



# 基于改进的度折扣方法研究社交网络 影响力最大化问题

夏欣<sup>1</sup>, 马闯<sup>2</sup>, 张海峰<sup>1\*</sup>

(1. 安徽大学数学科学学院 合肥 230601; 2. 安徽大学互联网学院 合肥 230601)

**【摘要】**在影响力最大化检测算法中, 度折扣算法是一个高效的启发式算法。针对现有度折扣算法中的不足, 该文对其计算期望影响力的公式进行修正, 提出了一阶改进的度折扣算法, 并进一步引入冗余弱化机制确保种子节点分散地分布在网络上, 得到了二阶改进的度折扣算法。基于独立级联模型, 在 4 个真实网络上与其他算法进行比较, 实验结果表明提出的两种算法使信息扩散速度更快、更广, 还能保证较低的时间复杂度。

**关键词** 度折扣算法; 期望影响力; 影响力最大化; 冗余弱化机制; 社会网络  
**中图分类号** TN92; O41 **文献标志码** A **doi**:10.12178/1001-0548.2020338

## An Improved Degree Discount Approach for Influence Maximization in Social Networks

XIA Xin<sup>1</sup>, MA Chuang<sup>2</sup>, and ZHANG Hai-feng<sup>1\*</sup>

(1. School of Mathematical Science, Anhui University Hefei 230601; 2. School of Internet, Anhui University Hefei 230601)

**Abstract** In the influence maximization detection algorithm, the degree discount algorithm is an efficient heuristic algorithm. Aiming at the shortcomings of the degree discount algorithm, the formula for calculating the expected influence is modified and the first-order improved degree discount algorithm is proposed. Furthermore, in order to ensure the seed nodes are scattered in the network, a redundancy weakening mechanism is introduced and then the second-order improved degree discount algorithm is constructed. Based on the independent cascade model, the proposed algorithms are compared with other algorithms in four real networks. The experimental results confirm that the proposed algorithms can ensure faster and wider information spreading with low time complexity.

**Key words** degree discount algorithm; expected influence; influence maximization problem; redundancy weakening mechanism; social networks

在线社交平台如 Twitter、Weibo、WeChat、Facebook 等已经成为人们生活中不可或缺的重要工具, 引起了不同领域的学者对社交网络的广泛关注与研究, 如社交网络上的影响力最大化问题、链路预测、社团发现及推荐系统等<sup>[1-4]</sup>。其中影响力最大化问题因其普遍的应用场景而被广泛研究, 如在广告投放和市场营销中如何用有限的成本选择具有较大传播力度的人物、地点投放广告, 使得了解并且购买该产品的用户最多, 从而获取最大收益。在疾病和谣言的传播链中, 控制网络中影响力大的节点可以避免其大规模快速传播<sup>[5-6]</sup>。

影响力最大化问题是指在一个网络中选择一定

数目的节点作为种子集, 通过激活这些种子节点, 使得信息在网络中传播, 最终希望网络中被激活的节点数最多。影响力最大化算法主要分为 3 类: 贪婪算法、中心性指标和启发式算法。

在贪婪算法中, 文献 [7] 将影响力最大化问题定义为一个离散的优化问题, 证明了在独立级联模型和线性阈值模型下该问题是一个 NP-Hard 问题。并进一步提出了近似比为  $(1-1/e)$  的爬山贪心算法, 在每一步迭代中都选择当前边际传播范围最广的节点加入种子集。文献 [8] 利用子模性提出了 CELF (cost-effective lazy forward) 算法, 该算法大大减少了计算时间。文献 [9] 提出了改进的 CELF++

收稿日期: 2020-08-21; 修回日期: 2021-02-23

基金项目: 国家自然科学基金 (61973001, 12005001); 安徽省自然科学基金 (2008085QF299)

作者简介: 夏欣 (1998-), 女, 主要从事复杂网络关键点识别等方面的研究。

通信作者: 张海峰, E-mail: haifengzhang1978@gmail.com

算法, 减少了不必要的计算次数。文献 [10] 提出 NewGreedy 算法, 以传播概率保留网络中的边, 根据子图的连通性考虑节点加入种子集的增益。由于 NewGreedy 算法无法保证子模性, 文献 [11] 提出了一个静态的贪婪算法 StaticGreedy 算法, 能够解决影响力最大化中的精确度可扩展性问题。贪婪算法虽然能找到影响力较大的种子节点, 但其时间复杂度太高, 不适用于大型网络。

第二类解决影响力最大化问题的方法是基于网络的拓扑结构对节点的影响力排序, 使用中心性指标对网络中的节点排序, 选择前  $K$  个中心性指标最大的节点作为种子节点<sup>[12]</sup>。基于网络结构定义的中心性指标有度中心性指标、接近度中心性及介数中心性等<sup>[13-16]</sup>。文献 [17] 系统回顾了关键点识别中的概念和度量, 并对问题和方法进行了分类, 通过广泛的实证分析, 比较不同方法的适用性。此外, 很多学者提出新的影响力排序方法, 如文献 [18] 认为种子节点应该分散在网络中, 定义了 GCC (generalized closeness centrality) 指标。文献 [19] 提出网络节点的 H 指数衡量节点的重要性。文献 [20] 通过构造算子说明了节点的度、H 指数和核数之间的关系, 该关系被定义为 DHC 定理。文献 [21] 结合节点的 K-shell 值和节点间最短路径长度, 将 K-shell 值看做节点重力, 提出重力中心性指标。

使用中心性指标选取种子节点, 忽略了节点间的重叠性, 可能造成传播过程中的冗余, 为了寻找更加高效的解决方法, 启发式算法被相继提出。如文献 [22] 考虑局部影响区域迭代节点的全局影响力提出了 IRIE (influence ranking influence estimation) 算法。文献 [23] 通过节点间的简单路径近似计算节点的传播影响力, 并且考虑不同的优化方法减少扩展中迭代调用的次数, 提出了 Simpath 算法。文献 [24] 介绍了识别网络中单个有影响力节点的方法、特点和性能, 同时介绍利用启发式算法寻找多个影响力节点的研究方法。文献 [25] 考虑种子集到其他节点的最短路径得到种子集影响范围, 提出 SPM (shortest-path model) 算法。文献 [26] 基于节点相似性提出了启发式的聚类算法, 选取聚类质心作为种子节点获得较大传播范围。文献 [10] 基于节点的度提出了启发式的度折扣 (degree discount, DD) 算法, 度折扣算法比贪婪算法的运行速度快了近千倍, 并且能达到与贪婪算法较为接近的效果, 因此受到广泛关注。

虽然度折扣算法具有快速、高效的特性, 但该

算法还有许多需要改进的地方, 这些不足之处是制约该算法性能进一步提升的关键因素。首先, 度折扣算法在计算节点的期望影响力时没有考虑邻居的差异性, 而是简单地认为每个未激活的邻居节点对该节点期望影响力的贡献是相同的, 导致计算期望影响力的公式不够精确。其次, 度折扣算法没有考虑节点之间共同邻居数的影响, 不能充分降低传播的冗余性, 比如节点  $i$  和  $j$  之间有许多共同邻居, 如果节点  $j$  已被选择为种子节点, 则节点  $i$  被感染的可能性也很高, 因为他们之间有多条可能的传播路径, 此时再选择节点  $i$  作为种子节点会导致传播的冗余效应<sup>[27]</sup>。针对以上两个问题, 本文首先对度折扣算法中计算期望影响力的公式进行修正, 然后基于共同邻居数引入冗余弱化效应进一步改进度折扣算法。

## 1 影响力最大化与传播模型

### 1.1 影响力最大化问题描述

用网络图  $G = (V, E)$  表示社交网络, 其中  $V$  是节点集, 代表社交网络中的用户;  $E$  是边集, 代表社交网络中用户之间的关系。影响力最大化问题就是在网络中找出  $K$  个节点作为种子集  $S$ , 然后按照规定的传播模型将信息传播给邻居节点, 使得最终能影响的节点数最多。在特定的传播模型  $M$  下, 任意种子节点集  $S$  的影响力可以表示为  $\sigma_M(S)$ 。因此, 网络上影响力最大化问题的形式化定义为:

$$S^* = \arg \max_{S \subseteq V, |S|=K} \sigma_M(S) \quad (1)$$

### 1.2 独立级联模型

在独立级联模型 (independent cascading model, IC) 中, 网络中的节点有两种可能的状态: 未激活态和激活态。同时为网络中的每条边  $(i, j)$  分配一个  $[0, 1]$  间的概率值  $\beta_{ij}$  作为信息在该条边上的传播概率。IC 模型具体传播过程如下:

1) 在  $t = 0$  时刻, 选定  $K$  个节点构成种子集  $S_0$ , 仅将种子集中节点设为激活状态, 其余节点为未激活状态。

2) 在  $t + 1$  时刻, 每一个激活节点  $i \in S_t$ , 以概率  $\beta_{ij}$  尝试激活邻居中的未激活节点  $j$ 。若激活成功, 则该邻居节点  $j$  由未激活状态转为激活状态, 否则仍保持未激活状态。并且每个激活节点只有一次激活邻居的机会, 无论是否成功激活, 该节点在下一轮将不具有激活能力。节点  $j$  有多个种子邻居时, 每个种子邻居都独立的激活节点  $j$ , 如种子节点  $i$  和  $h$  均为节点  $j$  的邻居, 则  $j$  被激活的概率为  $1 - (1 - \beta_{ij})$

$(1-\beta_{h_j})$ 。

3) 直到某个时刻, 网络中所有节点都不再具有激活其他节点能力时, 传播过程结束。

## 2 改进的度折扣算法

### 2.1 度折扣算法

度折扣算法的基本思想是: 假设节点  $j$  是节点  $i$  的邻居, 如果  $j$  已被选为种子节点, 那么在基于度中心性指标考虑节点  $i$  是否作为种子节点时, 应该对连边  $(i, j)$  打折扣, 因为  $i$  对  $j$  不能产生额外的影响。假设所有边的激活概率都相同, 均为  $\beta$ 。当节点  $i$  的邻居中有  $s_i$  个激活种子时,  $i$  被激活的概率为  $1-(1-\beta)^{s_i}$ , 此时  $i$  节点能被邻居节点激活, 其期望影响力与直接将  $i$  节点选为种子节点的期望影响力相同, 即此时选择节点  $i$  作为种子节点不增加额外的影响力 (对期望影响力的贡献为 0)。由于节点  $i$  没有被激活的概率为  $(1-\beta)^{s_i}$ , 当节点  $i$  被选为种子时, 其能激活的节点数为  $1+(d_i-s_i)\beta$ , 其中 “1” 表示节点  $i$  被激活, “ $(d_i-s_i)\beta$ ” 表示被激活的邻居数。因此考虑节点  $i$  选为种子时, 其产生的期望影响力为:

$$\frac{[1+(d_i-s_i)\beta](1-\beta)^{s_i}}{1+[d_i-2s_i-(d_i-s_i)s_i\beta]\beta} \triangleq A \quad (2)$$

当节点  $i$  邻居中没有种子节点时,  $i$  作为种子节点产生的期望影响力为:  $1+d_i\beta \triangleq B$ 。设  $\gamma$  是对邻居中每个种子节点的度折扣, 则  $s_i\beta\gamma \triangleq B-A$ , 可以得到  $\gamma = 2+(d_i-s_i)\beta$ , 因此, 当节点  $i$  有  $s_i$  个种子邻居时, 它的度折扣值定义为:

$$dd_i = d_i - s_i\gamma = d_i - 2s_i - (d_i - s_i)s_i\beta \quad (3)$$

度折扣算法的基本步骤为: 第一轮中没有种子节点, 所有节点的度都没有被折扣, 所以直接选择网络中度最大的节点作为第一个种子; 接下来每一轮根据式 (3) 计算每个未被激活节点的度折扣值  $dd_i$ , 并选择最大的一个节点加入种子集; 循环更新计算直到选出  $K$  个种子节点加入种子集  $S$ 。

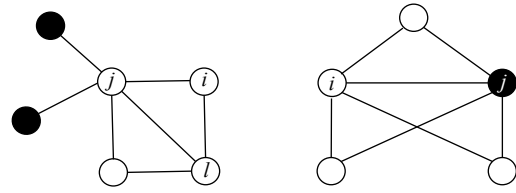
### 2.2 一阶改进的度折扣算法

度折扣算法认为所有非激活的邻居对节点  $i$  贡献的期望影响力都是相同的, 即为  $\beta$ 。实际上这些未激活邻居对节点  $i$  贡献的期望影响力是不同的, 如节点  $i$  有两个未激活邻居  $j$  和  $l$ , 如果节点  $j$  的邻居中已经有很多种子, 而节点  $l$  的邻居中没有种子节点, 如果节点  $j$  和  $l$  均被激活, 则节点  $l$  有更大的可能性是被  $i$  激活的, 而节点  $j$  很可能被其他节点激活。因此节点  $j$  和  $l$  对节点  $i$  期望影响力的贡献是截然不同

的, 如图 1a 所示, 其中实心黑色节点处于激活状态, 空心节点为未激活状态。基于此, 本文对计算期望影响力的公式进行如下修正 (IDD1 算法):

$$Ie_i = (1-\beta)^{s_i} \left[ 1 + \sum_{j \in \Gamma(i), j \notin S} a_{ij}\beta(1-\beta)^{s_j} \right] \quad (4)$$

式中,  $a_{ij}$  表示邻接关系,  $a_{ij} = 0$  表示节点  $i$  与  $j$  没有连边,  $a_{ij} = 1$  表示节点  $i$  与  $j$  有连边。在考虑节点  $i$  的期望影响力时需要考虑非激活邻居节点自身邻居中种子节点的数量, 若这些非激活邻居周围种子节点越多, 则对  $i$  贡献的期望影响力越小。



a.  $j$  和  $l$  对节点  $i$  影响力贡献不同 b.  $j$  通过直接连边或共同邻居激活  $i$

图 1 网络示例图

IDD1 算法的基本步骤为: 第一轮中直接选择网络中度最大的节点作为第一个种子; 接下来每一轮根据式 (4) 计算每个非激活节点的期望影响力  $Ie_i$ , 选择最大的一个节点加入到种子集; 循环更新计算直到选出  $K$  个种子加入  $S$ 。

### 2.3 二阶改进的度折扣算法

在选择多个节点作为种子节点时, 除了要考虑节点自身的重要性, 另一个重要因素是如何确保种子节点在网络上分布充分分散, 避免传播的冗余性。考虑如下情景: 节点  $i$  与节点  $j$  之间有许多共同邻居, 对于 IC 模型而言, 若节点  $j$  已经被选择为种子节点, 则节点  $i$  有很大的概率被激活 (通过直接连边激活和多个共同邻居激活), 此时再选择节点  $i$  作为种子节点对提高传播影响力的作用是非常有限的, 因此选择的种子节点之间应该避免有多个共同邻居的情况, 如图 1b 所示。基于此, 本文在 IDD1 算法基础上引入基于共同邻居数的冗余弱化机制, 保证种子节点的分布足够分散。令  $C_{ij}$  为节点  $i$  和  $j$  的共同邻居数, 定义节点  $i$  的影响期望力为  $Ce_i$ , 则:

$$Ce_i = \left[ 1 + \sum_{k \in \Gamma_i, k \notin S} a_{ik}\beta(1-\beta)^{s_k} \right] \prod_{j \in \Gamma_i \cap S} (1-\beta)e^{-\alpha C_{ij}} = (1-\beta)^{s_i} \left[ 1 + \sum_{k \in \Gamma_i, k \notin S} a_{ik}\beta(1-\beta)^{s_k} \right] e^{-\alpha \sum_{j \in \Gamma_i \cap S} C_{ij}} \quad (5)$$

式中,  $\alpha > 0$  是可调参数, 如果  $\alpha$  太大会导致所有邻居的贡献为 0, 若  $\alpha$  太小会导致邻居的贡献差异性被忽略, 本文取  $\alpha = 0.05$ 。

IDD2 算法的基本步骤如下: 首先选择网络中度最大的节点作为第一个种子; 之后每一轮根据式 (5) 计算其他未激活节点的  $Ce_i$ , 选择  $Ce_i$  值最大的节点作为种子; 循环更新计算直到选出  $K$  个种子加入  $S$ 。

### 3 实验与分析

本文在 4 个真实网络上对本文提出的 IDD1 算法和 IDD2 算法进行性能测试, 使用 IC 模型与一些启发式算法进行比较, 包括度折扣算法 (DD)、度中心性算法 (degree) 以及随机选择种子算法 (random)。本文没有与贪婪算法进行比较, 一方面是因为贪婪算法的时间复杂度太高, 对于规模上万的节点非常耗时。另一方面, 如文献 [10] 所述, 度折扣算法虽然是一种启发式算法, 但是该算法的性能接近于贪婪算法。实验结果表明了本文提出的 IDD1 算法和 IDD2 算法性能均优于度折扣算法, 因此更接近于贪婪算法。另外这两个算法均是基于局部网络结构的启发式算法, 与度折扣算法的时间复杂度接近。

#### 3.1 数据集与实验设置

本文使用的 4 个真实网络分别是 TAP、Ca-GrQc、PGP、Ca-CondMat。1) TAP 网络是通过串联亲和纯化实验生成的酵母蛋白结合网络 [28]; 2) Ca-GrQc 网络是广义相对论和量子宇宙学协作网络 [29]; 3) PGP 网络是一个加密的通信网络 [30]; 4) Ca-CondMat 网络是凝聚态物质科学协作网络 [29]。数据的基本结构信息如表 1 所示, 其中  $N$  是网络节点数,  $M$  是网络边数,  $\langle k \rangle$  是平均度值,  $\beta_{th}$  是网络中的传播阈值, 约等于  $\frac{\langle k \rangle}{\langle k^2 \rangle}$ 。

表 1 4 个经验网络的结构性质

网络	$N$	$M$	$\langle k \rangle$	$\beta_{th}$
TAP	1 373	6 833	9.950	0.061 1
Ca-GrQc	5 241	14 484	6.456	0.055 6
PGP	10 680	24 316	4.550	0.053 0
Ca-CondMat	23 133	93 439	8.078	0.045 3

本文使用 IC 模型作为传播模型, 每条边上的激活概率均为  $\beta$ , 从最终激活节点的比例、传播速度以及运行时间 3 方面比较不同算法的优劣性。为了保证结果的可靠性, 所有方法均独立传播 1 000

次取平均结果。

### 3.2 实验结果

#### 3.2.1 独立级联模型

本文在 4 个真实网络上比较了不同算法选择种子集的影响范围和传播速度, 并且定义了种子集的平均度和平均最短路径长度解释模型的有效性。

##### 1) 影响范围

图 2 显示了不同算法产生的最终传播范围随着种子数  $K$  的变化情况, 用  $\sigma(S)$  表示激活节点占网络的比例。前 3 个网络的传播率均设为 0.1, 由于第 4 个网络传播阈值较小, 传播率设置为 0.06。由图 2 可知: 度折扣算法比度中心性算法以及随机算法更能提高传播范围, 而本文提出的 IDD1 算法虽然只是对度折扣算法的公式进行简单修正, 但实验结果表明该算法在 4 个真实网络均优于度折扣算法。当考虑冗余弱化机制时, IDD2 算法能够更加明显地提高扩散范围。对于网络规模比较小的 TAP 网络而言, 在种子数较小时, IDD1 算法优势比较明显, 而对于较大的网络而言, IDD1 的优势在种子数较多时能更加凸显。

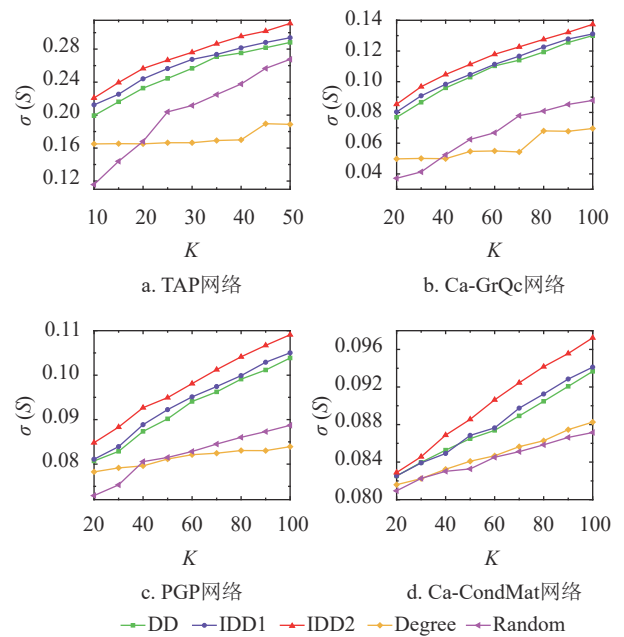


图 2 最终影响范围随着种子数的变化情况

图 3 是在固定种子数的情况下进一步研究最终传播范围随着不同传播概率  $\beta$  的变化情况, 其中 TAP 网络 20 个种子, Ca-GrQc 和 PGP 网络 50 个种子, Ca-CondMat 网络 70 个种子。实验结果再次表明, IDD1 算法比度折扣算法更能提高扩散范围, 而引入冗余弱化机制的 IDD2 算法则在所有网

络上都是效果最显著的。

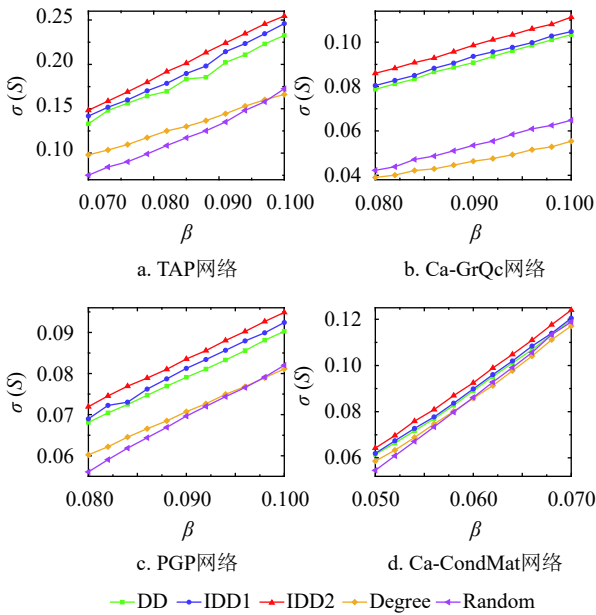


图3 最终影响范围随着传播率的变化情况

## 2) 传播速度

本文在固定种子数量以及传播率 $\beta$ 的情况下,比较不同算法对传播速度的影响,如图4所示。其中,TAP网络固定 $\beta=0.1$ , $K=20$ ;Ca-GrQc网络和PGP网络固定 $\beta=0.1$ , $K=50$ ;Ca-CondMat网络固定 $\beta=0.06$ , $K=70$ 。 $t$ 表示信息传播的迭代过程,用 $\sigma_t(S)$ 记录每次传播后网络中激活节点比例。从实验结果可以看出IDD1算法从一开始就以较快的速度扩散信息,并且最终传播范围也高于已有算法。

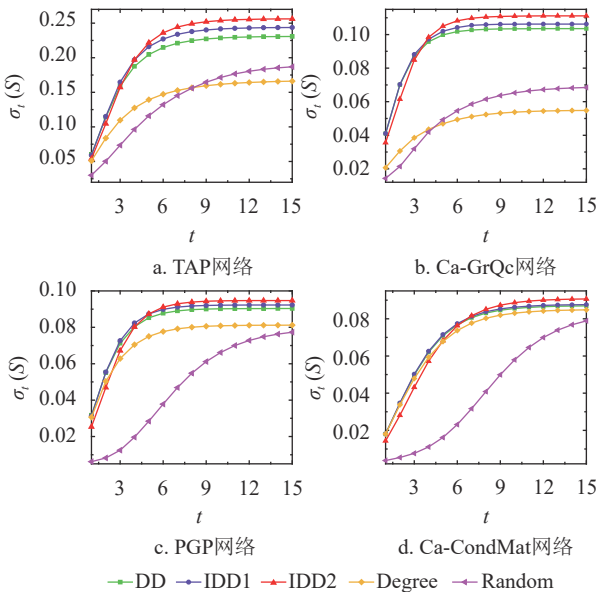


图4 影响范围随着时间的变化情况

IDD2算法在前几步传播过程中传播范围略低于度折扣算法和IDD1算法,这是由于IDD2算法选择的种子在网络中较为分散,因此需要经过一个短暂的酝酿过程。经过一小段时间后(3~6个时间步),IDD2算法的传播速度快速反超其他算法,并且得到的最终传播范围在所有算法中最广。从图2~图4可以发现,无论是最终传播范围还是传播速度,相比于其他算法,IDD1和IDD2算法都能取得更好的性能。

## 3) 种子集性质

为了解释为什么本文提出的两种算法能更有效地扩大传播范围,定义种子节点的平均度 $\langle d \rangle_S = \frac{1}{|S|} \sum_{i \in S} d_i$ ,其中 $|S|$ 是种子集大小。进一步定义种子节点之间的平均最短路径长度 $\langle L \rangle_S$ ,定义如下:

$$\langle L \rangle_S = \frac{1}{|S|(|S|-1)} \sum_{i,j \in S, i \neq j} L_{ij} \quad (6)$$

式中, $L_{ij}$ 是节点 $i$ 和节点 $j$ 的最短路径长度; $\langle L \rangle_S$ 指标越大,种子之间越分散。

此外,种子节点的平均度 $\langle d \rangle_S$ 随种子数的变化情况如图5所示,而种子节点间平均最短路径 $\langle L \rangle_S$ 随种子数的变化情况如图6所示。其中TAP网络、Ca-GrQc网络和PGP网络的传播率设为0.1,Ca-CondMat网络的传播率设为0.06。从图5和图6可以发现,虽然度中心性算法能保证种子节点的 $\langle d \rangle_S$ 最大,但是基于该算法得到的 $\langle L \rangle_S$ 值均小于度折扣算法以及IDD1和IDD2算法对应的值。反之,虽然随机算法能保证种子节点的 $\langle L \rangle_S$ 值最大,但是 $\langle d \rangle_S$ 值最小。度中心性算法和随机算法只能保证影响力最大化中两个关键因素中的某一条:1)种子节点自身很重要;2)种子节点足够分散地分布在网络上。度折扣算法以及改进算法得到的 $\langle d \rangle_S$ 值虽然小于度中心性算法对应的值,但是远高于随机算法得到的值,并且 $\langle L \rangle_S$ 大于度中心性算法对应的值。因而度折扣算法及改进算法平衡了种子节点自身的重要性和种子节点之间距离两个因素,可以更有效地扩大传播范围。尤为重要的是,虽然IDD1选择的种子节点的 $\langle d \rangle_S$ 值以及 $\langle L \rangle_S$ 值和度折扣算法对应的值都非常接近,但是该算法的传播范围更广、传播速度更快(如图2~图4)。因为IDD2算法充分考虑了冗余弱化机制,所以该算法得到的 $\langle L \rangle_S$ 值比其他非随机算法得到的值都大。当然,基于IDD2算法选出的种子节点的 $\langle d \rangle_S$ 值也一定程度

地变小, 但是远高于随机算法得到的  $\langle d \rangle_s$  值。IDD2 算法在 IC 模型中的最优性能也说明, 对于 IC 模型而言, 种子节点间的分散程度是能否保证影响力最大化的一个尤为重要的因素。

选出的种子节点仍包含了部分大度节点, IDD2 算法选取的种子节点则更加均匀地分散在网络中, 种子节点的度也相对减小。

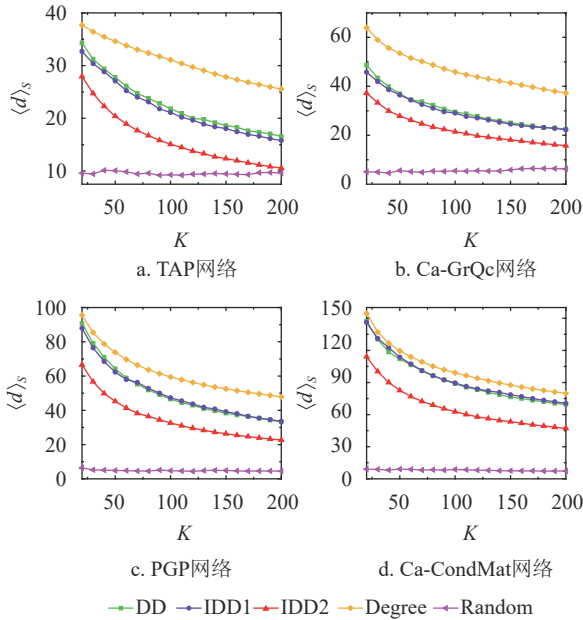


图5 种子节点间平均度随着种子数的变化情况

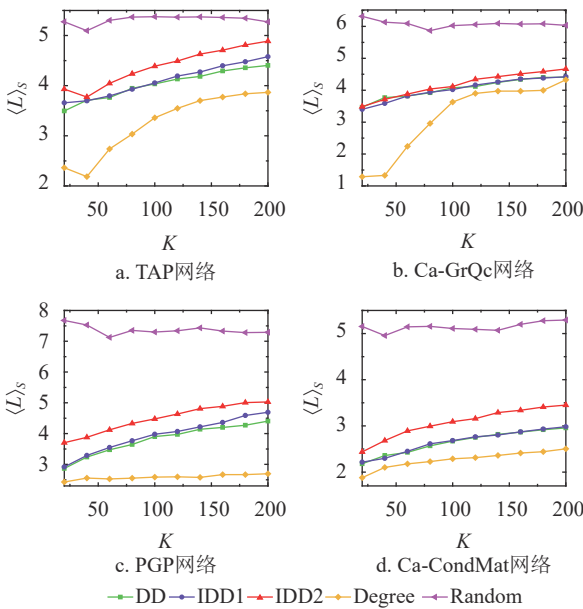
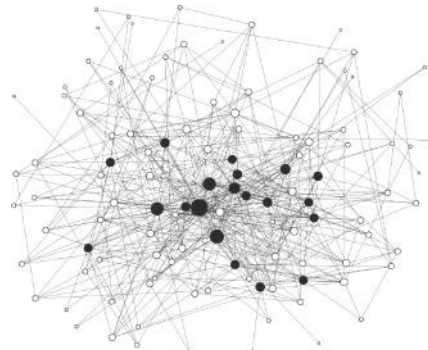
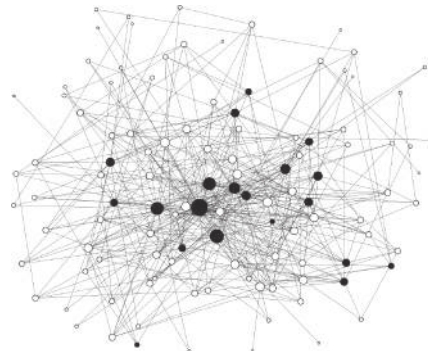


图6 种子节点间平均最短路径随着种子数的变化情况

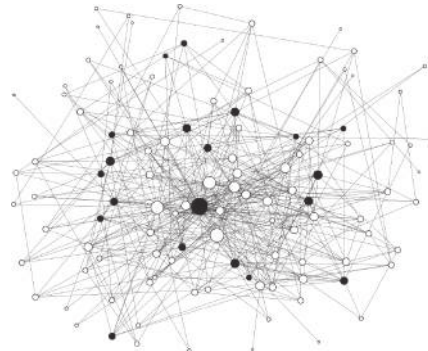
为了更为直观地表明改进算法选取出种子节点的特点, 本文对 Word 网络<sup>[31]</sup>进行可视化分析, 图7中节点的大小与度成正相关。分别使用度中心性算法、IDD1 算法和 IDD2 算法选取 20 个种子节点, 传播率设为 0.1。种子节点在网络中用实心黑点标记, 通过可视化可以直观地看出, IDD1 算法



a. Degree算法



b. IDD1算法



c. IDD2算法

图7 Word 网络种子节点可视化图

另外也测试了这些算法在线性阈值模型 (linear threshold model, LT) 中的性能, IDD1 算法效果仍优于度折扣算法, 然而 IDD2 算法在 LT 模型中效果一般, 这是由于当种子节点过于分散时, LT 模型的激活条件难以满足, 故不适用于 LT 模型。

### 3.2.2 未知传播率

在度折扣算法中, 作者假设节点真实传播率是已知的, 因而利用真实传播率计算节点的期望影响力, 但在实际情况下, 鉴于问题的复杂性, 很难知道真实传播率的大小。为了解决此问题, 一个科

学、折中的方法是用传播阈值近似代替真实传播率，这样做的合理性在于：传播率太小信息不能传播，而传播率太大时，任何一种方法都能使得信息大范围传播，此时再研究哪种方法能保证影响力最大没有太大意义。接下来的问题是：如果基于传播阈值计算式(3)~式(5)，IDD1算法和IDD2算法是否仍然具有更优异的表现。

图8显示了传播阈值近似代替传播率时最终影响范围。将前3个网络的真实传播率设为0.1，第4个网络的真实传播率设为0.06，而期望影响力是用 $\beta_{th}$ 代替 $\beta$ 计算得到的，然后比较了不同算法对最终影响范围 $\sigma(S)$ 的影响。从图8可以发现，对于IC模型而言，IDD1算法依然优于度折扣算法，IDD2算法在检查影响力最大化方面的性能更加突出。

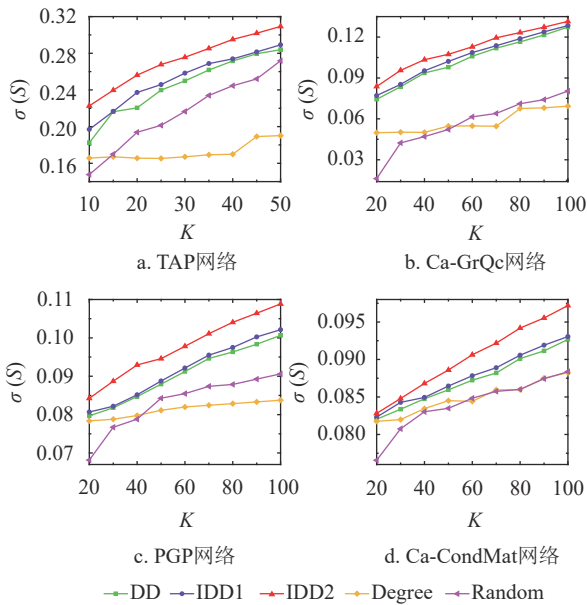


图8 传播阈值近似代替传播率时最终影响范围

### 3.2.3 敏感性分析

考虑IDD2算法中参数 $\alpha$ 的取值情况，本文进行了敏感性分析，其中TAP网络固定 $\beta=0.1$ ， $K=20$ ；Ca-GrQc网络和PGP网络固定 $\beta=0.1$ ， $K=50$ ，Ca-CondMat网络固定 $\beta=0.06$ ， $K=70$ 。针对不同的网络，如果网络较为稠密，则节点共同邻居数较多；若网络较为稀疏，共同邻居数很少。此时若 $\alpha$ 太大，指标将趋于0，若 $\alpha$ 太小，邻居的贡献差异性易被忽略，因此本文在 $[0,0.1]$ 的范围内观察IDD2算法的性能。当 $\alpha=0$ 时，IDD2算法退化为IDD1算法，即为IDD1算法的最终影响范围。从图9可以看出 $\alpha$ 取值在 $0.01\sim 0.1$ 时，IDD2算法效

果均优于IDD1算法，并且 $\alpha$ 取值为 $0.02\sim 0.1$ 时，效果相对稳定，因此本文将实验参数统一为0.05。

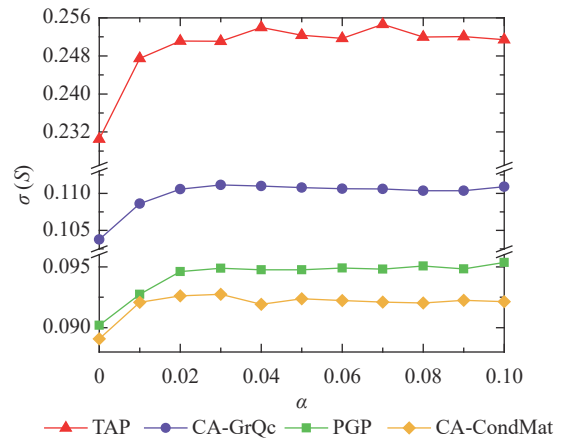
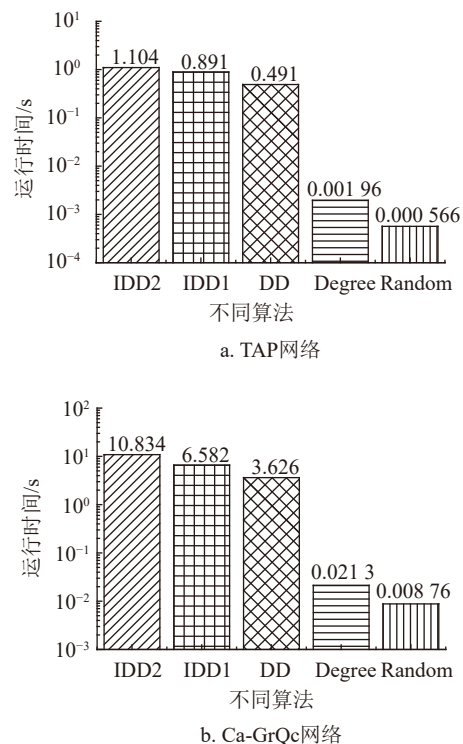


图9 参数 $\alpha$ 对IDD2算法的最终影响范围的影响

### 3.2.4 运行时间

图10比较了不同算法选择50个种子的运行时间，TAP网络、Ca-GrQc网络和PGP网络传播率设为0.1，Ca-CondMat网络设为0.06。由图可知，IDD1算法和IDD2算法的运行时间略高于度折扣算法。考虑冗余弱化机制的IDD2算法运行时间最长，这是因为在考虑节点度折扣时需要考虑每个邻居节点的种子邻居数，故复杂度要高于原度折扣算法，但度折扣算法相较于贪婪算法仍然具有高效性。



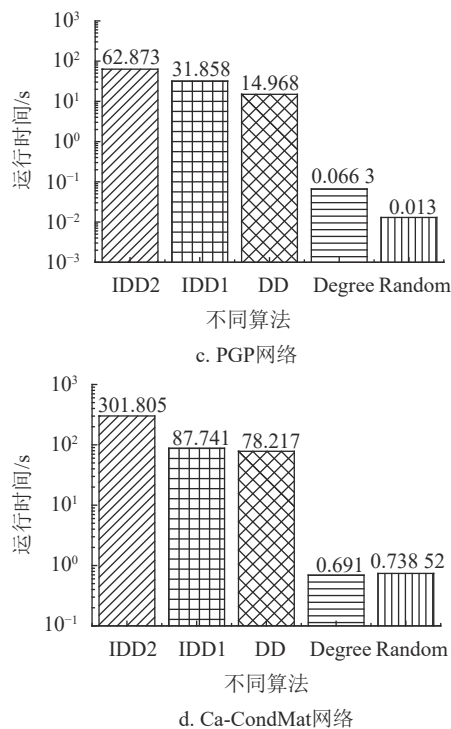


图10 不同算法选择种子节点的运行时间

## 4 结束语

研究社交网络影响力最大化问题具有重要的理论意义和应用价值, 在众多检测算法中, 由于度折扣算法具有快速、高效的性能, 受到学者们的广泛关注。考虑到度折扣算法在计算节点期望影响力时没有区别对待邻居中未激活节点的差异性, 本文对计算期望影响力的公式进行了修正, 得到了一阶改进的度折扣 (IDD1) 算法。为了进一步保证选择的种子节点充分分散地分布在网络上, 本文提出了二阶改进的度折扣 (IDD2) 算法。通过在 4 个真实网络上的实验结果表明, IDD1 算法影响的最终传播范围和传播速度均优于度折扣算法, 而 IDD2 算法在 IC 模型上扩大影响力范围的效果尤为突出。通过计算种子节点之间的平均度以及平均距离, 阐述了为何本文提出的算法能更有效地扩大传播范围。此外, 考虑到真实传播率难以获取, 本文利用传播阈值代替真实传播率计算期望影响力, 研究结果表明本文提出的算法依然有效。

## 参 考 文 献

- [1] 张静, 唐杰. 社会影响力分析综述[J]. 中国科学: 信息科学, 2017, 47: 967-979.  
ZHANG Jing, TANG Jie. Survey of social influence analysis and modeling[J]. *Sci Sin Inform*, 2017, 47: 967-979.
- [2] 周涛, 张子柯, 陈关荣, 等. 复杂网络研究的机遇与挑战[J]. 电子科技大学学报, 2014, 43(1): 1-5.  
ZHOU Tao, ZHANG Zi-ke, CHEN Guan-rong, et al. The opportunities and challenges of complex network research[J]. *Journal of University of Electronic Science and Technology of China*, 2014, 43(1): 1-5.
- [3] 陆豪放, 张千明, 周莹, 等. 微博中的信息传播: 媒体效应与社交影响[J]. 电子科技大学学报, 2014, 43(2): 167-173.  
LU Hao-fang, ZHANG Qian-ming, ZHOU Ying. Information spreading in microblogging systems: Media effect versus social impact[J]. *Journal of University of Electronic Science and Technology of China*, 2014, 43(2): 167-173.
- [4] LÜ Lin-yuan, MEDO M, YEUNG C H, et al. Recommender systems[J]. *Physics Reports*, 2012, 519(1): 1-49.
- [5] 周涛, 汪秉宏, 韩晓璞, 等. 社会网络分析及其在舆情和疫情防控中的应用[J]. 系统工程学报, 2010, 25(6): 742-754.  
ZHOU Tao, WANG Bing-hong, HAN Xiao-pu, et al. Social network analysis and its application in the prevention and control of propagation for public opinion and the epidemic[J]. *Journal of Systems Engineering*, 2010, 25(6): 742-754.
- [6] 鲍中奎, 张海峰. 二维 Kleinberg 网络上疾病传播的最优局部控制策略[J]. 电子科技大学学报, 2016, 45(3): 475-480.  
BAO Zhong-kui, ZHANG Hai-feng. Optimal local control strategy for the spreading of epidemic in two-dimensional Kleinberg networks[J]. *Journal of University of Electronic Science and Technology of China*, 2016, 45(3): 475-480.
- [7] 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*. New York: ACM, 2003: 137-146.
- [8] LESKOVEC J, KRAUSE A, GUESTRIN C, et al. Cost-effective outbreak detection in networks[C]// *Proceedings of the 13th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2007: 420-429.
- [9] GOYAL A, LU W, LAKSHMANAN L V S. CELF++: Optimizing the greedy algorithm for influence maximization in social networks[C]// *Proceedings of the 20th International World Wide Web Conference*. Hyderabad, India: [s.n.], 2011: 47-48.
- [10] CHEN Wei, WANG Ya-jun, YANG Si-yu. Efficient influence maximization in social networks[C]// *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Paris: ACM, 2009: 199-208.
- [11] CHENG Su-qi, SHEN Hua-wei, HUANG Jun-ming, et al. Staticgreedy: Solving the apparent scalability-accuracy dilemma in influence maximization[C]// *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management*. San Francisco: ACM, 2013: 509-518.
- [12] 朱军芳, 陈端兵, 周涛, 等. 网络科学中相对重要节点挖



- 掘方法综述[J]. 电子科技大学学报, 2019, 48(4): 595-603.
- ZHU Jun-fang, CHEN Duan-bin, ZHOU Tao, et al. A survey on mining relatively important nodes in network science[J]. Journal of University of Electronic Science and Technology of China, 2019, 48(4): 595-603.
- [13] GHOSH R, LERMAN K. Predicting influential users in online social networks[C]//Proceedings of the 4th KDD Workshop on Social Network Analysis. Washington: ACM, 2010: 3-5.
- [14] SABIDUSSI G. The centrality index of a graph[J]. *Psychometrika*, 1966, 31(4): 581-603.
- [15] FREEMAN L C. A set of measures of centrality based on betweenness[J]. *Sociometry*, 1977, 40(1): 35-41.
- [16] KITSAK M, GALLOS L K, HAVLIN S, et al. Identification of influential spreaders in complex networks[J]. *Nature Physics*, 2010, 6(11): 888-893.
- [17] LÜ Lin-yuan, CHEN Duan-bing, REN Xiao-long, et al. Vital nodes identification in complex networks[J]. *Physics Reports*, 2016, 650: 1-63.
- [18] LIU Huan-li, MA Chuang, XIANG Bing-bing, et al. Identifying multiple influential spreaders based on generalized closeness centrality[J]. *Physica A: Statistical Mechanics and its Applications*, 2018, 492: 2237-2248.
- [19] KORN A, SCHUBERT A, TELCS A. Lobby index in networks[J]. *Physica A: Statistical Mechanics and Its Applications*, 2009, 388(11): 2221-2226.
- [20] LÜ Lin-yuan, ZHOU Tao, ZHANG Qian-ming, et al. The H-index of a network node and its relation to degree and coreness[J]. *Nature Communications*, 2016, 7: 10168.
- [21] MA Ling-ling, MA Chuang, ZHANG Hai-feng, et al. Identifying influential spreaders in complex networks based on gravity formula[J]. *Physica A*, 2016, 451: 205-212.
- [22] JUNG K, HEO W, CHEN W. IRIE: Scalable and robust influence maximization in social networks[C]//Proceedings of the 12th IEEE International Conference on Data Mining, Washington: IEEE, 2012: 918-923.
- [23] GOYAL A, LU W, LAKSHMANAN L V S. Simpath: An efficient algorithm for influence maximization under the linear threshold model[C]//Proceedings of the 11th IEEE International Conference on Data Mining. Italy: IEEE, 2011: 211-220.
- [24] PEI S, MAKSE H A. Spreading dynamics in complex networks[J]. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(12): P12002.
- [25] KIMURA M, SAITO K. Tractable models for information diffusion in social networks[C]//Proceedings of the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases. Berlin: [s.n.], 2006: 259-271.
- [26] BAO Zhong-kui, LIU Jian-guo, ZHANG Hai-feng. Identifying multiple influential spreaders by a heuristic clustering algorithm[J]. *Phys Lett A*, 2017, 381: 976.
- [27] ZHANG Ya-ming, SU Yan-yuan, LI Wei-gang, et al. Identifying multiple influential spreaders with local relative weakening effect in complex networks[J]. *Europhys Lett*, 2018, 124(2): 28001.
- [28] GAVIN A C, BÖSCHE M, KRAUSE R, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes[J]. *Nature*, 2002, 415(6868): 141-147.
- [29] LESKOVEC J, KLEINBERG J, FALOUTSOS C. Graph evolution: Densification and shrinking diameters[J]. *ACM Transactions on Knowledge Discovery from Data*, 2007, 1(1): 1-41.
- [30] BOGUNÁ M, PASTOR-SATORRAS R, DÍAZ-GUILERA A, et al. Models of social networks based on social distance attachment[J]. *Physical Review E*, 2004, 70(5): 056122.
- [31] NEWMAN M E J. Finding community structure in networks using the eigenvectors of matrices[J]. *Physical Review E*, 2006, 74(3): 036104.

编辑 蒋晓