*. New York: ACM Press, 2010. 641–650. [doi: 10.1145/1772690.1772756]
- [8] Larusso N, Bogdanov P, Singh A. Identifying communities with coherent and opposing views. In: *Proc. of the 15th Annual Graduate Student Workshop in Computing*. Santa Barbara: UCSB, 2010. 31–32. <http://gswc.cs.ucsb.edu/2010/proceedings.pdf>
- [9] Pang B, Lee L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2008,2(1-2):1–135. [doi: 10.1561/1500000011]
- [10] Cartwright D, Harary F. Structural balance: A generalization of Heider's theory. *Psychological Review*, 1956,63(5):277–293. [doi: 10.1037/h0046049]

- [11] Huang ZX, Qiu YH. A multiple-perspective approach to constructing and aggregating citation semantic link network. *Future Generation Computer Systems*, 2010,26(3):400–407. [doi: 10.1016/j.future.2009.07.006]
- [12] Srinivasan A. Local balancing influences global structure in social networks. *Proc. of the National Academy of Sciences*, 2011, 108(5):1751–1752. [doi: 10.1073/pnas.1018901108]
- [13] Davis JA. Clustering and structural balance in graphs. *Human Relations*, 1967,20(2):181–187. [doi: 10.1177/001872676702000206]
- [14] Szell M, Lambiotte R, Thurner S. Multirelational organization of large-scale social networks in an online world. *Proc. of the National Academy of Sciences*, 2010,107(3):13636–13641. [doi: 10.1073/pnas.1004008107]
- [15] Chiang KY, Natarajan N, Tewari A, Dhillon IS. Exploiting longer cycles for link prediction in signed networks. In: *Proc. of the 20th ACM Int'l Conf. on Information and Knowledge Management*. New York: ACM Press, 2011. 1157–1162. [doi: 10.1145/2063576.2063742]
- [16] Bonacich P, Lloyd P. Calculating status with negative relations. *Social Networks*, 2004,26(4):331–338. [doi: 10.1016/j.socnet.2004.08.007]
- [17] Bonacich P. Power and centrality: A family of measures. *American Journal of Sociology*, 1987,92(5):1170–1182. [doi: 10.1086/228631]
- [18] Bonacich P. Some unique properties of eigenvector centrality. *Social Networks*, 2007,29(4):555–564. [doi: 10.1016/j.socnet.2007.04.002]
- [19] Kamvar SD, Schlosser MT, Garcia-Molina H. The EigenTrust algorithm for reputation management in P2P networks. In: *Proc. of the 12th Int'l Conf. on World Wide Web*. New York: ACM Press, 2003. 640–651. [doi: 10.1145/775152.775242]
- [20] de Kerchove C, van Dooren P. The PageTrust algorithm: How to rank Web pages when negative links are allowed. In: *Proc. of the SIAM Int'l Conf. on Data Mining*. Philadelphia: SIAM, 2008. 346–352. <https://www.siam.org/proceedings/datamining/2008/dm08.php>
- [21] Harary F, Kabell JA. A simple algorithm to detect balance in signed graphs. *Mathematical Social Sciences*, 1980,1(1):131–136. [doi: 10.1016/0165-4896(80)90010-4]
- [22] Maybee JS, Maybee SJ. An algorithm for identifying Morishima and anti-Morishima matrices and balanced digraphs. *Mathematical Social Sciences*, 1983,6(1):99–103. [doi: 10.1016/0165-4896(83)90050-1]
- [23] Kunegis J, Schmidt S, Lommatzsch A, Lerner J, de Luca EW, Albayrak S. Spectral analysis of signed graphs for clustering, prediction and visualization. In: *Proc. of the SIAM Int'l Conf. on Data Mining*. Philadelphia: SIAM, 2010. 559–570. <https://www.siam.org/proceedings/datamining/2010/dm10.php>
- [24] Zachary WW. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, 1977,33(4):452–473.
- [25] Katz L. A new status index derived from sociometric analysis. *Psychometrika*, 1953,18(1):39–43. [doi: 10.1007/BF02289026]
- [26] Harary F. On the measurement of structural balance. *Behavioral Science*, 1959,4(4):316–323.
- [27] Harary F. A matrix criterion for structural balance. *Naval Research Logistics Quarterly*, 1960,7(2):195–199. [doi: 10.1002/nav.3800070208]
- [28] Facchetti G, Iacono G, Altafini C. Computing global structural balance in large-scale signed social networks. *Proc. of the National Academy of Sciences*, 2011,108(52):20953–20958. [doi: 10.1073/pnas.1109521108]
- [29] Galam S. Fragmentation versus stability in bimodal coalitions. *Physica A: Statistical Mechanics and its Applications*, 1996,230(1-2):174–188. [doi: 10.1016/0378-4371(96)00034-9]
- [30] Antal T, Krapivsky PL, Redner S. Dynamics of social balance on networks. *Physical Review E*, 2005,72(3):036121. [doi: 10.1103/PhysRevE.72.036121]
- [31] Marvel SA, Strogatz SH, Kleinberg JM. Energy landscape of social balance. *Physical Review Letters*, 2009,103(19):198701. [doi: 10.1103/PhysRevLett.103.198701]
- [32] Shen HW. *Community Structure of Complex Networks*. Berlin: Springer-Verlag, 2013. 1–17.
- [33] Doreian P, Mrvar A. A partitioning approach to structural balance. *Social Networks*, 1996,18(2):149–168. [doi: 10.1016/0378-8733(95)00259-6]
- [34] Doreian P, Batagelj V, Ferligoj A. *Generalized Blockmodeling*. Cambridge: Cambridge University Press, 2005. 295–325.
- [35] Doreian P. A multiple indicator approach to blockmodeling signed networks. *Social Networks*, 2008,30(3):247–258. [doi: 10.1016/j.socnet.2008.03.005]
- [36] Brusco M, Doreian P, Mrvar A, Steinley D. Two algorithms for relaxed structural balance partitioning: Linking theory, models and data to understand social network phenomena. *Sociological Methods and Research*, 2011,40(1):57–87. [doi: 10.1177/0049124110384947]



- [37] Mrvar A, Doreian P. Partitioning signed two-mode networks. *The Journal of Mathematical Sociology*, 2009,33(3):196–221. [doi: 10.1080/00222500902946210]
- [38] Banerjee S, Sarkar K, Gokalp S, Sen A, Davulcu H. Partitioning signed bipartite graphs for classification of individuals and organizations. In: *Proc. of the Social Computing, Behavioral—Cultural Modeling and Prediction*. LNCS, Berlin: Springer-Verlag, 2012. 196–204. [doi: 10.1007/978-3-642-29047-3\_24]
- [39] Inohara T. Quasi-Clusterability of signed graphs with negative self evaluation. *Applied Mathematics and Computation*, 2004,158(1): 201–215. [doi: 10.1016/j.amc.2003.09.005]
- [40] Inohara T. Signed graphs with negative self evaluation and clusterability of graphs. *Applied Mathematics and Computation*, 2004, 158(2):477–487. [doi: 10.1016/j.amc.2003.09.015]
- [41] Gómez S, Jensen P, Arenas A. Analysis of community structure in networks of correlated data. *Physical Review E*, 2009,80(1): 016114. [doi: 10.1103/PhysRevE.80.016114]
- [42] Traag VA, Bruggeman J. Community detection in networks with positive and negative links. *Physics Review E*, 2009,80(3): 036115. [doi: 10.1103/PhysRevE.80.036115]
- [43] Li X, Chen HC, Li S. Exploiting emotions in social interactions to detect online social communities. In: *Proc. of the Pacific Asia Conf. on Information Systems*. Atlanta: AIS, 2010. 1426–1437.
- [44] Bansal N, Blum A, Chawla S. Correlation clustering. *Machine Learning*, 2002,56(1-3):89–113.
- [45] Demaine ED, Immorlica N. Correlation clustering with partial information. In: *Proc. of the Approximation, Randomization, and Combinatorial Optimization*. LNCS, Berlin: Springer-Verlag, 2003. 1–13. [doi: 10.1007/978-3-540-45198-3\_1]
- [46] Yang B, Cheung WK, Liu JM. Community mining from signed social networks. *IEEE Trans. on Knowledge and Data Engineering*, 2007,19(10):1333–1348. [doi: 10.1109/TKDE.2007.1061]
- [47] Kong LQ, Yang ML. Improvement of clustering algorithm FEC for signed networks. *Journal of Computer Applications*, 2011,31(5): 1395–1399 (in Chinese with English abstract). [doi: 10.3724/SP.J.1087.2011.01395]
- [48] Sharma T, Charls A, Singh PK. Community mining in signed social networks—An automated approach. In: *Proc. of the Int'l Conf. on Computer Engineering and Applications*. Singapore: IACSIT Press, 2009. 152–157. <http://cpfd.cnki.com.cn/Article/CPFDTOTAL-CDYA200906001028.htm>
- [49] Wang ZG, Thorngate W. Sentiment and social mitosis: Implications of Heider's balance theory. *Journal of Artificial Societies and Social Simulation*, 2003,6(3):26–45.
- [50] Antal T, Krapivsky PL, Redner S. Social balance on networks: The dynamics of friendship and enmity. *Physica D: Nonlinear Phenomena*, 2006,224(2):130–136. [doi: 10.1016/j.physd.2006.09.028]
- [51] Kulakowski K, Gawroński P, Groniek P. The Heider balance—A continuous approach. *Int'l Journal of Modern Physics C*, 2005, 16(5):707–716. [doi: 10.1142/S012918310500742X]
- [52] Gawroński P, Groniek P, Kulakowski K. The Heider balance and social distance. *Acta Physica Polonica B*, 2005,36(8):2549–2558.
- [53] Gawroński P, Kulakowski K. Heider balance in human networks. *AIP Conf. Proc.*, 2005,779(1):93–95.
- [54] Marvel SA, Kleinberg J, Kleinberg RD, Strogatz SH. Continuous-Time model of structural balance. *Proc. of the National Academy of Sciences*, 2011,108(5):1771–1776. [doi: 10.1073/pnas.1013213108]
- [55] Gawroński P, Kulakowski K. A numerical trip to social psychology: Long-Living states of cognitive dissonance. *Computational Science*, 2007,4490:43–50. [http://link.springer.com/chapter/10.1007/978-3-540-72590-9\\_6](http://link.springer.com/chapter/10.1007/978-3-540-72590-9_6)
- [56] Radicchi F, Vilone D, Yoon S, Meyer-Ortmanns H. Social balance as a satisfiability problem of computer science. *Physical Review E*, 2007,75(2):026106. [doi: 10.1103/PhysRevE.75.026106]
- [57] Radicchi F, Vilone D, Meyer-Ortmanns H. Universality class of triad dynamics on a triangular lattice. *Physical Review E*, 2007, 75(2):021118. [doi: 10.1103/PhysRevE.75.021118]
- [58] Malekzadeh M, Fazli MA, Khalidabadi PJ, Safariy M, Rabieey HR. Social balance and signed network formation games. In: *Proc. of the 5th Int'l Workshop on Social Network Mining and Analysis*. New York: ACM Press, 2011. <http://users.cis.fiu.edu/~lzheng001/activities/KDD2011Program/workshops/SNAKDD2011/SNAKDD2011-Proceedings.pdf#page=127>
- [59] Victor P, Cornelis C, De Cock M, da Silva PP. Gradual trust and distrust in recommender systems. *Fuzzy Sets and Systems*, 2009, 160(10):1367–1382. [doi: 10.1016/j.fss.2008.11.014]
- [60] Victor P, Cornelis C, De Cock M, Teredesai AM. Trust- and distrust- based recommendations for controversial reviews. *Intelligent Systems*, 2011,26(1):48–55. [doi: 10.1109/MIS.2011.22]
- [61] Ma H, Lyu MR, King I. Learning to recommend with trust and distrust relationships. In: *Proc. of the 3rd ACM Conf. on Recommender Systems*. New York: ACM Press, 2009. 189–196. [doi: 10.1145/1639714.1639746]
- [62] Chen CC, Wan YH, Chung MC, Sun YC. An effective recommendation method for cold start new users using trust and distrust networks. *Information Sciences*, 2009,224:19–36. [doi: 10.1016/j.ins.2012.10.037]

- [63] Symeonidis P, Tiakas E, Manolopoulos Y. Transitive node similarity for link prediction in social networks with positive and negative links. In: Proc. of the 4th ACM Conf. on Recommender Systems. New York: ACM Press, 2010. 183–190. [doi: 10.1145/1864708.1864744]
- [64] Liu HF, Lim EP, Lauw HW, Le MT, Sun AX, Srivastava J, Kim YA. Predicting trusts among users of online communities: An opinions case study. In: Proc. of the 9th ACM Conf. on Electronic Commerce. New York: ACM Press, 2008. 310–319. [doi: 10.1145/1386790.1386838]
- [65] DuBois T, Golbeck J, Srinivasan A. Predicting trust and distrust in social networks. In: Proc. of the 2011 IEEE Int'l Conf. on Social Computing. Washington: IEEE Computer Society, 2011. 418–424. [doi: 10.1109/PASSAT/SocialCom.2011.56]
- [66] Zolfaghar K, Aghaie A. Mining trust and distrust relationships in social Web applications. In: Proc. of the 2010 IEEE Int'l Conf. on Intelligent Computer Communication and Processing. Washington: IEEE Computer Society, 2010. 73–80. [doi: 10.1109/ICCP.2010.5606460]
- [67] Borzymek P, Sydow M. Trust and distrust prediction in social network with combined graphical and review-based attributes. In: Proc. of the Agent and Multi-Agent Systems: Technologies and Applications. LNCS, Berlin: Springer-Verlag, 2010. 122–131. [doi: 10.1007/978-3-642-13480-7\_14]
- [68] Mani S, Cautis B, Abdesslem T. Building a signed network from interactions in Wikipedia. In: Proc. of the Databases and Social Networks. New York: ACM Press, 2011. 19–24. [doi: 10.1145/1996413.1996417]
- [69] Mishra A, Bhattacharya A. Finding the bias and prestige of nodes in networks based on trust scores. In: Proc. of the 20th Int'l Conf. on World Wide Web. New York: ACM Press, 2011. 567–576. [doi: 10.1145/1963405.1963485]
- [70] Li RH, Yu JX, Huang X, Cheng H. A framework of algorithms: Computing the bias and prestige of nodes in trust networks. PLoS ONE, 2013,7(12):e50843.
- [71] Chiluka N, Andrade N, Pouwelse J, Sips H. Leveraging trust and distrust for sybil-tolerant voting in online social media. In: Proc. of the 1st Workshop on Privacy and Security in Online Social Media. New York: ACM Press, 2012. 1–8. [doi: 10.1145/2185354.2185355]
- [72] Khopkar T, Li X, Resnick P. Self-Selection, slipping, salvaging, slacking, and stoning the impacts of negative feedback at eBay. In: Proc. of the 6th ACM Conf. on Electronic Commerce. New York: ACM Press, 2005. 223–231. [doi: 10.1145/1064009.1064033]
- [73] Inohara T. Clusterability of groups and information exchange in group decision making with approval voting system. Applied Mathematics and Computation, 2003,136(1):1–15. [doi: 10.1016/S0096-3003(02)00008-5]
- [74] Nishikawaa T, Mottera AE. Network synchronization landscape reveals compensatory structures, quantization, and the positive effect of negative interactions. Proc. of the National Academy of Sciences, 2010,107(23):10342–1034. [doi: 10.1073/pnas.0912444107]
- [75] Brandes U, Kenis P, Lerner J, van Raaij D. Network analysis of collaboration structure in Wikipedia. In: Proc. of the 18th Int'l Conf. on World Wide Web. New York: ACM Press, 2009. 731–740. [doi: 10.1145/1526709.1526808]
- [76] Kermarrec AM, Thraves C. Can everybody sit closer to their friends than their enemies. In: Proc. of the Mathematical Foundations of Computer Science. LNCS, Berlin: Springer-Verlag, 2011. 388–399. [doi: 10.1007/978-3-642-22993-0\_36]

#### 附中参考文献:

- [47] 孔令旗,杨梦龙.符号网络聚类算法 FEC 的改进.计算机应用,2011,31(5):1395–1399. [doi: 10.3724/SP.J.1087.2011.01395]



程苏琦(1989—),女,江西南城人,博士生,主要研究领域为网络拓扑分析,网络科学,信息传播.

E-mail: chengsuqi@gmail.com



张国清(1965—),男,博士,副研究员,博士生导师,CCF 高级会员,主要研究领域为网络科学,网络拓扑及应用,大数据分析 & 传输优化.

E-mail: gqzhang@ict.ac.cn



沈华伟(1982—),男,博士,副研究员,CCF 会员,主要研究领域为网络科学,社会网络分析,数据挖掘.

E-mail: shenhuawei@ict.ac.cn



程学旗(1971—),男,博士,研究员,博士生导师,CCF 高级会员,主要研究领域为网络科学,网络与信息安全,互联网搜索与服务.

E-mail: cxq@ict.ac.cn