

基于博弈论的符号网络社团发现算法

王帅辉^{1,2} 胡谷雨³ 潘雨¹ 张志越² 张海峰² 潘志松³

1 陆军工程大学研究生院 南京 210007

2 海军航空大学第三飞行训练基地 河北 秦皇岛 066000

3 陆军工程大学指挥控制工程学院 南京 210007

(wsh91876@sina.com)

摘要 社团结构作为复杂网络的中尺度特征,对于深入理解网络的结构和属性具有重要的意义。与无符号网络不同,符号网络包括正边和负边,分别代表了友好和敌对的关系。在形成社团时,节点通常会选择与朋友在同一社团内,而与对手在不同的社团。基于这种思想,构建了一种用于符号网络中社团发现的博弈论模型,设计了一种社团发现算法。实验结果表明,该算法在非重叠社团和重叠社团的识别中都具有卓越的性能。另外,对算法的运行效率进行了分析,提出了一种优化方法,有效地提高了算法的运行效率。

关键词:符号网络;社团发现;博弈论;重叠社团

中图法分类号 TP393

Community Detection in Signed Networks with Game Theory

WANG Shuai-hui^{1,2}, HU Gu-yu³, PAN Yu¹, ZHANG Zhi-yue², ZHANG Hai-feng² and PAN Zhi-song³

1 Graduate School, Army Engineering University of PLA, Nanjing 210007, China

2 The Third Flight Training Base of Naval Aeronautical University, Qinhuangdao, Hebei 066000, China

3 Command and Control Engineering College, Army Engineering University of PLA, Nanjing 210007, China

Abstract As a meso-scale feature of complex networks, community structure is of great significance for understanding the structure and property of networks. Unlike unsigned networks, signed networks include positive and negative edges, which represent friendly and hostile relations, respectively. When forming a community, a node usually chooses to be in the same community with friends, but in different communities with enemies. Based on this idea, a game theory model for community detection in signed networks is constructed, and a related algorithm is designed. Experimental results show that the algorithm performs well in identifying non-overlapping and overlapping communities. In addition, the efficiency of the algorithm has been verified, and an optimization method, which can effectively improve the efficiency of the proposed algorithm, is proposed.

Keywords Signed networks, Community detection, Game theory, Overlapping community

1 引言

符号网络指包含正边和负边的网络。正边表示积极、信任或朋友关系,负边代表了消极、不信任或敌对关系。现实世界中的很多系统都能抽象为符号网络,如微信和 Twitter 等社交媒体中用户之间的关系网络等。因此,近年来,符号网络受到了越来越多的关注。

社团结构是复杂网络中节点集合内部连接稠密,而集合之间连接稀疏的一种重要的中尺度特征。社团发现是复杂网络分析的一个基本任务,有助于其他社会计算任务的实现,并被应用于许多实际问题的求解^[1]。自从 Newman 将现代数据分析技术引入社团发现领域以来^[2-3],大量的基于无符号网

络的社团发现方法和技术被提出来。在无符号网络的社团发现中,连接的强度是主要的考虑因素。而在符号网络中,不仅要考虑连接的强度,还要考虑负边的影响,认为社团内为正边、社团间为负边的结构平衡理论^[4]在符号网络的社团发现中发挥着重要的作用。因此,针对无符号网络的社团发现算法不能直接应用于符号网络中,需要针对符号网络的特点,研究适用于符号网络的社团发现方法。

在形成社团时,符号网络中的节点通常是理性的,为追求自身利益最大化,它们会选择与朋友(正边)在一个社团,而与敌人(负边)在不同的社团。这与博弈论中参与者理性、自私的特征相吻合。因此,本文将符号网络中的社团发现问题建模为博弈论模型,通过达到博弈的均衡状态得到社团划分结果。

基金项目:国家重点研发计划(2017YFB0802800);国家自然科学基金(61473149)

This work was supported by the National Key Research Development Program of China (2017YFB0802800) and National Natural Science Foundation of China (61473149).

通信作者:潘志松(hotpz@sina.com)

2 相关工作

总结现有文献中的方法,关于符号网络的社团发现问题,主要有3种研究思路:1)将社团发现问题转化为目标优化问题进行求解。Zhao等^[5]使用统计推断的方法实现了对符号网络的社团发现,建立了符号网络的概率模型,利用参数估计的期望最大化算法计算社团结构。Li等^[6]提出基于改进的非负矩阵分解(NMF)的符号网络社团发现算法,该算法将符号网络映射到低维空间,然后在新的空间中进行社团发现。这种方法的性能非常卓越,因此将其作为基准算法与本文方法进行对比,以验证本文方法的性能。Girdhar和Bharadwaj^[7]综合考虑网络中连接的强度和连接的符号,利用模块度、frustration和社会影响力因子等指标构建了一种基于多目标优化的符号社交网络社团发现方法。2)将符号网络的社团发现问题进行分段处理,首先对正边网络采用传统的无符号网络的社团发现方法,然后利用符号网络的负边信息构建评价函数,对社团发现结果进行调整和修正。Yang等^[8]将FEC算法拓展应用到符号网络中,首先随机地从一个还未标识其社团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合的相应概率,在游走过程中忽略掉所有负边,然后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社区。FEC算法的时间复杂度低,但是初始种子节点的选择、随机游走步长的设定会造成性能不稳定。3)对无符号网络社团发现方法进行改造,建立适用于符号网络的社团发现方法。Gomez等^[9]在传统模块度的基础上,提出了符号模块度的概念。Anchuri等^[10]设计了基于符号模块度优化的社团发现方法。Kunegis等^[11]对谱聚类方法进行扩展,提出了符号谱聚类的社团发现方法。

对于博弈理论在社团发现问题中的应用,已有学者进行了相关研究。Chen等^[12]首次将博弈论引入到社团发现问题中,建立社团生成博弈框架,为社团发现问题提供了新思路。随后,Alvari等^[13]在此基础上,通过改进增益函数和决策规则,改善了社团发现的结果。Lung等^[14]利用非合作博弈的纳什均衡建模了社交网络中的社团发现问题。Chopade等^[15]在Lung等工作的基础上,修改了模块度和拟合函数,构建了可用于大规模复杂网络的社团发现方法。然而,这些社团发现方法都是基于无符号网络的,不能直接用于解决符号网络中的社团发现问题。

3 社团生成博弈模型

3.1 符号定义

为了便于表述,表1列出了下文将用到的符号,并给出了这些符号的定义。在复杂网络研究中,通常记网络为 $G=(V,E)$, $m=|E|$ 为网络的总边数, $n=|V|$ 为网络的节点数。尽管本文方法能够应用于有向、加权网络中,但本文只对无向、无权网络进行研究。所有可能的社团集合记为 $[k]=\{1,2,\dots,k\}$ 。定义 $S=(S_1,S_2,\dots,S_n)$ 为全局策略空间,是所有节点社团标签的矢量。 S_i 为节点 i 的策略空间,为所有可能社团集合的子集,即 $S_i\subseteq[k]$ 。也就是说, S_i 是节点 i 的社团标签。 u_i , g_i 和 l_i 分别为节点 i 的效用函数、增益函数和损失函数,满足关系 $u_i=g_i-l_i$ 。

表1 符号定义

Table 1 Definition of symbols

符号	定义
G	无向无权图
V,E	节点集合,边集合
m,n	边数和节点数
A	邻接矩阵
a_{ij}	邻接矩阵中的元素
S	全局策略空间
S_i	节点 i 的策略空间
g_i	节点 i 的增益函数
l_i	节点 i 的损失函数
u_i	节点 i 的效用函数

3.2 数学模型

直观来讲,符号网络中的节点作为博弈的参与者,在形成社团时力求达到社团内的朋友(正边)最多、敌人(负边)最少,并使社团间的敌人(负边)最多、朋友(正边)最少。基于这一思想,我们定义节点 i 的增益函数为:

$$g_i = \frac{1}{m} \sum_{j \in V, j \neq i} [a_{ij} * \delta(i, j) - a_{ij} * (1 - \delta(i, j))] \quad (1)$$

其中,如果 i 与 j 在同一社团,则 $\delta(i, j) = 1$,否则 $\delta(i, j) = 0$ 。 a_{ij} 取值为1,0和-1,当 i 与 j 之间是正边时, $a_{ij} = 1$;为负边时, $a_{ij} = -1$;若 i 与 j 无连接时, $a_{ij} = 0$ 。式(1)各项的具体物理含义是,前一项表示社团内的正边数与负边数之差,后一项(含负号)表示社团间负边数与正边数之差。

借鉴chen等^[12]的方法,定义损失函数为:

$$l_i = \frac{1}{m} (|s_i| - 1) \quad (2)$$

其中, $|s_i|$ 表示节点 i 归属的社团数。那么,本文所述博弈的效用函数可表示为:

$$\begin{aligned} u_i &= g_i - l_i \\ &= \frac{1}{m} \left(\sum_{j \in V, j \neq i} [a_{ij} * (2\delta(i, j) - 1)] - (|s_i| - 1) \right) \end{aligned} \quad (3)$$

4 算法设计与分析

根据上文构建的社团生成博弈模型,我们设计了一种社团发现算法,称为边最大化博弈算法(简记为EM-Game),可用于符号网络中非重叠社团和重叠社团的划分。该算法在初始阶段将每个节点作为独立的社团,然后对随机选取的节点,基于式(3)的效用函数从加入、离开、切换操作中选择最优的决策。对所有的节点重复这个操作,直到最终达到局部纳什均衡状态。详细的算法流程如算法1所示。

算法1 EM-Game

输入:网络 G 的邻接矩阵

输出:最终的社团划分结果 C_1, C_2, \dots, C_k

1. 初始化每个节点为独立的社团
2. 计算每个节点的效用函数值
3. While not convergent do
4. 随机选取一个节点 i
5. 得到节点 i 分别执行加入、离开和切换操作的最佳策略
6. 根据效用函数值选择节点 i 的最佳操作
7. 对比节点 i 决策改变后的效用值 u_i' 与当前效用值 u_i
8. if $u_i' > u_i$ then
9. $u_i \leftarrow u_i'$
10. 更新节点 i 的社团划分
11. else 无动作
12. end if
13. end while

在算法1中,节点 i 可以选择加入新社团、离开当前所在社团和从一个社团切换到另一社团这3种操作,并且在每个操作上的策略空间是有限的。因此,计算节点执行每个操作时的所有决策是可行的,然后选取每个操作中具有最大效用函数值的决策作为该操作的最佳策略。详细的算法流程如算法2所示。受篇幅所限,本文只在算法2中列出了切换操作的详细步骤,加入和离开操作与算法2相似,未在文中列出。值得注意的是,我们将节点 i 与其邻居节点的效用值总和作为最佳策略的选取标准,而没有使用节点 i 的单独效用值。这与实际情况中个人利益与集体利益的相互依存关系是一致的。在博弈过程中,个体在追求自身利益最大化的同时会考虑其他个体的策略,尤其是与自己相邻或相关的个体的策略对自身利益的影响。因此,将综合效用值作为策略选取标准,会减少节点策略的重复执行,从而降低计算量,提高社团划分效率。

算法2 切换社团操作

输入:随机选取一个节点 i

输出:最佳切换社团策略及其对应的综合效用函数值

1. 计算当前综合效用函数值 u_{switch}
2. 将与节点 i 在不同社团的邻居节点的社团作为待加入社团集 CV1
3. 将节点 i 所在社团作为待离开社团集 CV2
4. For j in CV1 do
5. For k in CV2 do
6. 对节点 i 执行加入社团 j 操作
7. 对节点 i 执行离开社团 k 操作
8. 计算综合效用函数值 u'_{switch}
9. if $u'_{\text{switch}} > u_{\text{switch}}$ then
10. $u_{\text{switch}} \leftarrow u'_{\text{switch}}$
11. 存储该策略
12. end if
13. end for
14. end for

5 算法性能分析与优化

为验证本文算法的性能,我们在人工符号网络数据集上进行了大量的实验。本文采用的符号网络生成器是在 Lancichinetti-Fortunato-Radicchi(LFR)人工合成网络生成器^[16]的基础上引入 p_+ 和 p_- 两个参数,分别用于调节社团间正边和社团内负边出现的概率。由该生成器生成的人工合成符号网络记为 $SLFR(N, k, \max k, \min c, \max c, t_1, t_2, on, om, \mu, p_+, p_-)$ 。其中, N 表示节点数, k 表示节点平均度, $\max k$ 表示节点最大度, $\min c$ 和 $\max c$ 分别表示社团的最小和最大成员数量, t_1 和 t_2 分别代表节点度分布和社团大小分布的负指数, on 和 om 分别表示重叠节点的数目和重叠节点归属社团的个数。 μ 是社团结构指示参数,其值越大,社团结构越模糊。

为对比验证本文算法的性能,我们从3种符号网络社团发现的研究思路中各选取了一种,分别是基于约束的 semi-NMTF(Res-NMTF)算法^[6]、FEC^[8]和符号谱聚类(SSC)^[11]算法。采用的性能评价指标包括符号模块度(Q_{sn})^[9]、标准化互信息(NMI)^[17]和基于重叠社团的标准化互信息(NMI_{max})^[18]。

5.1 在非重叠社团发现中的性能

为测试本文算法在非重叠社团检测中的性能,我们使用

上文提到的符号网络生成器产生了一组人工符号网络,其参数设置如表2所列。其中,社团结构指示参数 μ 间隔 0.1 分别取 5 个值, p_+ 和 p_- 间隔 0.1 分别取 3 个值。然后,分别使用 EM-Game 和其他 3 种算法对这些网络进行社团检测,得到如图 1 和图 2 所示的算法性能对比结果。

表 2 人工合成符号网络参数信息

Table 2 Details of synthetic signed networks

N	k	$\max k$	$\min c$	$\max c$	t_1	t_2	μ	p_+	p_-
500	15	50	20	50	2	1	0.1~0.5	0~0.2	0~0.2

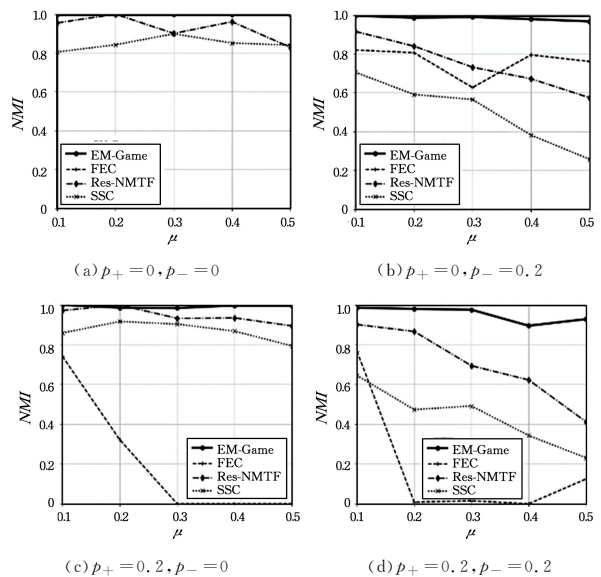


图 1 基于 NMI 指标的几种社团发现算法的性能对比

Fig. 1 Comparison between Em-Game and existing algorithms on metrics of NMI

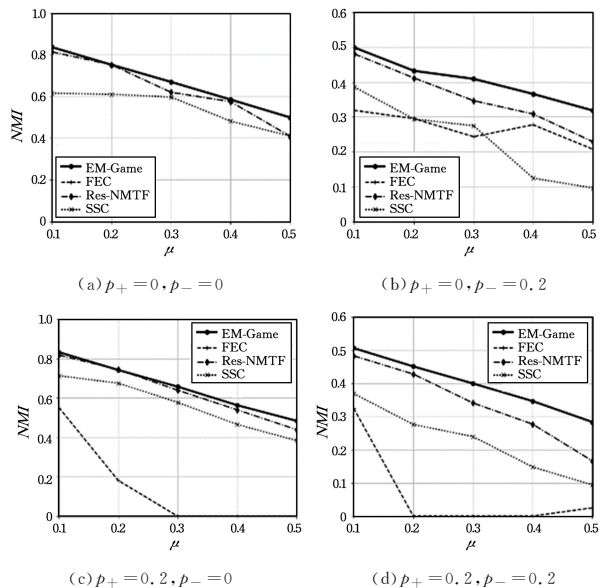


图 2 基于 Q_{sn} 指标的几种社团发现算法的性能对比

Fig. 2 Comparison between Em-Game and existing algorithms on metrics of Q_{sn}

由图 1、图 2 可见,本文算法在所有情况下的 NMI 和 Q_{sn} 指标上都优于其他 3 种算法。随着 μ 值的增大,社团结构逐渐模糊,EM-Game 算法的性能也随之减弱,在网络中存在噪声时表现得更加明显,尤其是在 Q_{sn} 指标上算法性能呈现近似线性的减弱。其他算法也存在这种现象,且在网络中存在噪

声时,性能减弱的程度较本文算法更加明显。

5.2 算法优化与分析

由上文可见,本文算法在符号网络的非重叠社团发现中具有较好的性能,能够准确地进行非重叠社团划分。但是,在实验中发现,本文算法的运行时间较长,计算效率较低。对此,我们进行了详细的实验,分析了影响算法效率的原因,提出了一种优化方案。本实验的实验平台搭建在 Ubuntu 16.04 版本操作系统上,硬件配置为 Intel(R) Xeon(R) CPU E5-2630,16GB 内存。

分析算法 2 发现,在切换操作中,节点的待加入社团数目是与其不在同一社团的邻居节点所归属的社团数目,当节点数目较多时,该待加入社团数将会很大,是影响算法运行效率的主要因素。因此,我们对此进行优化,将上述待加入社团中具有最多节点的社团作为最终的待加入社团,这样将大大减小计算量,同时这也与实际网络中个体的从众效应一致。对比优化前后的算法运行时间得到如图 3 所示的结果。图 3 的结果来自于对参数为 $\mu=0.1, p_+=0, p_-=0$ 的网络的分析结果。从图中可以发现,优化后算法的运行效率得到了显著的改善,尤其是随着节点数目的增加,算法运行速度提高得愈加明显。另外,对比算法优化前后的社团划分准确度指标发现,优化后的算法在运行效率提升的同时获得了与优化前几乎相当的准确度(见图 4(a))。

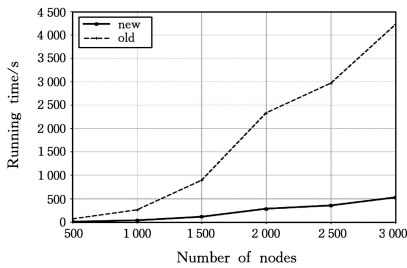
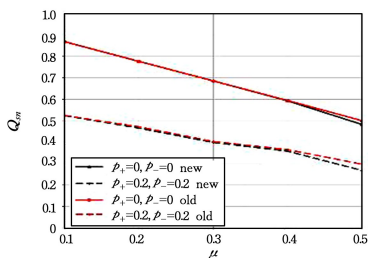
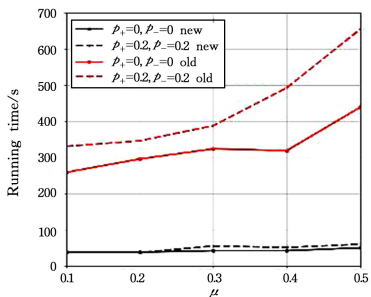


图 3 算法优化前后运行时间对比结果

Fig. 3 Running time comparison before and after algorithm optimization



(a) Q_m 性能指标



(b) 运行时间

图 4 在不同噪声情况下算法性能随 μ 的变化

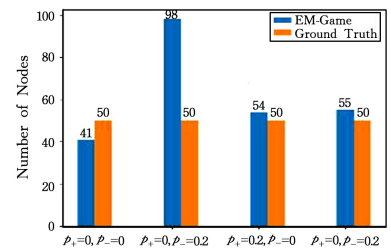
Fig. 4 Performance of algorithm varies with μ under different noise levels

图 4 描述了节点数为 1000 的网络在不同噪声情况下社团划分性能随 μ 的变化情况。由图可见,优化后的算法仅在社团结构模糊后(如图 4(a)所示,当 $\mu=0.5$ 时的结果)或噪声增大后稍有下降(如图 4(a)所示,当 $p_+=0.2, p_-=0.2$ 时的结果),但性能减弱的幅度很小,与算法优化后带来的运算速度的提升相比显得微不足道(如图 4(b)所示)。图 4(b)分析了在不同噪声情况下本文算法优化前后的运行时间随社团结构指标 μ 的变化情况。由图可见,优化前算法的运行时间随着 μ 的增大呈非常明显的增加趋势,且随着社团内和社团间噪声的增加而变大。与之相比可以发现,优化后算法的运行时间随社团结构模糊程度的增加而稍有增大,但变化不是很明显,且网络中的噪声对优化后算法的运行效率的影响也相对减弱。

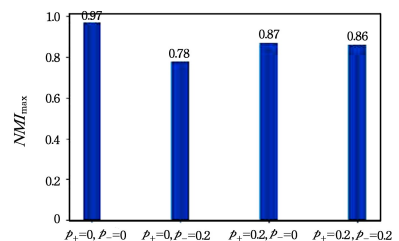
5.3 在重叠社团发现中的性能

从式(3)可以看出,本文的社团发现模型从本质上讲是针对重叠社团的,对于非重叠的社团发现只是该算法的一种特例。因此,我们对本文算法在重叠社团发现中的性能进行了实验,进一步验证了本文算法在符号网络中进行社团发现的卓越性能。为验证算法性能,我们引用适用于重叠社团的扩展 NMI 指标^[18],即 NMI_{max} 作为算法的性能指标。

利用上文所述生成人工符号网络的方法,设置 on 和 om 参数分别为 50 和 2,其他参数采用与表 2 相同的参数生成节点数为 500 的包含重叠社团的人工符号网络。对 $\mu=0.1$ 的网络采用 EM-Game 算法进行重叠社团发现,得到如图 5 所示的社团发现结果。图 5(a)给出了算法得到的重叠社团中节点的个数与真实值的对比情况。由图 5(a)可见,算法较准确地识别了属于重叠社团的节点。但是,可以明显地看到,当 $p_-=0.2$,即社团内负边出现的概率为 0.2 时,算法在重叠区域内识别出了几乎是真实情况两倍的节点数。这是因为当社团内出现负边时,节点通过加入更多的社团来达到效用的最大化。另外,我们发现,与社团间正边噪声相比,本文算法的性能对社团内的负边噪声更加敏感。尽管网络中噪声对本文算法的性能有一定影响,但是通过图 5(b)可以看到,本文算法在 NMI_{max} 指标上几乎都达到了 0.8 以上,具有良好的重叠社团发现性能。



(a) 重叠社团的节点个数



(b) NMI_{max} 性能指标

图 5 EM-Game 算法在人工网络上的重叠社团发现性能

Fig. 5 Performance of EM-Game algorithm on synthetic networks

结束语 本文提出了一种基于博弈论的用于符号网络中社团发现的方法,该方法既可以进行非重叠社团的发现,也能在不改变模型及任何参数的情况下用于重叠社团的识别。我们通过大量的实验验证了算法的性能。另外,我们还对算法的运行效率进行了分析,对算法进行了优化,提高了算法的运行速度,并在实验中验证了算法效率的提升。尽管本文算法在解决符号网络中的社团发现问题时具有良好的性能,但是我们也可以看到,在处理 3 000 个节点的网络时,即便是优化后的算法也需要长达 500 s 的处理时间。由此可见,本文算法最大的问题在于难以应用于大规模的网络。因此,下一步可以对算法的运行效率进行进一步优化,使其能够用于大规模社交网络数据的挖掘与处理。

参考文献

- [1] TANG L, LIU H. Community detection and mining in social media [J]. *Synthesis Lectures on Data Mining and Knowledge Discovery*, 2010, 2(1): 1-137.
- [2] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks[J]. *Proceedings of the National Academy of Sciences*, 2002, 99(12): 7821-7826.
- [3] NEWMAN M E J, GIRVAN M. Finding and evaluating community structure in networks[J]. *Physical Review E*, 2004, 69(2): 026113.
- [4] HEIDER F. Attitudes and cognitive organization[J]. *The Journal of Psychology*, 1946, 21(1): 107-112.
- [5] ZHAO X, YANG B, LIU X, et al. Statistical inference for community detection in signed networks [J]. *Physical Review E*, 2017, 95(4-1): 042313.
- [6] LI Z, CHEN J, FU Y, et al. Community Detection Based on Regularized Semi-Nonnegative Matrix Tri-Factorization in Signed Networks[J]. *Mobile Networks & Applications*, 2017(2): 1-9.
- [7] GIRDHAR N, BHARADWAJ K K. Community Detection in Signed Social Networks Using Multiobjective Genetic Algorithm [J]. *Journal of the Association for Information Science and Technology*, 2019, 70(8).
- [8] YANG B, CHEUNG W, LIU J. Community Mining from Signed Social Networks[J]. *IEEE Transactions on Knowledge & Data Engineering*, 2007, 19(10): 1333-1348.
- [9] GÓMEZ S, JENSEN P, ARENAS A. Analysis of community structure in networks of correlated data[J]. *Physical Review E*, 2009, 80(1): 016114.
- [10] ANCHURI P, MAGDONISMAIL M. Communities and Balance in Signed Networks: A Spectral Approach [C]// *IEEE/ACM International Conference on Advances in Social Networks Analysis & Mining*. ACM, 2012: 235-242.

ternational Conference on Advances in Social Networks Analysis & Mining. ACM, 2012: 235-242.

- [11] KUNEGIS J, SCHMIDT S, LOMMATZSCH A, et al. Spectral Analysis of Signed Graphs for Clustering, Prediction and Visualization [C] // *Siam International Conference on Data Mining*. 2010: 559-570.
- [12] CHEN W, LIU Z, SUN X, et al. A game-theoretic framework to identify overlapping communities in social networks[J]. *Data Mining & Knowledge Discovery*, 2010, 21(2): 224-240.
- [13] ALVARI H, HASHEMI S, HAMZEH A. Discovering overlapping communities in social networks: A novel game-theoretic approach[M]. IOS Press, 2013.
- [14] LUNG R I, GOG A, CHIRA C. A Game Theoretic Approach to Community Detection in Social Networks[M]. *Nature Inspired-Cooperative Strategies for Optimization (NICSO 2011)*. Springer Berlin Heidelberg, 2011: 121-131.
- [15] CHOPADEV P, ZHAN J. A Framework for Community Detection in Large Networks Using Game-Theoretic Modeling [J]. *IEEE Transactions on Big Data*, 2017, PP(99): 1-1.
- [16] LANCICHINETTI A, FORTUNATO S. Community detection algorithms: a comparative analysis[J]. *Phys Rev E Stat Nonlin Soft Matter Phys*, 2009, 80(2): 056117.
- [17] DANON L, DIAZGUILERA A, DUCH J, et al. Comparing community structure identification[J]. *Journal of Statistical Mechanics: Theory and Experiment*, 2005, 2005(9).
- [18] MCDAID A F, GREENE D, HURLEY N. Normalized mutual information to evaluate overlapping community finding algorithms[J]. *arXiv: 1110. 2515*, 2011.



WANG Shuai-hui, born in 1984, Ph.D candidate. His main research interests include graph data mining and graph neural networks.



PAN Zhi-song, born in 1973, Ph.D, professor. His main research interests include pattern recognition, machine learning and neural networks.