

# 大规模网络中多传播源的重叠影响力问题研究

周明洋<sup>1,2</sup>, 付忠谦<sup>1\*</sup>, 廖好<sup>2</sup>

(1. 中国科技大学电子科学与技术系, 安徽合肥 230027; 2. 深圳大学计算机系, 广东深圳 518060)

**摘要:** 计算机和 Internet 的快速发展推动了网络科学的研究。信息传播是网络研究领域的热点, 信息传播的一个关键问题是初始信息源的选择, 传统的方法通过衡量节点重要性的指标(节点度、介数、邻居节点的重要性等), 然后根据单一节点的重要性从大到小排序, 依次选择重要节点。针对单一节点的传播能力, 传统的方法具有很好的效果, 但是在多传播源选择时效果并不好。本文首先分析了传统方法的不足, 并指出多节点的综合影响力没有提高是由于节点影响力的重叠效应。在此基础上, 本文提出了一个衡量多节点综合影响力的指标, 然后通过贪婪算法选择多个传播源节点。实际网络上的仿真实验表明, 针对传染病模型(SIR), 本文提出的方法能够选择更优的多传播源节点。进一步研究发现这些驱动节点间的平均距离更大, 稀疏性更高, 表明增加传播源节点的稀疏性能够降低节点的重叠影响力, 因此综合影响力得到提高。

**关键词:** 传播; 扩散; 节点影响力; 复杂网络

**中图分类号:** O414, TP311      **文献标识码:** A      doi:10.3969/j.issn.0253-2778.2016.01.005

**引用格式:** ZHOU Mingyang, FU Zhong qian, LIAO Hao. Overlapping influence of multiple spreaders in complex networks[J]. Journal of University of Science and Technology of China, 2016, 46(1):28-35.

周明洋, 付忠谦, 廖好. 大规模网络中多传播源的重叠影响力问题研究[J]. 中国科学技术大学学报, 2016, 46(1):28-35.

## Overlapping influence of multiple spreaders in complex networks

ZHOU Mingyang<sup>1,2</sup>, FU Zhong qian<sup>1\*</sup>, LIAO Hao<sup>1</sup>

(1. Department of Electronic Science and Technology, University of Science and Technology of China, Hefei 230027, China;  
2. Department of Computer Science, Shenzhen University, Shenzhen, 518060)

**Abstract:** With the development of computer technology and the Internet, network science is attracting many scientists from various fields. One field in network science is epidemic spreading, in which the key problem is the selection of source spreaders. Conventional methods select spreaders according to the importance of nodes (degree, betweenness and so on) and nodes with high importance are selected. Traditional methods perform well in characterizing the spreading ability of single nodes, but poorly in multiple nodes. An analysis is made and the reasons poor performance of multiple spreaders is attributed to the overlapping influences that decrease the overall spreading ability of multiple nodes. Then, an improved method is proposed to suppress the overlapping influences. The validity of the proposed method

收稿日期:2015-08-27;修回日期:2015-09-29

作者简介:周明洋,男,1987年生,博士生.研究方向:复杂网络、网络控制及数据挖掘. E-mail: zmy123@mail.ustc.edu.cn

通讯作者:付忠谦,副教授. E-mail: zqfu@ustc.edu.cn

is illustrated in four real-world networks in which the method could select better multiple spreaders. Further, it was found that improving the sparsity could reduce the overlapping influence of multiple spreaders, which enhances the overall spreading ability of nodes.

**Key words:** spreading; diffusion; influence of nodes; complex network

## 0 引言

计算机的快速发展使得分析大规模网络成为可能,而 Internet 尤其是 Web 2.0 为网络科学提供了海量和廉价的网络数据<sup>[1-2]</sup>. 这些数据包含 Internet 结构网络、社交网络、生物网络、维基百科知识关系网络等. 同时,这些来自不同领域、不同背景的网络在结构上和功能上呈现出一些共性,如无标度特征、等级结构、富人俱乐部特征、弱鲁棒性等,并存在着一些相似传播的动力学行为. 在经典的疾病传播模型中,节点会以一定的概率跟随邻居节点的状态,或者维持自身的状态不变,比较典型的应用有社会网络中的疾病传播<sup>[3]</sup>;在线社交网络的舆论传播<sup>[4]</sup>;计算机网络中的病毒传播<sup>[5]</sup>等. 随着在线社交网络的急剧膨胀,网络传播的最大化问题受到了广泛的关注,传播的最大化是指在传播源数量确定的情况下,通过选择不同的传播源节点,使得在稳态情况下有尽可能多的节点跟随源节点的状态<sup>[6-7]</sup>. 如在新商品宣传成本受限的情况下,通过少量名人的口碑宣传,能够达到大的宣传覆盖面. 单一节点的影响力可能很大,但是这并不代表多个重要节点的综合影响力一定最大;同时选择影视圈的 10 个名人的宣传效果肯定不如同时选择多个领域的名人(可以分别来自影视圈、体育圈、音乐圈等),宣传途径的多样化(电视、广播、网络等)实际上扩大了传播的覆盖面,因而综合影响力得到提高. 由于传播的最大化问题直接关系到新商品的推广效率、新闻的时效性等<sup>[8]</sup>,因而具有重大的研究价值和现实利益. 实际生活中,不同背景、不同领域重要节点的识别需要额外的节点属性信息,但是很多网络用户信息不全或者仅知道拓扑连接关系,在信息受限的情况下如何进行重要节点识别是当前的热点.

在重要节点的选择过程中,传统的方法主要通过节点的影响力选择,而节点影响力通过结构或者功能特征来刻画<sup>[9-10]</sup>. 通过网络结构衡量节点影响力时,最基本的方式是通过节点的度进行刻画,度大的节点在无标度 BA 模型网络中具有很好的传播能力. 而在实际网络中,Kitsak 等发现大度节点的影

响力也可能很低,进而发现  $K$ -壳可以更好地描述节点的中心性,并且  $K$ -壳越高的节点影响力越大<sup>[11]</sup>. 单个  $K$ -壳高的点影响力高,并不表示多个高  $K$ -壳节点也具有高影响力,这是由于节点影响力的重叠效应. 进一步,Zeng 等通过考虑节点的  $K$ -壳与外围节点的连接关系区分  $K$ -壳相同节点的影响力<sup>[12]</sup>. Liu 等考虑节点与网络中心( $K$ -壳最高的节点)的距离,并将此作为衡量节点重要性的指标,改进的算法极大提高了网络传播能力<sup>[13]</sup>. 针对节点的功能特征,基本的方法通过节点在传播过程中的重要性刻画影响力,如 PageRank 和 LeaderRank 方法<sup>[14-15]</sup>. 在此基础上,Chen 等通过考虑节点的聚类特征发现高聚类性会降低节点的传播能力,然后将聚类系数加入节点影响力. 仿真实验发现,考虑聚类系数的节点影响力更准确<sup>[16]</sup>. Cha 等根据社交网络中用户的行为(追随、转发、评论等)进行角色划分,分别进行影响力分析<sup>[17]</sup>. 进一步,Henderson 等根据节点的角色进行角色分类,不同于传统的数据聚类,具有相似功能但不一定连接的点会被分为一类,分类结果可直接用于重要节点识别<sup>[18]</sup>. Bakshy 等根据用户的转发次序进行划分,该方法可以直接追溯到信息源和重要影响力节点<sup>[19]</sup>.

除了这些常用的方法外,还有一些基于其他中心性指标的方法<sup>[14]</sup>. 如半局部中心性、离中心性、接近中心性、Katz 中心性等<sup>[20-22]</sup>以及根据随机游走理论演化出来的其他方法,比如特征向量指标、Hits 等. 这些方法都是从单个节点影响力的角度出发,多传播源节点选择时依次选择影响力高的节点. 由于存在影响力重叠效应,低影响力节点的组合也可能具有更好的效果. 如 Zou 等发现在无标度网络的牵引控制中,控制节点数量比较大时,控制度小的节点效果可能更好<sup>[23]</sup>. 这表明控制过程中,多个小度节点的组合在控制问题中的综合影响力更大,间接表明高影响力节点的组合不一定最优,因此有必要从多节点的综合影响力角度分析多节点的传播能力.

与传统的方法不同,本文从多节点的综合影响力出发,首先提出多节点的综合影响力计算指标,多节点的综合影响力不等于单一节点的影响力之和,

然后通过贪婪算法求得多个传播源节点. 该方法选择的单个传播源节点影响力可能并不是最大, 但是多个节点的重叠影响力被降低, 因而综合影响力得到了提高.

## 1 传播模型与节点重叠影响力

### 1.1 传播模型

传统的传播模型主要指 SI、SIS、SIR 或者 SIRS 等<sup>[24]</sup>.  $S$  (susceptible)——易感状态,  $I$  (infected)——感染状态,  $R$  (removed 或 recovered)——移除状态. 对于一个给定的网络  $G=(V, E)$ , 其中  $V$  表示顶点集合,  $E$  表示边集合, 初始时刻假设所有的顶点处于  $S$  状态, 然后一部分节点设置为  $I$  状态(初始传播源), 由于节点间的连接关系, 易感节点受感染邻居的影响逐渐变为感染节点, 同时感染节点也会以一定的概率恢复进入  $R$  或者  $S$  状态. 如果模型中仅存在  $S$  和  $I$  状态, 并且只存在从  $S$  到  $I$  状态的转变, 则为 SI 模型; 如果感染节点也会从  $I$  状态转移至  $S$  状态, 则为 SIS 模型. 进一步, 如果还存在  $R$  状态, 感染节点会恢复并进入免疫状态, 则为 SIR 模型; 如果免疫节点经过一段时间会重新变为易感节点, 则为 SIRS 模型.

由于本文主要利用 SIR 模型验证节点的传播能力, 因此主要介绍 SIR 模型. 在均匀网络中 SIR 的传播模型可以方程(1)表示<sup>[24]</sup>:

$$\left. \begin{aligned} \frac{dS}{dt} &= -\alpha SI + \beta R \\ \frac{dI}{dt} &= \alpha SI - \gamma I \\ \frac{dR}{dt} &= \gamma I - \beta R \end{aligned} \right\} \quad (1)$$

式中,  $\alpha$  为易感节点转为感染节点的概率,  $\beta$  为免疫节点转为易感节点的概率,  $\gamma$  为感染节点恢复并变为免疫节点的概率.

SIR 模型初始时刻包含易感节点和一部分感染节点, 之后感染节点以一定的概率  $p$  激活邻居节点, 假设一个易感节点有  $k$  个邻居, 其中  $k_1$  个邻居为感染节点, 那么在下一时刻该节点转变为感染节点的概率为  $1 - (1 - p)^{k_1}$ , 同时易感节点在每一步转变为免疫节点的概率为  $q$ . 为了更好地关注初始感染源对传播的影响, 仿真实验中设置为  $\beta = 0$ , 表示免疫节点不会变为易感节点.

传播问题的关键是选择合适的初始感染节点. 对于确定数量  $L$  的传播源节点, 传播最大化的问题

实际上是选择合适的传播源, 经过一定的时间, 网络中感染节点和免疫节点的数量达到稳定, 感染节点和免疫节点的数量之和可以表示传播源的传播能力. 当传播源数量为  $L$  时, 对于任意一种传播源组合, 若要判断这是最优方案, 存在  $CL_N$  种组合需要对比, 因此最优传播源选择问题是 NP 难问题; 一个合理的方式通过启发式算法选择初始传播源.

### 1.2 节点的重叠影响力

实际网络中节点之间存在重叠影响力, 如图 1 所示, 节点 1~5 属于全连通子图, 他们的度比网络中剩余节点的度大很多, 同时也具有更高的  $K$ -壳. 从单一节点的角度考虑这些节点的影响力都很高(从网络拓扑结构上看, 节点 2 和 4 比 1, 3, 5 的影响力略高); 但是如果选择多个重要节点作为传播源(假设选择 5 个), 全部选择节点 1~5 不太合理, 主要是由于节点 1~5 紧密连接, 重叠影响力较大, 合理的方式是选择方案中包含节点 7, 因为节点 7 属于影响力较大, 但与节点 1~5 影响力重叠较小. 从  $K$ -核的角度, Kitsak 等解释了单一节点的影响力, 认为节点 1~5 的  $K$ -壳较高, 处于网络中心, 单一节点影响力大. 对于多个重要节点没有给出完美的解决方案, 因此传统方法很容易忽略掉节点 7.

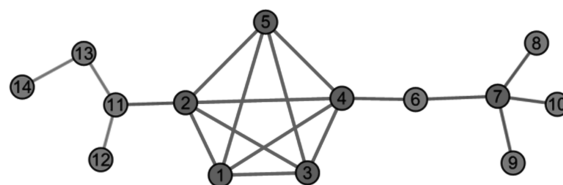


图 1 重叠影响力示意图

Fig. 1 Illustration of overlapping influence

注意到无论是基于节点度、 $K$ -壳、介数等方法或者基于传播模型的各种改进算法, 其核心思想都是识别网络中影响力高的节点, 很少关注多节点的重叠影响力, 因此有必要系统研究节点的重叠影响力, 降低重叠的影响, 从而提高多节点的综合影响力, 并在此基础上进行多传播源节点识别.

## 2 基于重叠影响力的多传播源选择方案

### 2.1 传播源节点选择方法

由于传统方法很少考虑节点影响力的重叠效应, 基于此, 本文在传统方法的基础上, 从综合影响力的角度出发设计多传播源选择方案.

具体地讲, 传统方法衡量节点的重要性时通常

根据节点的度、介数、 $K$ -核、PageRank 指标等. 以节点度指标为例, 单一节点的影响力通过节点与邻居的连接关系确定. 在多个传播源节点中, 传播源节点之间存在连边关系, 并且由于网络中存在“富人俱乐部”现象, 大度节点之间的连边比较稠密, 然而传播源或者感染节点之间的连边对于扩散没有意义. 实际上在多传播源中, 单一节点的传播能力应该表示为该节点与易感节点之间的连接关系, 该节点与其他感染节点之间的关系应该忽略. 基于此, 本文提出一个新的刻画多传播源节点综合影响力的指标  $\varphi$ ,

$$\varphi = \sum_{i \in S} w_i - \sum_{i, j \in S} w_{ij} \quad (2)$$

式中,  $S$  为传播源节点集合,  $w_i$  表示节点  $i$  的影响力 ( $w_i$  可以根据需要表示为节点度、介数、 $K$ -壳等),  $w_{ij}$  为节点  $i$  和  $j$  的重叠影响力.

本文中主要考虑两类方法, 基于度的综合影响力和基于介数的综合影响力. 在基于度的方法中,  $w_i$  定义为节点的度, 如果节点  $i$  和  $j$  之间有连边,  $w_{ij} = 2$ , 否则  $w_{ij} = 0$ .  $w_{ij} = 2$  是因为传播源  $i$  和  $j$  之间的连边在传播中无意义, 但在计算节点的影响力时被累加 2 次. 在基于介数的方法中,  $w_i$  定义为节点的介数,  $w_{ij}$  定义为边介数的 2 倍.

在给定传播源数量  $L$  的情况下, 初始时刻仅包含重要性最高的节点, 然后通过贪婪算法, 在每一步增加一个节点, 这个新增加的节点保证能够最大化  $\varphi$ . 具体算法如算法 2.1 所示,

**算法 2.1** Multiple\_spreader\_selection\_method

输入: 原始网络连接矩阵  $\mathbf{A} = \{a_{ij}\}_{N \times N}$ , 矩阵中元素为 1 或者 0 以及传播源节点数量  $L$ .

输出: 传播源节点索引数组.

第一步 网络数据预处理. 删除网络中的孤立节点和小节点簇, 仅保留最大的连通子图.

第二步 数据初始化. 根据节点的重要性指标, 分布计算节点的重要性(度、介数)和边的重要性, 并初始化传播源, 初始传播源仅包含重要性最高的节点(最大度、介数等节点).

第三步 通过贪婪算法. 在每一步中新增一个节点, 新增节点在每一步中都能够最大化目标函数  $\varphi$ . 重复此步骤, 直到选择的传播源达到设定的数量  $L$  为止.

第四步 将选择的传播源节点与原始网络的索引对应, 并返回.

## 2.2 评估标准

评估不同传播源传播能力的最重要指标是感染节点的比例, 在传播过程中一部分节点会转化为免疫节点, 因此传播能力表示为感染节点和免疫节点的和, 具体为:

$$\tau = \frac{N_I + N_R}{N} \quad (3)$$

式中,  $N_I$  表示为感染节点数量,  $N_R$  表示为免疫节点数量,  $N$  为网络总节点数. 对于不同的传播源节点,  $\tau$  越大越好, 表示节点的传播能力越强.

除了传播能力之外, 在研究传播源时更重要的是传播源的特点. 本文主要研究两个特点: 一是传播源节点间的平均距离  $d$ ; 二是与传统方法相比传播源的变化情况, 即不同方法选择的传播源的交集.

$$\sigma = \frac{S_1 \cap S_2}{S_1 \cup S_2} \quad (4)$$

式中,  $S_1$  与  $S_2$  分别是二种不同的传播源选择方案. 本文由于采用基于节点度和介数的改进方案, 主要比较两类相似性: 基于节点度改进的方案与基于度方法选择的节点相似性; 基于节点介数改进的方案与基于介数方法选择的节点相似性.

## 3 实验结果和分析

### 3.1 数据集描述

为了验证方案的有效性, 本文提出的方法分别在 4 个数据集上进行了有效测试. 这四个数据集来自斯坦福的公开数据库<sup>[25]</sup>, 分别为: as-Caida, ca-GrQc, ca-HepTh, facebook. as-caida 为 Internet AS 层的连接网络, ca-GrQc 是 Arxiv 上的广义相对论方向合作关系, ca-HepTh 是 Arxiv 上的高能物理方向合作关系网络, facebook 是 facebook 网络上的一个小的关系网络. 原始网络经过一定的预处理, 所有边都认为无权无向的, 同时删除孤立节点, 仅保留最大连通子图. 此外, 为了降低计算复杂度, 提取了网络的 4-核, 由于网络的自相似性, 子网络与原始网络有相似的结构特征<sup>[26-27]</sup>. 网络具体特征如表 1 所示.

### 3.2 实验结果和分析

为了验证算法的有效性, 本文在 SIR 模型中进行仿真实验, 其中  $S$  到  $I$  的感染概率为  $p$ ,  $I$  到  $R$  的恢复概率为  $q$ , 仿真中每一时间步传播扩散一次, 时间步长为  $T$ . 本文设置  $p=0.2$ ,  $q=0.7$ ,  $T=40$ . 实验中基于度的方法依次选择度从大到小的节点, 基

表 1 网络数据集

Tab. 1 Database of network

数据集	节点 (V)	边数 (E)	4-核节点 (V <sub>4</sub> )	4-核边数 (E <sub>4</sub> )
as-caida	26 475	53 381	2 175	12 967
ca-GrQc	4 158	13 422	1 471	8 302
ca-HepTh	8 638	24 896	3 178	14 814
facebook	1 034	26 749	978	26 641

于介数的方法依次选择介数从大到小的节点,基于  $K$ -壳的方法根据  $K$ -壳的大小进行选择. 本文提出的方法选择传播源节点时根据算法 2.1 选择,并基于节点度和介数有两种变形,分别表示为  $\varphi$ -degree 和  $\varphi$ -betweenness. 同时在仿真实验中,  $\delta = \frac{I_0}{N}$  表示初始感染节点比例. 实验首先观察不同方法下感染节点与传播源数量的关系,然后分析传播源节点的稀疏程度(源节点平均距离),最后分析不同方法选择

的传播源集合的重叠程度.

图 2 描绘了在  $p=0.2, q=0.7$ , 时间步长  $T=40$  时网络中感染过的节点比例与传播源数量的变化情况,其中横坐标为传播源节点的比例,纵坐标为感染过的节点比例,为了提高传播能力,  $\tau$  越大越好. 实验结果为 100 次实验的平均值. 图 2 中,经过改进后的度和介数方法受感染的节点比例增加很多,并且都比基于  $K$ -核的方法高. Kitsak 等认为  $K$ -核高的节点传播能力强,但是从多个节点的角度考虑,高  $K$ -核的节点重叠影响力很大,因而其综合影响力并不高. 需要注意的是在图 2(d)中,  $\delta < 0.35$  时,  $\tau$  和基于介数的方法基本重合. 这主要是因为本文目标函数的缺陷,虽然本文提出的目标函数可以降低多传播源节点的重叠影响力,但并没有完全消除重叠影响力. 当  $\delta > 0.35$  时,  $\tau$  快速增长,表明多传播源的综合影响力得到极大提高,也表明其重叠影响力被降低,证明了方案的可行性.

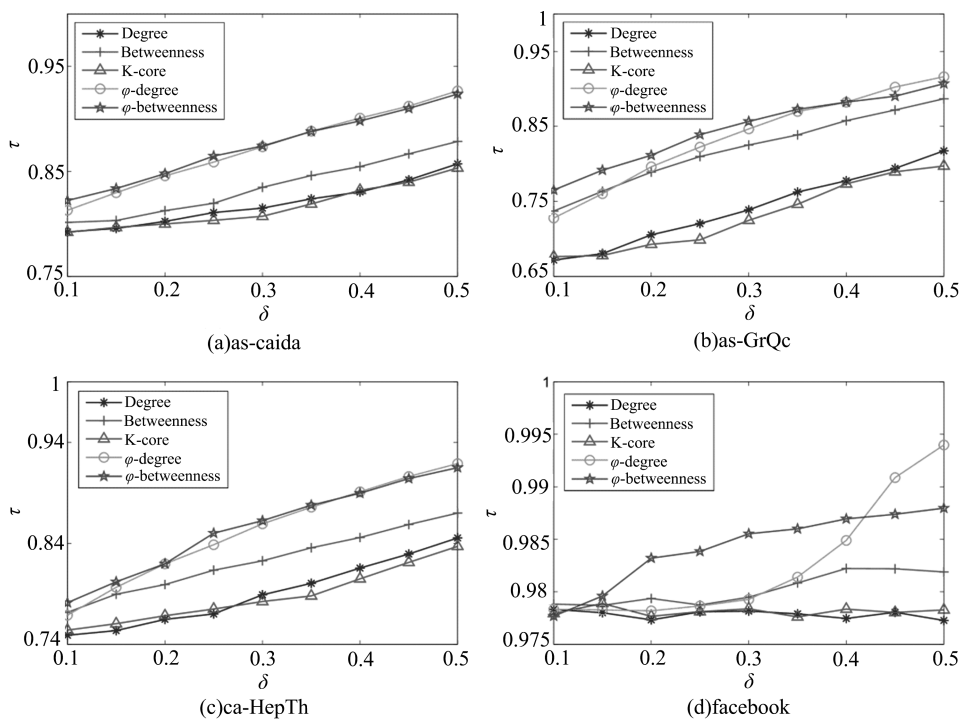


图 2 感染节点和免疫节点比例与初始传播源比例的关系

Fig. 2 The proportion of infected and recovered nodes as a function of source spreaders

为了更好地刻画节点重叠影响力,图 3 描绘了传播源节点之间的平均距离与传播源节点比例的变化情况. 图 3 中,经过改进的算法选择的传播源节点具有更高的平均距离,表明传播源在网络中分布比较稀疏,因此他们的重叠影响力比较低. 注意到基于度、介数和  $K$ -核的方法平均距离随着  $\delta$  的增长而增

长;在 as-caida 网络中,改进的算法平均距离随着  $\delta$  的增长而降低. 总体来讲,改进算法选择的驱动节点间平均距离更大,基于度、介数和  $K$ -核从网络的核心向外围逐渐选择重要节点,并且改进的方法首先选择非常稀疏的重要节点,然后选择的节点处于这些稀疏节点之间,因此平均距离比传统方法高.

为了更好地观察图 3 中中差异,图 4 描绘了经过  $\varphi$  改进的度方法与传统度方法的差异以及经过  $\varphi$

改进介数方法与传统介数方法的差异. 经过改进的方法与传统方法选择的驱动节点重合率很低(低于

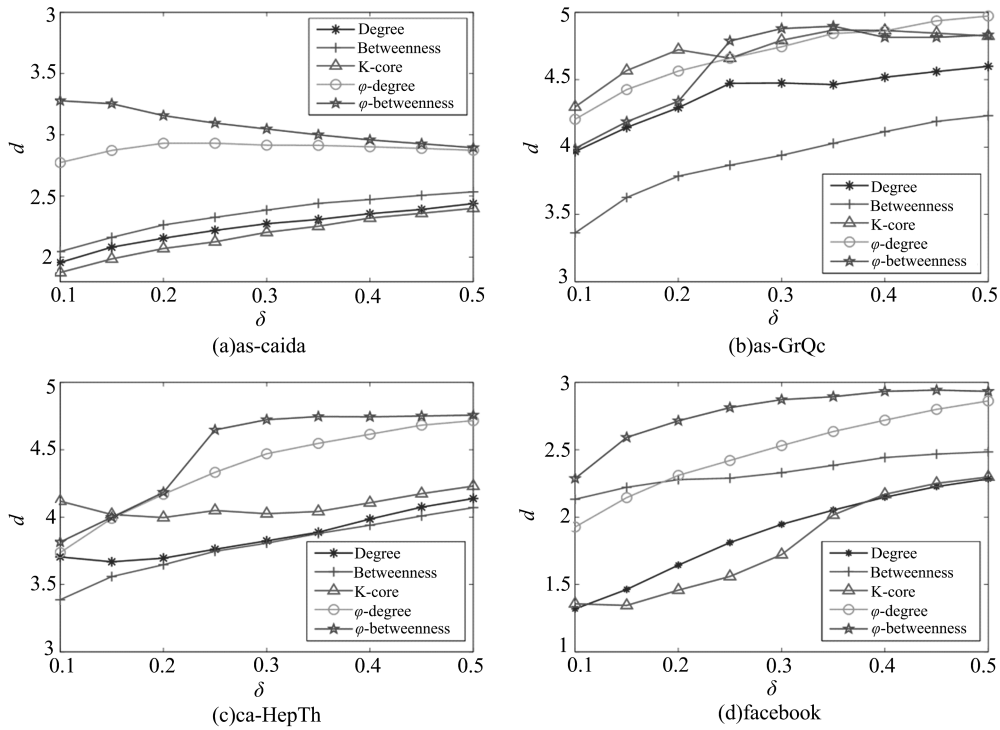


图 3 传播源节点的平均距离  
Fig. 3 The average distance of source spreaders

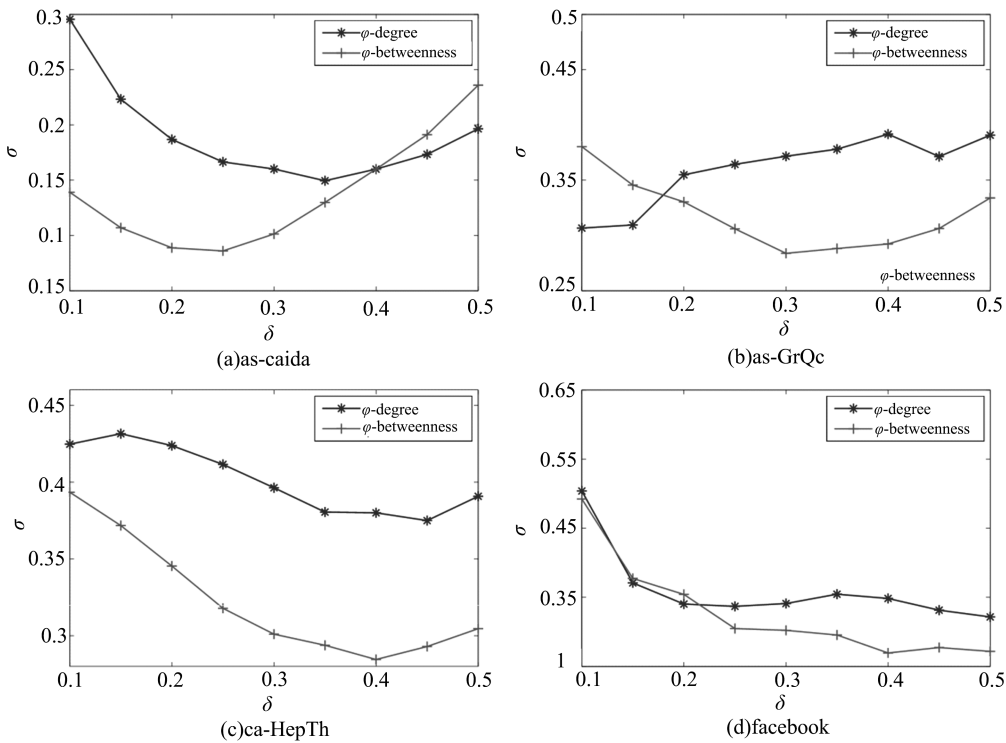


图 4 改进算法与传统方法选择的传播源节点的交集  
Fig. 4 The intersection between improved methods and traditional methods

0.5),但是在一些网络中(如 facebook 中,  $\delta < 0.15$ )传播能力相似(图 2(d)),表明多传播源选择方案的多样性,改进算法选择的传播源中单个节点影响力不高,但是其综合影响力也能达到很高,甚至超过高影响力节点的组合.同时,即使在传播源数量很少时,不同方法选择的传播源重合度也很低(小于 0.5),表明改进的方法能够选择更加稀疏、重叠影响力更低的传播源节点.

## 4 结论

本文从多节点的综合影响力角度出发,提出一种选择多传播源节点的算法,并对选择的传播源节点特征进行了分析.实验发现,该算法选择的节点传播能力更强.传统的算法根据节点的重要性排序,依次选择重要性高的节点.不同于传统算法,本文考虑传播源节点之间的重叠影响力,并提出了一种消除重叠影响力的方法,在此基础上通过贪婪算法选择传播源节点.在四个实际网络上的传播实验中,多传播源的传播能力得到显著提高,证明了方法的有效性.实验中还发现,本文设计的目标函数不是最优,如何设计更优良的目标函数以降低多传播源节点间的重叠影响力是多传播源选择问题的关键,也是本文的后续工作.虽然本文提出的目标函数不是最优,但是这个目标函数也适合于传统的重要性选择算法,如 PageRank、Hits 等,这些传统的方法都可以用本文的方法进行改进.本文从新的角度理解传播源的选择问题:从重叠影响力角度出发提高传播源的综合传播能力,为传播源选择提供了一种新思路.

### 参考文献(References)

- [1] Boccaletti, S., et al., Complex networks: Structure and dynamics. *Physics reports*, 2006, 424(4): p. 175-308.
- [2] Strogatz S H. Exploring complex networks [J]. *Nature*, 2001, 410(6825): 268-276.
- [3] Diekmann O, Heesterbeek J A P. *Mathematical epidemiology of infectious diseases: Model building, analysis and interpretation* [M]. Chichester: John Wiley & Sons, 2000.
- [4] Chierichetti F, Lattanzi S, Panconesi A. Rumor spreading in social networks [C]// *International Colloquium on Automata, Languages and Programming*. Springer, 2009: 375-386.
- [5] Staniford S, Paxson V, Weaver N. How to own the internet in your spare time [C]// *Proceedings of the 11th USENIX Security Symposium*. Berkeley, USA: ACM Press, 2002:149-167.
- [6] Kempe D, Kleinberg J, Tardos É. Maximizing the spread of influence through a social network [C]// *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge discovery and Data Mining*. Washington, USA: ACM Press, 2003: 137-146.
- [7] Chen W, Wang Y, Yang S. Efficient influence maximization in social networks [C]// *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Paris, France: ACM Press, 2009: 199-208.
- [8] 苏晓萍, 申情, 宋玉蓉, 等. 利用社会网络上最有影响力节点实现高效病毒营销 [J]. *小型微型计算机系统*, 2014, 35(8): 1803-1807.  
Su Xiaoping, Shen Qing, Song Yurong, et al. Improving the efficiency of viral of marketing in a social network using influential nodes [J]. *Journal of Chinese Computer Systems*, 2014, 35(8): 1803-1807.
- [9] Kempe D, Kleinberg J, Tardos É. Influential nodes in a diffusion model for social networks [C]// *International Colloquium on Automata, Languages and Programming*. Springer, 2005: 1127-1138.
- [10] Chen D B, Lü L Y, Shang M S, et al. Identifying influential nodes in complex networks [J]. *Physica A: Statistical Mechanics and its Applications*, 2012, 391(4): 1777-1787.
- [11] Kitsak M, Gallos L K, Havlin S, et al. Identification of influential spreaders in complex networks [J]. *Nature Physics*, 2010, 6(11): 888-893.
- [12] Zeng A, Zhang C J. Ranking spreaders by decomposing complex networks [J]. *Physics Letters A*, 2013, 377(14): 1031-1035.
- [13] Liu J G, Ren Z M, Guo Q. Ranking the spreading influence in complex networks. *Physica A: Statistical Mechanics and its Applications*, 2013, 392(18): 4154-4159.
- [14] 任晓龙, 吕琳媛, 网络重要节点排序方法综述. *科学通报*, 2014, 59(13): 1175-1197.  
Ren Xiaolong, Lü Linyuan. Review of ranking nodes in complex networks [J]. *Chinese Science Bulletin*, 2014, 59(13): 1175-1197.

- [15] 刘建国, et al., 复杂网络中节点重要性排序的研究进展. 物理学报, 2013. 62(17): p. 178901-178901.
- [16] Chen D B, Gao H, Lü L, et al. Identifying influential nodes in large-scale directed networks: the role of clustering[J]. Plos One, 2013, 8(8): e77455.
- [17] Cha M, Haddadi H, Benevenuto F, et al. Measuring user influence in twitter: The million follower fallacy [C]// Proceedings of the 4th International Conference on Weblogs and Social Media. Washington, USA: ACM Press, 2010, 14: 30.
- [18] Henderson K, Henderson K, Gallagher B, et al. Rolx: Structural role extraction & mining in large graphs[CP]// Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Beijing, China: ACM Press, 2012: 1231-1239.
- [19] Bakshy E, Hohman J M, Mason W, et al. Everyone's an influencer: Quantifying influence on twitter[C]// Proceedings of the 4th ACM International Conference on Web Search and Data Mining. Hong Kong, China: ACM Press, 2011: 65-74.
- [20] Hou B N, Yao Y P, Liao D B. Identifying all-around nodes for spreading dynamics in complex networks[J]. Physica A: Statistical Mechanics and Its Applications, 2012. 391(15): 4012-4017.
- [21] Gao S, Ma J, Chen Z M, et al. Ranking the spreading ability of nodes in complex networks based on local structure[J]. Physica A: Statistical Mechanics and Its Applications, 2014, 403: 130-147.
- [22] Fu Y H, Huang C Y, Sun C T. Identifying super-spreader nodes in complex networks[J]. Mathematical Problems in Engineering, 2015, 2015: 675713(1-8).
- [23] Zou Y L, Chen G R. Choosing effective controlled nodes for scale-free network synchronization [J]. Physica A: Statistical Mechanics and its Applications, 2009. 388(14): 2931-2940.
- [24] 汪小帆, 李翔, 陈关荣. 网络科学导论[J]. 北京: 高等教育出版社, 2012.
- [25] Leskovec J, Krevl A. SNAP Datasets: Stanford Large Network Dataset Collection [EB/OL]. <http://snap.stanford.edu/data>.
- [26] Song C M, Havlin S, Makse H A. Self-similarity of complex networks[J]. Nature, 2005, 433 (7024): 392-395.
- [27] Leskovec J, Lang K J, Dasgupta A, et al. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters[J]. Internet Mathematics, 2009. 6(1): 29-123.