

文章编号:2095-6134(2019)03-0425-08

简报

社交网络关键节点检测的积极效应问题\*

王新栋<sup>1</sup>, 于 华<sup>1</sup>, 江 成<sup>2†</sup>

(1 中国科学院大学工程科学学院, 北京 100049; 2 首都经济贸易大学信息学院, 北京 100070)  
(2017 年 12 月 29 日收稿; 2018 年 4 月 18 日收修改稿)

Wang X D, Yu H, Jiang C. Positive effect of key player detection in social networks[J]. Journal of University of Chinese Academy of Sciences, 2019, 36(3): 425-432.

**摘 要** 在社交网络中, 识别有影响力的关键节点对于调控网络至关重要, 是网络科学最前沿热点的研究内容。然而, 现有方法大多基于局部特征进行求解, 缺乏对网络整体结构的建模。为有效地解决这个问题, 针对社交网络关键节点检测积极效应问题 KPP-POS (key player problem positive), 在 KPP-POS 的检测指标  $D_R$  的基础上, 建立关键节点积极效应模型的 0-1 整数线性规划模型 (0-1 integer linear programming key players problem positive effects model, IP-KPP-POS), 进而提出一种计算复杂度较低且精确度较高的局部搜索启发式算法。最后通过多种人造网络和真实网络的实验分析, 验证 IP-KPP-POS 模型在解决社交网络关键节点检测积极效应问题上的正确性和有效性。

**关键词** 社交网络; 0-1 整数线性规划; 关键节点; 网络优化; 启发式算法

**中图分类号:** TP399      **文献标志码:** A      **doi:** 10. 7523/j. issn. 2095-6134. 2019. 03. 017

Positive effect of key player detection in social networks

WANG Xindong<sup>1</sup>, YU Hua<sup>1</sup>, JIANG Cheng<sup>2</sup>

(1 School of Engineering Science, University of Chinese Academy of Sciences, Beijing 100049, China;  
2 School of Information, Capital University of Economics and Business, Beijing 100070, China)

**Abstract** Identifying influential nodes has been one of the most intensive studies among network analysis, and it is essential to control social networks. However, most of the existing methods are based on local features and lack the modeling of the overall network structure. In order to solve the key player problem positive (KPP-POS) problem effectively, we propose a 0-1 integer linear programming model (IP-KPP-POS) based on the detection standard  $D_R$  of KPP-POS. Then, we design a local search heuristic algorithm that significantly reduces the computational complexity and simultaneously achieves high accuracy. Finally, the effectiveness of our methods are validated by experiments with various synthetic networks and real-world networks.

**Keywords** social networks; 0-1 integer linear programming; critical nodes; network optimization; heuristic algorithm

\* 国家自然科学基金(71450009)和首都经济贸易大学 2018 年度科研基金资助

† 通信作者, E-mail: jiangcheng@cueb.edu.cn

随着科学技术的进步和社会物质的富足,人类社会的分工合作趋于精细化和网络化。人们处于各种形形色色的复杂网络中,比如,与生活密切相关的社会网络<sup>[1]</sup>和铁路运输网<sup>[2]</sup>,威胁人类健康的流行病疾病传播网络<sup>[3]</sup>,以及威胁人类社会安定的恐怖组织网络<sup>[4]</sup>等等。现实网络中的节点地位存在明显的差异,关键节点相对于其他节点能更大程度地影响网络的结构和功能<sup>[5-6]</sup>。如何正确评价个人在社交网络中的地位、分析相互作用影响和关联关系,以及群体行为的效应等问题已经成为复杂网络研究中一项具有重要意义的课题。特别地,随着大数据时代的来临,关键节点的检测问题研究已经成为研究热点。其中复杂网络上节点积极效应研究涉及动力学、拓扑学、运筹学等多个学科的交叉应用,应用情景广泛,如社交网络中的谣言传播<sup>[7]</sup>、通讯网络中的病毒传播<sup>[8]</sup>、电力网络中的相继故障<sup>[9]</sup>,以及经济网络中的危机扩散<sup>[10]</sup>,等等。

在国内,众多学者们从不同的实际问题出发设计出各种各样的方法。孙睿和罗万伯<sup>[11]</sup>对网络舆情中节点重要性评估方法的研究现状进行综述,并从基于节点属性和网络拓扑结构两个方面,总结节点重要性的评价模型和方法。Hu 等<sup>[12]</sup>将 K 壳分解法与社区结构相结合,提出一种改进指标,并在流行病传染模型领域上实证分析,验证该方法得到的关键传染源,相比 K 壳分解方法得到的传染源,具有更强的传播能力。赵之滢等<sup>[13]</sup>在复杂网络理论、动力系统理论与控制理论结合的基础上,开展复杂动态网络同步与控制理论的研究。利用网络中节点所关联的社团数目衡量节点的传播能力,提出一种基于网络社团结构的节点影响力度量方法,通过在 Facebook 社区、邮件通讯和博客平台等多个互联网领域数据集上实验评测,证实此方法能更有效地检测出互联网领域内信息传播能力较强的关键人物。

在国外,多个领域的学者对复杂网络关键点挖掘的相关理论基础进行了探索性研究。Borgatti<sup>[14]</sup>率先提出 KPP-POS (key player problem positive) 的概念,即寻找一组关键节点用作信息扩散的种子可以最快速让信息在整个网络中扩散。Hussain<sup>[15]</sup>提出适应一般性系统的关于 KPP 问题的条件概率和熵度量的统计方法。为了更有效地找到最佳解决方案,Hamill 等<sup>[16]</sup>提出基于对关键节点选择的有条件约束的线性规划问题。在

线性规划问题基础上,McGuire 和 Deckro<sup>[17]</sup>进一步扩展到加权 WKPP-Pos 以便考虑节点的权重和边缘的关系权重。Yang<sup>[18]</sup>通过考虑网络的结构及其节点和边缘的属性,提出一个广义的关键节点问题 GKPP (generalized key player problem)。

综上所述,利用数学规划的模型和思想研究网络关键节点检测问题属于学术前沿,相关理论和方法的研究也鲜有见闻。因此,本文基于网络的整体结构,考虑关键节点对网络积极效应的影响,采用数学规划的方法建立更加精确的模型并设计高效的求解算法,通过对这部分问题进行深入研究,完善关键节点检测理论和方法。

本文的主要贡献如下:

1) 基于对 KPP-POS 检测指标  $D_R$  的优化,建立 0-1 整数线性规划的关键节点积极效应模型 (0-1 integer linear programming key players problem positive effects model, IP-KPP-POS),用来解决社交网络关键节点检测积极效应的问题,并对该模型进行精确求解。

2) 设计局部搜索启发式算法对 IP-KPP-POS 模型进行求解,以适应不同规模网络的关键节点检测问题。并通过在特定随机图、多种类型随机网络以及真实标准网络数据集上的综合实验分析,验证本文所提出的模型在多个性能指标上的明显优势。

## 1 相关工作

复杂网络关键节点的选择标准是多样的,有时需要初始免疫的节点在传染病扩散时能最大限度地保护全体人群,而有时需要破坏节点造成最广泛的连锁故障等等。因此,找到一个最佳量化节点在每种情况下的重要性的通用指标是不可能的<sup>[19]</sup>。Borgatti<sup>[14]</sup>最早给出社交网络关键节点消极效应和积极效应的正式定义,并提出  $D_F$  指标和  $D_R$  指标分别用来衡量网络的离散程度和网络的聚合程度。 $D_F$  和  $D_R$  指标具体表示如下:

$$D_F = 1 - \frac{2 \sum_{i>j} \frac{1}{d_{ij}}}{n(n-1)}, \quad D_R = \frac{\sum_j \frac{1}{d_{kj}}}{n-k}. \quad (1)$$

$D_F$  指标很好地刻画了网络的离散程度,但直接对  $D_F$  指标进行优化存在计算上的不可行性。为解决当前所面临的问题,Arulselvan 等<sup>[20]</sup>针对社交网络关键节点消极效应问题建立 0-1 整数线性规划模型并设计启发式算法进行求解,取得了

较好的实验效果。然而,该模型只考虑连通分支内节点的个数,而忽视了连通分支内部的结构,不能有效优化连通分支的内部结构。为了更好地优化连通分支的内部结构,Jiang 等<sup>[21]</sup>对  $D_F$  指标进行近似,综合考虑剩余网络的一步连通性和两步连通性,开创性地建立 0-1 二次约束二次规划模型并设计算法进行求解,填补了将  $D_F$  指标融入数学规划模型研究的空白。

然而到目前为止,关于如何将积极效应指标  $D_R$  融入数学规划的模型中,采用系统的方法检测关键节点研究的更是接近空白。本文在现有研究的基础上,基于对 KPP-POS 检测指标  $D_R$  的优化,针对社交网络关键节点检测积极效应问题,建立 0-1 整数线性规划的关键节点积极效应模型,并设计一种计算复杂度较低且精确度较高的局部搜索启发式算法对网络的关键节点进行识别。

本文所研究的社交网络关键节点检测积极效应问题(KPP-POS)具体是指:给定无向图  $G(V, E)$  和正整数  $K$ ,从节点集合  $V$  中选择  $K$  个节点子集,使其关联节点的总数最多,从而达到最大传播效应的目的。这个问题的本质是计算一组节点成员( $KP$  集)与另一组节点成员(网络的其余部分)之间的连接(内聚)强度。函数指标  $D_R$  就很好地描述了集合  $K$  的成员与网络的剩余部分( $V-K$ )之间的内聚程度。

其中,  $d_{Kj}$  定义为从集合  $K$  到剩余部分节点  $j$  的最小距离,分子表示为从  $KP$  集到所有剩余节点的最小距离的倒数之和,分母功能是标准化度量指标,即  $D_R \in [0, 1]$ 。 $D_R$  可看作是  $KP$  集合节点达到其余所有节点的加权比例,并且到达其他节点的权值和它们最小距离成反比。当且仅当距离为 1 时,节点被赋予最大权值 1。因此,当每个外部节点与  $KP$  集合的至少一个成员相邻(即  $KP$  集合是支配集合)时,  $D_R$  取最大值 1。当  $KP$  集合的任何成员都不与外部节点集的成员相邻,即  $KP$  集完全离散时,得到最小值 0。

## 2 IP-KPP-POS 模型建立

公式(1)给出衡量网络聚合程度的指标  $D_R$ ,虽然  $D_R$  考虑了距离加权,但是  $D_R$  指标包含最短路径  $d_{Kj}$  的求解,直接把  $D_R$  作为目标进行优化复杂度高,求解困难。另外,对于一步可达路径  $d_{ij} = 1, 1/d_{ij} = 1$ ;对于两步可达路径  $d_{ij} = 2, 1/d_{ij} = 1/2$ ;

$d_{ij} \geq 3, 1/d_{ij}$  相应减少为  $1/3, 1/4, \dots, 0$ 。在资源有限或成本约束条件下,可以假设两步和两步以上可达路径在实际中是不连通的,在此假设条件下,对  $D_R$  的优化就简化为对一步可达路径节点对的连通性能最大化优化问题。

基于对 KPP-POS 的检测指标  $D_R$  的优化,建立考虑网络内部距离加权的深入优化的 0-1 整数线性规划的关键节点积极效应模型(0-1 Integer linear programming key players problem positive effects model, IP-KPP-POS),具体模型表示如下。

$$S_{\text{KPP-POS}} = \{n \mid y_n = 1\}. \tag{2}$$

$$\max \sum_i \sum_j a_{ij} \cdot x_{ij}. \tag{3}$$

$$\text{s.t.} \quad \sum_{j=1}^N x_{ij} = 1, \forall i \in V. \tag{4}$$

$$\sum_{j=1}^N y_j = K. \tag{5}$$

$$x_{ij} \leq y_j, \forall i, j \in V. \tag{6}$$

$$x_{ij}, y_j \in \{0, 1\}, \forall i, j \in V. \tag{7}$$

$$a_{ij} \in \{0, 1\}, \forall i, j \in V. \tag{8}$$

模型变量解释如表 1 所示。

表 1 模型变量解释  
Table 1 Explanations of model variables

变量名	变量解释
$i$	网络的其余部分节点
$j$	网络的关键节点
$N$	网络节点
$V$	网络所有节点集合
$K$	关键节点个数
$a_{ij}$	节点 $i$ 和 $j$ 之间是否有边相连,相连则等于 1,否则为 0
$x_{ij}$	如果节点 $i$ 指派给节点 $j$ (即节点 $j$ 对节点 $i$ 有影响),则等于 1,否则为 0
$y_j$	如果节点 $j$ 被选为关键节点,则等于 1,否则为 0

约束(4)迫使每个网络的其余部分节点最多只受一个关键节点影响;约束(5)在资源有限条件下限制网络关键节点的总数量;约束(6)允许只有当前位置为关键节点时,点  $i$  才能受点  $j$  影响。

为快速求解,本文采用 MATLAB 中内置的 Integer Linear Programming 算法包对 IP-KPP-POS 模型求解;已有文章证明,网络关键节点检测问题是 NP-complete 的<sup>[20]</sup>。因此,为适应大规模网络的求解,本文在第 3 节中设计启发式算法进行求解。

3 启发式算法设计

3.1 贪婪算法设计

最近,已有学者提出针对社交网络关键节点检测积极效应问题求解的 KPP-POS 贪婪算法,并通过实验证实了该算法的有效性<sup>[14]</sup>。KPP-POS 贪婪算法如表 2 所示。

表 2 关键节点检测的贪婪算法

Table 2 Greedy algorithm for key players detection
Procedure KPP-POS Greedy algorithm
Input: Graph G and the number of key players k
Output: the critical nodes set S
1. Select k nodes at random to populate set S
2. Set F=fit using appropriate key player metric
3. For each node <i>u</i> in S and each node <i>v</i> not in S
a. DELTAF=improvement in fit if <i>u</i> and <i>v</i> were swapped
4. Select pair with largest DELTAF
a. If DELTAF <=zero then terminate
b. Else, swap pair with greatest improvement in fit and
set F = F + DELTAF
5. Go to step 3
6. return S

这个算法首先随机选择 *k* 个节点作为关键节点填充集合 *S*,然后设置合适的初始目标值。随后对集合 *S* 中的每个节点 *u* 和每个不在集合 *S* 中的节点 *v* 进行交换,如果目标值得到改善,那就进行交换并更新目标值,反之取消交换,直到目标值不再得到优化时程序终止。

由于此贪婪算法选取起始关键节点集采用的方式,所以运行结果容易陷入局部最优解。为有效地解决该问题,通常使用数十个随机起始集重复实验,最终选取实验结果最好的目标值对应的关键节点集 *S*,但同时也增加整个求解过程的运行时间,即减缓全局收敛速度。

3.2 贪婪结合局部搜索的启发式算法

为了能够求解较大规模社交网络关键节点检测积极效应问题,本文在贪婪框架内结合局部搜索方法,设计了一个启发式算法。如表 3 所示。

这个算法首先选择能够使得目标值最优的第一个节点,即所有节点中度最大的一个节点。然后在删除已选择的关键节点的剩余网络中,加入可以使目标值最优的下一个节点,通过局部搜索方法,每次从已选择作为关键节点的集合中取出一个节点,与剩余网络中的一个节点置换,如果可以增加目标值(目标值为 max),那么进行置换;

表 3 关键节点检测的启发式算法

Table 3 Heuristic algorithm for key players detection
Procedure IP-KPP-POS Heuristic algorithm
Input: Graph G and the number of key players K
Output: the critical nodes set S*
1. Initialization $CNK = \varnothing$ .
2. While( $ CNK  \neq K$ ) do
3. $c \leftarrow \underset{i,j \in V}{\operatorname{argmax}} \{a_{ij} \cdot x_{ij}\}$
4. $CNK \leftarrow CNK \cup \{c\}$
5. $V \leftarrow V \setminus \{c\}$
6. End While
7. $S^* \leftarrow CNK$
8. For $[i,j] \in V * V$ do
9.If $i \in CNK$ and $j \notin CNK$ then
10. $CNK \leftarrow CNK \setminus \{i\}$
11. $CNK \leftarrow CNK \cup \{j\}$
12.If $obj(CNK) > obj(S^*)$ then
13. $S^* \leftarrow CNK$
14.Else
15. $CNK \leftarrow CNK \setminus \{j\}$
16. $CNK \leftarrow CNK \cup \{i\}$
17.End If
18.End If
19. End For
20. return $S^*$ ;

反之,取消置换。同理,按照该方法对集合中的其他节点做置换尝试,直至选择的关键节点个数达到预先设定的 *K* 值。

该启发式算法起始集选择使得目标值最优的节点,这种起始集的选择减少了得到局部最优解的可能并且加快了全局收敛速度。显而易见,该启发式算法的主要消耗在局部搜索增强操作上,复杂度为 *O*(*N*)。

4 实验分析

本章节将通过在特定随机图,多种类型随机网络和真实标准网络数据集上同 KPP-POS 贪婪算法综合实验对比分析,对提出的 IP-KPP-POS 模型、IP-KPP-POS 启发式算法进行验证。其中,IP-KPP-POS 模型、IP-KPP-POS 启发式算法都是使用 MATLAB 语言编程开发的,所有的模型和算法均在配置为 Windows 操作系统的台式电脑平台上运行。平台的具体配置为:主频 2.6 GHz 的骁龙 CPU;32 G 的内存。

4.1 特定 E-R 随机图

如图 1 所示,本文在节点规模为 15,边规模为 41,连接密度为 30% 的 ER 随机生成图上进行阐述。以 *D<sub>R</sub>* 指标作为网络积极效应的评测标准,

根据前文对  $D_R$  的定义,当  $D_R$  值越大,说明选择的一组节点 ( $KP$  集) 的成员与另一个节点 (网络的其余部分) 的成员之间的连接或内聚越强,即更具有重要性;反之亦然。实验结果如表 4 所示。

表 4 特定 E-R 随机图上的实验分析  
Table 4 Experimental analysis on a specific E-R random graph

Graph				IP-KPP-POS			IP-KPP-POS 启发式算法				KPP-POS 贪婪算法 <sup>[14]</sup>			
$N$	$E$	$K$	Obj	$KP$ 集	$T/ms$	$D_R$	Obj	$KP$ 集	$T/ms$	$D_R$	Obj	$KP$ 集	$T/ms$	$D_R$
15	41	1	9	{12}	15	<b>0.821 4 *</b>	9	{12}	<b>0.5</b>	<b>0.821 4 *</b>	9	{12}	5	<b>0.821 4 *</b>
15	41	2	12	{13,14}	17	<b>0.961 5 *</b>	12	{6,14}	<b>0.7</b>	<b>0.961 5 *</b>	12	{6,14}	7	<b>0.961 5 *</b>
15	41	3	12	{5,6,14}	18	<b>1 *</b>	12	{12,6,14}	<b>1.0</b>	<b>1 *</b>	12	{14,6,15}	10	<b>1 *</b>

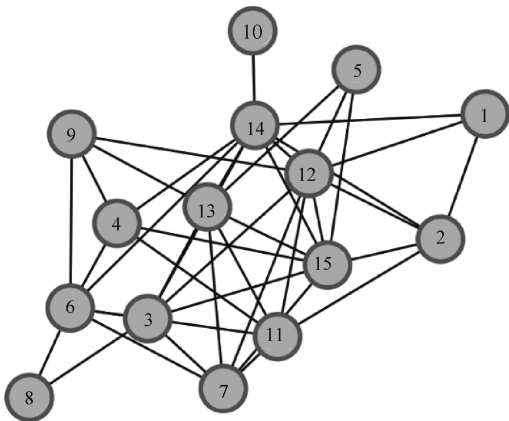


图 1 特定 E-R 随机图  
Fig.1 A specific E-R random graph

表 4 中,前两列分别表示实验中网络的节点数和边数,第 3 列表示关键节点的个数  $K$  值;Obj 表示对应方法求解所得到的目标值, $KP$  集表示求解得到的关键节点集, $T$  表示方法求解时间,单位

为 ms;而  $D_R$  表示衡量网络凝聚程度的优化效果指标。上标“\*”表示针对同一网络优化效果更好的一方。

当  $K$  值分别取 1、2 和 3,即网络中作为关键节点的节点数目分别为 1、2 和 3 时,3 种方法的  $D_R$  值都相同。这意味着在关键节点数目分别为 1、2 和 3 时,IP-KPP-POS 模型和 IP-KPP-POS 启发式算法跟 KPP-POS 贪婪算法在网络积极效应优化效果一致,但 IP-KPP-POS 启发式算法的运行时间要明显优于 IP-KPP-POS 模型和 KPP-POS 贪婪算法。

4.2 多种类型随机网络

为测试不同  $K$  值下,IP-KPP-POS 模型和 IP-KPP-POS 启发式算法的性能,针对节点数为 20、50、70、120,边数分别为 56、383、683、1 538,连接密度为 30% 的随机网络图,同 KPP-POS 贪婪算法进行实验对比分析,实验结果如表 5 所示。

表 5 多种类型随机网络上的实验分析  
Table 5 Experimental analysis on various types of random networks

Graph			IP-KPP-POS			IP-KPP-POS 启发式算法			KPP-POS 贪婪算法 <sup>[14]</sup>		
$N$	$E$	$K$	Obj	$T/ms$	$D_R$	Obj	$T/ms$	$D_R$	Obj	$T/ms$	$D_R$
20	56	1	9	17.5	<b>0.736 8 *</b>	9	<b>0.8</b>	<b>0.736 8 *</b>	9	6.7	<b>0.736 8 *</b>
20	56	2	14	19	<b>0.888 9 *</b>	14	<b>1.2</b>	<b>0.888 9 *</b>	14	7.8	<b>0.888 9 *</b>
20	56	3	16	30	<b>0.970 6 *</b>	16	<b>1.5</b>	<b>0.970 6 *</b>	16	9.7	<b>0.970 6 *</b>
50	383	1	24	47	<b>0.744 9 *</b>	24	<b>1.6</b>	<b>0.744 9 *</b>	24	8.3	<b>0.744 9 *</b>
50	383	2	36	108	<b>0.875 0 *</b>	36	<b>2.4</b>	<b>0.875 0 *</b>	36	12.3	<b>0.875 0 *</b>
50	383	3	43	224	<b>0.957 4 *</b>	43	<b>3.0</b>	<b>0.957 4 *</b>	43	18.5	<b>0.957 4 *</b>
70	683	1	28	49	<b>0.702 9 *</b>	28	<b>1.8</b>	<b>0.702 9 *</b>	28	10.7	<b>0.702 9 *</b>
70	683	2	47	157	<b>0.845 6 *</b>	47	<b>3.2</b>	<b>0.845 6 *</b>	44	19.8	0.823 5
70	683	3	57	9.0e4	<b>0.925 4 *</b>	55	<b>4.6</b>	0.910 4	55	28.4	0.910 4
100	1538	1	44	63	<b>0.722 2 *</b>	44	<b>2.7</b>	<b>0.722 2 *</b>	44	16.3	<b>0.722 2 *</b>
100	1538	2	69	4.3e3	<b>0.852 0 *</b>	68	<b>4.6</b>	0.846 9	67	30.6	0.841 8

表 5 中,前两列分别表示实验中网络的节点数和边数,第 3 列表示关键节点的个数  $K$  值;Obj 表示对应方法求解所得到的目标值, $T$  表示方法

求解时间,单位为 ms;而  $D_R$  表示衡量网络凝聚程度的优化效果指标。上标“\*”表示针对同一网络优化效果更好的一方。例如,第 1 行至第 3 行

表示对于一个节点数为 20,边数为 56 的网络图,当  $K$  值分别为 1、2、3 时,IP-KPP-POS 模型、IP-KPP-POS 模型启发式算法和 KPP-POS 贪婪算法的优化效果对比分析。

从第 9 行可以看出,当  $N=70, E=683, K=3$  时,IP-KPP-POS 模型、IP-KPP-POS 模型的启发式算法和 KPP-POS 贪婪算法的最优值为分别为 57、55 和 55,网络的  $D_R$  值分别为 0.925 4、0.910 4 和 0.910 4,IP-KPP-POS 模型的启发式算法优化效果和 KPP-POS 贪婪算法相同,但是 IP-KPP-POS 模型的启发式算法的求解时间远远优于 KPP-POS 贪婪算法和 IP-KPP-POS 模型的求解时间,而且 IP-KPP-POS 模型的优化效果明显优于 IP-KPP-POS 模型的启发式算法和 KPP-POS 贪婪

算法。

从整体的对比实验结果来看,虽然 IP-KPP-POS 模型的求解时间要比 KPP-POS 贪婪算法的求解时间长,但是 IP-KPP-POS 模型的求解精度要比 KPP-POS 贪婪算法更高;同时在同样的优化效果下,IP-KPP-POS 模型的启发式算法的求解时间远远优于 KPP-POS 贪婪算法和对应的 IP-KPP-POS 模型的求解时间。此外,IP-KPP-POS 模型的启发式算法和 IP-KPP-POS 模型求解的结果大部分一致,这也证实了启发式算法的正确性。

4.3 真实恐怖网络数据集

为验证 IP-KPP-POS 模型在真实网络上的性能,本节在如图 2 所示的真实恐怖网络数据集上进行实验分析,实验结果如表 6 所示。

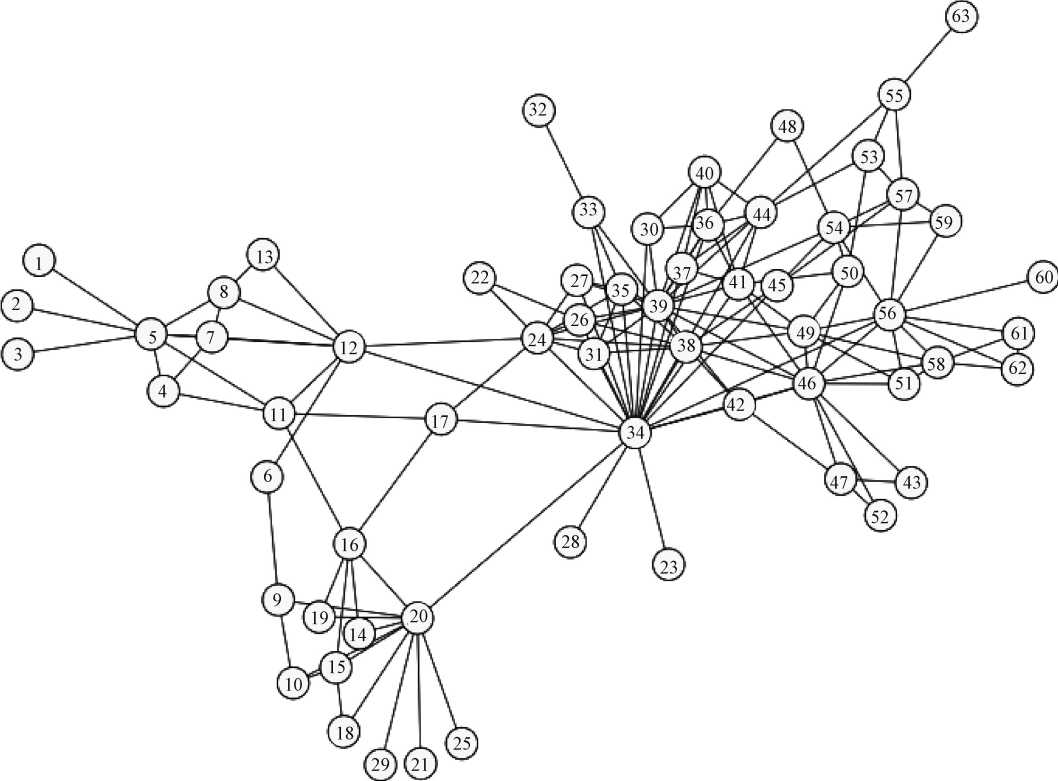


图 2 “9.11”恐怖网络数据集  
Fig.2 “9.11” terrorist network dataset

Krebs 的 911 劫持者数据集:以下数据集是由 Krebs 使用关于 9/11 恐怖分子的开源文献编制的<sup>[22]</sup>。该数据集包含与 911 事件相关的恐怖主义袭击事件人员之间的所有联系。数据集包含 63 个节点和 154 条边。

当  $K$  值取 3,也就是网络中作为关键节点的节点数目为 3 时,Borgatti 的 Key Player Set 模型<sup>[14]</sup>和 McGuire 的 WKPP-Pos 模型<sup>[17]</sup>选择的节

表 6 真实恐怖网络数据集上的实验比较		
Table 6 Comparison on a real terror network dataset		
方法	KP 集	$D_R$
Borgatti's Key Player Set <sup>[14]</sup>	{5, 34, 46}	0.788 9
McGuire's WKPP-POS <sup>[17]</sup>	{5, 34, 46}	0.788 9
IP-KPP-POS	{20, 34, 56}	<b>0.811 1*</b>

点集都是 5、34 和 46,  $D_R$  值为 0.788 9;而本文所建立的 IP-KPP-POS 模型选择的是节点集是 20、

34 和 56,  $D_R$  值为 0.811 1, 优化效果明显好于 0.788 9。通过比较可以看出, 本文建立的 IP-KPP-POS 模型在真实恐怖网络数据集上网络积极效应优化效果相比其他方法明显占优。

#### 4.4 Facebook 数据集

由于 IP-KPP-POS 模型的算法复杂度较高, 求解算法又受到 MATLAB 的内置约束限制, 目前在现有平台只能求解节点规模小于 100 的网络。为验证在大规模网络上的性能, 本章节在开源评测数据集网络“Facebook 数据集”<sup>[23]</sup>上对比 IP-KPP-POS 启发式算法和 KPP-POS 贪婪算法, 该数据集包含 347 个节点和 5 038 条边。实验结果如表 7 所示。

表 7 Facebook 数据集上的实验比较  
Table 7 Comparisons on the Facebook dataset

K	KPP-POS 贪婪算法 <sup>[14]</sup>			IP-KPP-POS 启发式算法		
	Obj	T/ms	$D_R$	Obj	T/ms	$D_R$
1	77	520	<b>0.485 6</b> *	<b>77</b>	<b>51</b>	<b>0.485 6</b> *
2	104	870	0.534 1	102	<b>86</b>	<b>0.536 5</b> *
3	129	1. 2e3	<b>0.618 0</b> *	129	<b>122</b>	<b>0.618 0</b> *
4	153	1. 5e3	<b>0.654 3</b> *	151	<b>165</b>	0.652 4
5	168	2. 1e3	0.677 7	169	<b>204</b>	<b>0.679 6</b> *
6	185	2. 4e3	0.713 1	182	<b>236</b>	<b>0.714 6</b> *
7	199	2. 7e3	<b>0.735 8</b> *	194	295	0.733 3
8	206	3. 2e3	0.751 7	208	<b>337</b>	<b>0.755 7</b> *
9	215	3. 9e3	0.766 8	216	<b>387</b>	<b>0.768 2</b> *
10	223	4. 1e3	<b>0.785 4</b> *	223	<b>427</b>	0.783 6

从表 7 可以看出, 当网络规模增加到 347 时, IP-KPP-POS 启发式算法的平均运行时间为 2. 3e2, 而 KPP-POS 贪婪算法的平均运行时间为 2. 2e3, KPP-POS 贪婪算法大约是 IP-KPP-POS 启发式算法平均运行时间的 10 倍。从  $D_R$  的结果对比来看, IP-KPP-POS 启发式算法所选择的关键节点集, 优于 KPP-POS 贪婪算法所选择的关键节点集。

对于更大规模的网络, 论文同样进行了实验分析。由于 KPP-POS 贪婪算法的复杂度远高于 IP-KPP-POS 启发式算法的复杂度。随着网络规模的增加, KPP-POS 贪婪算法的计算运行时间大幅增加。当网络规模为 958 时, IP-KPP-POS 启发式算法运行时间平均为 1. 7e5, 而 KPP-POS 贪婪算法的运算时间平均为 2. 0e6。因此, 在现有的计算资源有限的平台约束下, 因缺少对照实验数据, 很难进行更大规模网络数据集的验证评测。在未来研究中, 可以增加计算资源, 在更大规模的网络数据集上进行实验对照分析。

#### 4.5 实验结果

为了评测 IP-KPP-POS 模型以及 IP-KPP-POS 启发式算法的有效性和正确性, 同 KPP-POS 贪婪算法在特定 E-R 随机图, 多种类型随机网络以及真实标准网络数据集上进行对比实验。实验表明, 本文所提出的模型方法在多个性能指标上具有明显优势。特别地, 在小数据集上, IP-KPP-POS 模型以及 IP-KPP-POS 启发式算法在网络积极效应优化方面同样出色; 但随着数据规模的扩大, IP-KPP-POS 模型检测到的关键节点在网络积极效应优化精确度上比 IP-KPP-POS 启发式算法和 KPP-POS 贪婪算法的精确度更高一些, 但是 IP-KPP-POS 启发式算法的求解时间要明显优于 IP-KPP-POS 模型和 KPP-POS 贪婪算法的求解时间。

#### 5 总结

在本文中, 我们提出一种新的方法 IP-KPP-POS 解决复杂网络关键节点检测积极效应问题。由于问题的 NP 难特性, 又对该模型的求解算法深入研究。一方面, 利用 0-1 整数线性规划方法对模型求解, 满足小规模网络的快速精确求解; 另一方面, 设计贪婪局部搜索的启发式算法, 满足大规模网络的节点检测。通过在 E-R 随机网络、多种类型随机图、真实恐怖网络、Facebook 数据集等多种数据集上同 KPP-POS 贪婪算法进行对比实验, 验证模型的正确性和有效性。实验结果表明, 本文提出的 IP-POS-CNP 模型, 在准确性能上明显高于传统的 KPP-POS 方法, 并且所设计的对应的启发式算法在求解速度上具有显著优势, 更加适合大规模网络关键节点的快速检测。本文的研究成果可为各领域的关键节点相关问题提供决策支持。在未来的研究中, 本文将对有向有权图中关键节点的挖掘进一步探索研究。

#### 参考文献

[ 1 ] Borgatti P, Mehra A, Brass J, et al. Network analysis in the social sciences [J]. science, 2009, 323 (5 916): 892-895.

[ 2 ] 王伟, 刘军, 蒋熙, 等. 中国铁路网的拓扑特性[J]. 北京交通大学学报, 2010, 34(3): 148-152.

[ 3 ] Zhou T, Fu Z Q, Wang B H. Epidemic dynamics on complex networks [J]. Progress in Natural Science, 2005, 16 (5): 452-457.

[ 4 ] Latora V, Marchiori M. How the science of complex networks can help developing strategies against terrorism [J]. Chaos,

Solitons and Fractals,2004,20(1):69-75.

[ 5 ] Albert R,Jeong H,Barabasi A L. Error and attack tolerance of complex networks [J]. Nature,2000,406(6794):378-382.

[ 6 ] Kurant M,Thiran P,Hagmann. Error and attack tolerance of layered complex networks [J]. Physical Review E,2007,76(2):026103.

[ 7 ] 朱冠桦,蒋国平,夏玲玲. 社交网络上从众现象对谣言传播影响的研究[J]. 计算机科学,2016,43(2):135-139.

[ 8 ] 韩江漫. 基于动态复杂网络技术的病毒传播控制策略研究[J]. 计算机与数字工程,2017,45(10):2004-2008.

[ 9 ] 曹照. 基于相继故障的复杂脑网络研究[D]. 兰州:兰州理工大学,2016.

[10] 薛红艳. 金融危机通过资本市场对我国经济扩散研究[D]. 保定:河北大学,2014.

[11] 孙睿,罗万伯. 网络舆论中节点重要性评估方法综述[J]. 计算机应用研究,2012,29(10):3606-3608.

[12] Hu Q C,Gao Y, Ma P F, et al. A new approach to identify influential spreaders in complex networks [C] // Web-Age Information Management, Lecture Notes in Computer Science, New York:Springer,2013,62(14):99-104.

[13] 赵之滢,于海,朱志良,等. 基于网络社团结构的节点传播影响力分析[J]. 计算机学报,2014(4):753-766.

[14] Borgatti S P. Identifying sets of key players in a social network [J]. Computational & Mathematical Organization Theory, 2006,12(1):21-34.

[15] Hussain D M A. Investigation of key-player problem in terrorist networks using Bayes conditional probability [M]. Handbook of Social Network Technologies and Applications, 2010:523-547.

[16] Hamill J T, Deckro R F, Chrissis J W, et al. Analysis of layered social networks [J]. IO Sphere, 2008(1):27-33.

[17] McGuire R M,Deckro R F. The weighted key player problem for social network analysis [J]. Military Operations Research, 2015,20(2):35-53.

[18] Yang J. Generalized key player problem[J]. Computational & Mathematical Organization Theory,2015,21(1):24-47.

[19] Lü L, Chen D, Ren X L, et al. Vital nodes identification in complex networks[J]. Physics Reports,2016,650:1-63.

[20] Arulselman A, Commander C W, Eleftheriadou L, et al. Detecting critical nodes in sparse graphs [J]. Computers and Operations Research,2009,36(7):2193-2200.

[21] Jiang C, Wang J Y, Yu H, et al. An optimal approach for critical node problem using semidefinite programming [J]. Physica A: Statistical Mechanics and its Application,2017,471:315-324.

[22] Krebs V E. Uncloaking terrorist networks [J/OL]. First Monday,2002,7(4)[2017-12-20]. <http://firstmonday.org/ojs/index.php/fm/article/view/941/86>.

[23] McAuley J, Leskovec J. Learning to discover social circles in ego networks [C] // International Conference on Neural Information Processing Systems. Curran Associates Inc,2012:539-547.