

DOI: 10.11992/tis.201906039

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.tp.20190722.1416.004.html>

## 基于渗流模型的影响力最大化算法

花勇, 陈伯伦, 朱国畅, 袁燕, 金鹰

(淮阴工学院 计算机与软件工程学院, 江苏 淮安 223003)

**摘要:**多数社交网络影响力最大化算法的研究只关注于所选种子节点集合的影响力是否最优, 忽略网络自身传播影响力的固有能力。本文对网络进行渗流模拟, 计算渗流后网络的主连通分量随着传播概率改变的趋势, 并且求得主连通分量大小增加开始变快的相变点, 从而计算网络自身传播影响力的固有能力。通过相变值与种子节点集合大小的换算, 求得当前网络最佳的种子节点集合大小。将种子节点集合大小限制在最佳大小范围内即可获得最佳的影响力。在 karateclub、football、highschool 和 socdolphins 社交网络数据集上进行实验, 验证了该方法的有效性。

**关键词:**社交网络; 影响力最大化; 种子节点集合; 渗流; 传播概率; 主连通分量; 相变点; 相变值

**中图分类号:** TP301.6 **文献标志码:** A **文章编号:** 1673-4785(2019)06-1262-09

中文引用格式: 花勇, 陈伯伦, 朱国畅, 等. 基于渗流模型的影响力最大化算法 [J]. 智能系统学报, 2019, 14(6): 1262-1270.

英文引用格式: HUA Yong, CHEN Bolun, ZHU Guochang, et al. An influence maximization algorithm based on percolation model[J]. CAAI transactions on intelligent systems, 2019, 14(6): 1262-1270.

## An influence maximization algorithm based on percolation model

HUA Yong, CHEN Bolun, ZHU Guochang, YUAN Yan, JIN Yin

(School of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian 223003, China)

**Abstract:** Most of the influence maximization algorithms in social networks only focus on whether the influence of the seed node set selected is the optimal, and ignore the inherent ability of social network's propagating influence. Using percolation simulation, we calculate the change trend of the giant component of the network generation after percolation with propagation probability, and derive the starting point at which the size of the giant component increases fastest, that is, the phase point. The phase value shows the inherent ability of the network propagating influence. The optimal seed set size of the network can be calculated through conversion of the phase value and the size of the seed set. We can obtain the optimal influence by limiting the size of the seed set to the optimal size. We performed experiments on karate club, football, high school, and soc-dolphins, verifying the effectiveness of the algorithm.

**Keywords:** social network; influence maximization; seed set; percolation; propagation probability; giant component; phase point; phase value

随着社交网络中信息量的快速增长, 信息传播速度的不断加快, 其信息传播构建了一种分布式传播机制<sup>[1]</sup>以及节点的合作机制<sup>[2]</sup>, 即信息在用户之间的扩散会受到用户影响力的影响<sup>[3]</sup>。因此, 开展影响力分析研究显得十分重要。影响力

最大化问题是影响力分析的重要课题之一。2015 年, Morone 和 Makse 在 Nature 上对社交网络中影响力最大化问题进行了深入探讨<sup>[4]</sup>。影响力最大化问题解决的是如何衡量网络中节点重要性的问题, 其经典应用之一是病毒营销<sup>[5-8]</sup>, 也就是通过口口相传效应进行产品的销售<sup>[9-10]</sup>。

影响力最大化问题最早由 Kempe 等<sup>[11]</sup>率先提出。Kempe 等使用独立级联模型与线性阈值模型对社交网络中影响力的传播进行建模, 并且证

收稿日期: 2019-06-21. 网络出版日期: 2019-07-23.

基金项目: 国家自然科学基金项目 (61602202); 江苏省自然科学基金项目 (BK20160428).

通信作者: 陈伯伦. E-mail: [chenbolun1986@163.com](mailto:chenbolun1986@163.com).

明在社交网络中寻找具有最佳影响力的种子节点集合是 NP-Hard 问题。而且他们提出使用简单的贪婪算法寻找具有最佳影响力的种子节点集合,获得了  $(1-1/e)$  的近似保证。影响力最大化问题中的关键问题是如何衡量节点传播影响力的能力,也就是节点所具有的影响力。在最初的影响力最大化问题研究中,一些基于网络拓扑结构的中心性方法被提出,例如度中心性和点介数中心性。在随后的影响力最大化问题研究中,多数算法使用次模性质<sup>[12-15]</sup>,即通过大量迭代来计算边际收益,从而近似求得问题的最优解,但是存在算法时间复杂度较高的问题。Chen 等<sup>[16]</sup>提出了度折损启发式算法,即优先选择度大的节点作为种子节点,一旦某节点被选取为种子节点,那么其邻居节点被选为种子节点的概率将大大降低。Lee 等<sup>[17]</sup>利用多数节点平均只能影响到 2 阶邻居节点的现象,提出了 2-hop 贪婪算法,即利用节点在 2 阶邻居范围内的影响力更新节点的边际收益。Chen 等<sup>[16]</sup>提出 MIA(maximum influence arborescence)最大影响力树算法,利用 MIIA(maximum influence in-arborescence)影响力最大入树和 MIOA(maximum out-arborescence)影响力最大出树构建节点的影响力传播路径,通过节点的影响力最大化路径计算得分并选取种子节点,相较于原始的贪婪算法取得了较低的时间复杂度,但因为 MIA 需要对每个节点建立树,所以空间复杂度相较于其他算法较高。JUNK 等<sup>[18]</sup>提出 IRIE(influence ranking influence estimation)算法,即算法整体被分为两部分,影响力排名使用一种算法,影响力模拟再使用一种算法。在最近影响力最大化问题的研究中,许多优秀的算法被提出,其中 Kitsak 等<sup>[19]</sup>提出最具影响力的节点往往不是具有较大连接的节点,而是处于网络核心位置的节点,通过 k-shell 分解<sup>[20]</sup>分析网络中节点核数与节点影响力的关系,为影响力最大化问题提供了新的解决思路。Gao 等<sup>[21]</sup>根据节点的影响力与其局部结构的关系,提出了一种局部结构中心性的方法,利用节点以及其邻居的拓扑结构和中心性来衡量节点的影响力,此算法在评估节点影响力方面更加准确。王等<sup>[22]</sup>提出的多种群随机差分粒子群优化算法和刘等<sup>[23]</sup>提出的改进萤火虫算法也可很好的应用到影响力最大化问题当中。

在上述的影响力最大化算法中,研究人员只关注所选种子节点影响力是否最佳,旨在研究更加优秀的算法近似求得影响力最优的种子节点集合,而忽略了种子节点集合的大小和网络固有的传播影响力能力的关系。本文从网络传播影响力

能力的角度出发研究影响力最大化问题,从而得出网络所适合的种子节点集合的大小。本文提出种子节点集合大小并不是越大越好,而是每个网络都存在一个种子节点个数的上限,一旦超过这个上限,随着种子节点个数大小的增加,种子节点集合的影响力是趋于饱和的。为研究网络传播影响力的固有能力,本文使用渗流<sup>[24-26]</sup>的思想,即对网络进行渗流模拟,得出网络由大量零散的团块趋向于形成一个主团块的相变值,从而得出网络所适合的种子节点集合的大小,即提出一种基于渗流模型的影响力最大化种子节点集合大小选取算法来获得最优的种子节点集合大小,并对算法结果进行了分析。

## 1 问题描述

### 1.1 影响力最大化问题

本文主要研究无向网络传播影响力的能力,定义无向网络  $G=(V,E)$ ,其中  $V$  为无向网络  $G$  中的节点集合, $E$  为无向网络  $G$  中边的集合。定义  $n=|V|$  为网络  $G$  的节点个数, $m=|E|$  为网络  $G$  边的个数。影响力最大化问题就是在网络  $G$  中寻找大小为  $k$  的种子节点集合,使得这  $k$  个种子节点在网络  $G$  中传播的影响力是最大的,我们定义种子节点集合为  $S$ 。在影响力最大化问题中,我们面临着两个尤为重要的问题,即影响力是如何定义的,以及影响力在网络中是如何传播的。独立级联模型(independent cascade model)是经典的模拟影响力传播的模型,在之前多数影响力最大化问题的研究中,研究人员都使用独立级联模型作为影响力传播模型。而节点或者节点集合的影响力,我们使用影响力函数进行求解。

### 1.2 独立级联模型

独立级联模型<sup>[27-28]</sup>是经典的用于模拟影响力传播的算法,模型描述如下:在网络  $G$  中,节点集合  $V$  中的节点  $v$  存在两种状态,即一种是激活状态另一种是未激活状态。假设  $t$  时刻,节点  $v$  已经处于激活状态,那么节点  $v$  会尝试以概率  $p$  去激活其邻居节点  $u$ ,如果激活成功,那么节点  $u$  会从  $t+1$  时刻开始,一直处于激活状态。如果激活失败,那么从  $t+1$  时刻开始,节点  $u$  再也不能被节点  $v$  尝试激活。我们定义影响力传播模型的初始时刻为  $t=0$  时刻,定义集合  $A_0$  中的节点处于激活状态,那么集合  $A_t$  为  $t$  时刻被激活的节点集合。如果存在某一时刻  $c+1$ ,集合  $A_{c+1}$  为空集,那么独立级联模型终止运行。当独立级联模型终止时,我们可以获得在此过程中激活的节点集合  $A_{\text{total}} = \{A_0, A_1, A_2, \dots, A_c\}$ 。

### 1.3 影响力函数

定义影响力函数为  $I(x)$ : 将有限集合映射到非负整数域上的函数。在网络中使用独立级联模型模拟影响力传播的过程当中, 种子集合  $S$  为初始时刻已经激活的节点集合, 在影响力传播过程的每个离散时刻都会因集合  $S$  激活一些未激活的节点, 在影响力传播过程结束时我们可以得到在此过程中激活的节点集合  $A_{\text{total}}$ , 即集合  $A_{\text{total}}$  是被种子节点集合  $S$  影响的节点集合。我们令种子节点集合  $S$  的影响力为  $I(S)$ , 其值为  $|A_{\text{total}}|$ , 即种子节点集合  $S$  的影响力是在影响力传播过程中被集合  $S$  影响到的节点的个数。

## 2 算法思想与步骤

本文主要研究影响力最大化问题中种子节点集合大小选取的问题。在之前的研究中, 研究人员一般选取 5~50 个节点作为种子节点, 并观察算法选取出的种子节点集合的影响力, 旨在通过改进算法得到更优的种子节点集合。而本文从网络传播影响力的固有能力的角度出发, 发现网络在选取种子节点时, 并非越多越好, 而是一定数量的种子节点就能达到最优的影响力, 即在网络中存在一个最优大小的种子节点集合, 即我们所说的网络传播影响力的固有能力的角度。为了研究网络传播影响力的固有能力的角度, 本文借助渗流模型对网络进行模拟分析, 提出一种基于渗流模型的影响力最大化算法, 即在不同的传播概率  $p$  下对网络进行渗流模拟, 通过建立传播概率  $p$  与渗流模拟后网络的最大连通子图大小的函数关系, 最终求得当前网络所适合的种子节点集合的大小。具体算法步骤如下:

**算法** 基于渗流模型的影响力最大化种子节点集合大小选取算法

**输入** 上三角邻接矩阵  $G'$ , 传播概率数组  $p_{\text{list}}$ , 网络中节点的数量  $n$ , 矩阵  $C$ , 模拟次数  $R$ ;

**输出** 最优种子节点集合大小  $k'$

- 1) For  $i=1:1:\text{len}(p_{\text{list}})$
- 2) For  $j=1:1:R$
- 3) 根据  $p_{\text{list}}(i)$  对  $G'$  进行渗流模拟, 形成渗流后网络  $GP$ , 并且获得  $GP$  的最大连通子图  $GP'$ ;
- 4)  $C(j,i)=\text{num}(GP')$ ;
- End
- End
- 5)  $C'=\text{avg}(C)$ ;

6) 对  $p_{\text{list}}$  和  $C$  进行多项式拟合, 求得拟合函数  $F(x)$ ;

7) 求  $F(x)$  的导函数  $dF(x)$ ;

8) 通过函数  $dF(x)$  求得相变值  $p_c$ ;

9) 最优种子集合大小  $k' = p_c \times n$

本文提出一种基于渗流模型的影响力最大化种子节点集合大小选取算法。算法的输入为: 无向网络  $G$  的上三角邻接矩阵  $G'$ , 传播概率数组  $p_{\text{list}}$ , 网络中节点的数量  $n$ , 矩阵  $C$  和模拟次数  $R$ 。因为本文主要研究无向网络, 在对网络进行渗流模拟时, 为了保持边的一致性, 所以使用网络  $G$  的上三角邻接矩阵  $G'$ 。  $p_{\text{list}}$  是大小为  $1 \times 1000$  的一维数组, 其中数组元素为  $0.001 \sim 1$  的数字, 且相邻元素相差  $0.001$ 。  $C$  是大小为  $100 \times 1000$  的矩阵。因为在网络  $G$  上进行渗流的结果具有随机性, 所以我们在传播概率  $p \in p_{\text{list}}$  的情况下进行  $R$  次渗流。算法的输出为最优种子节点集合大小  $k'$ 。具体的算法步骤描述如下:

1) 函数  $\text{len}(x)$  用来计算数组的长度, 所以  $\text{len}(p_{\text{list}})$  的值等于  $1000$ , 此步骤具体是: 在传播概率  $p \in p_{\text{list}}$  的情况下, 对网络  $G'$  进行渗流;

2) 采用当前传播概率  $p$  对网络  $G'$  进行  $R$  次渗流;

3) 渗流模型的定义如下: 在网络  $G'$  中, 网络每条边具有统一的传播概率值  $p$ 。我们对每条边产生独立的随机值  $p_r$ , 如果  $p_r < p$ , 那么此边处于占有状态, 如果  $p_r > p$ , 那么此边处于非占有状态, 也就是此边从网络中删除。通过改变统一的传播概率值  $p$ , 那么存在一个值  $p_c$ , 当  $p > p_c$  时,  $GP$  中的节点倾向于紧密的连接在一起, 形成一个主团块。当  $p < p_c$  时,  $GP$  中的节点倾向于形成多个零散的小团块。  $p_c$  便称为网络  $G$  的相变值。图 1 为渗流实验例图, 图 1(a) 为 2 维晶格网络, 网络中边的概率统一为  $p$ 。图 1(b) 到图 1(d) 为渗流模拟后的网络, 并且边的概率逐渐提高, 我们发现当边的概率较低时, 渗流模拟后的网络倾向于形成零散的团块, 当边的概率较大时, 渗流模拟后的网络倾向于形成一个主团块, 其存在一个过度值, 即网络由零散的团块到形成一个主团块的值, 即相变值。图 1(d) 为 2 维晶格网络发生渗流时的状态, 相变值  $p_c=0.6$ 。在 3) 中, 通过对网络  $G'$  进行渗流模拟, 可以得到渗流后的网络  $GP$  并且计算求得  $GP$  的最大连通子图  $GP'$ ;

4) 函数  $\text{num}(x)$  用于计算网络中节点的个数, 所以此步骤是将  $GP'$  的节点个数存入到  $C$  中;



5) 函数  $\text{avg}(x)$  用于计算矩阵每列的平均值, 所以此步骤是对传播概率  $p$  下  $R$  个  $\text{num}(\text{GP}')$  求平均值, 并且存入到  $\mathbf{C}'$  中, 也就是说, 每个传播概率  $p$ , 我们都会对  $\mathbf{G}'$  进行  $R$  次渗流模拟, 从而产生  $R$  个  $\text{num}(\text{GP}')$ , 最后将其求平均, 即得到每个传播概率  $p$  所对应  $\text{GP}'$  的平均大小;

6) 采用多项式拟合的方法对  $p_{\text{list}}$  和  $\mathbf{C}'$  进行拟合。多项式拟合是使用多项展开式去近似数据点的函数关系, 并使用最小二乘法来得到多项展开式的系数, 最终求得数据点函数关系的方法。多项式拟合公式为

$$F(x) = a_0 + a_1x + a_2x^2 + \cdots + a_lx^l \quad (1)$$

式中:  $a_0$  到  $a_l$  为使用最小二乘法求取的系数;  $l$  为多项式的阶数。本文式 (1) 中的  $x$  为  $p_{\text{list}}$  中的元素,  $\mathbf{C}'$  中的元素为  $F(x)$  的函数值, 通过多项式拟合函数求得  $F(x)$  的系数, 从而求得  $p_{\text{list}}$  和  $\mathbf{C}'$  的拟合函数  $F(x)$ ;

7) 求函数  $F(x)$  的导函数  $dF(x)$ ;

8) 相变值  $p_c$  在  $p_m$  的左邻域中, 其中  $dF(p_m)$  的值为函数  $dF(x)$  的最大值,  $dF(p_c)$  的值为靠近  $dF(p_m)$  最近的最小值, 相变值  $p_c$  为函数  $F(x)$  变化速率开始加快的起点位置;

9) 求取最优种子集合大小。

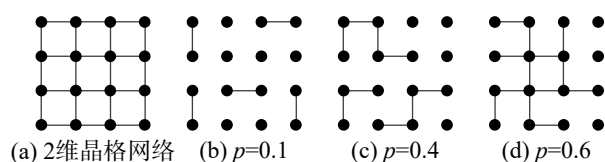


图1 渗流实验例图

Fig. 1 The example of percolation experiment

### 3 实验结果及分析

本文提出的一种基于渗流模型的影响力最大化种子节点集合大小选取算法在4个公共数据集上进行实验, 数据集分别为 KarateClub<sup>[29]</sup>、Football<sup>[30]</sup>、HighSchool<sup>[31]</sup> 和 SocDolphins<sup>[32]</sup>。因本文主要研究无向网络, 所以必要时对数据集进行了无向化处理。KarateClub 数据集是 1970 年美国大学生空手道俱乐部 34 名成员之间朋友关系的社交网络。Football 是 2000 年美式足球秋季常规赛大学之间的比赛网络。HighSchool 是 2013 年 12 月法国马赛高中学生友谊联系的社交网络。SocDolphins 是宽吻海豚之间的社交网络。其中数据集 KarateClub、HighSchool 和 SocDolphins 中的边表示成员之间拥有相对频繁的联系, Football 数据集中的边表示球队之间会有比赛安排。不同数据集的拓扑属性如表 1 中所示, 其中节点数为网络

中节点的总数, 边数为网络中边的总数, 最大度数为网络中边数的最大值, 平均度为度的平均值。同配系数是描述大度节点之间相连接的能力, 其值越靠近 1 说明其同配性越好; 聚类系数是描述节点之间连接成团的能力, 其值越大说明网络中的节点更有可能产生连接; 网络密度描述了网络实际存在边数与网络可容纳边数的比值, 也就是节点之间相互连边的密集程度, 其值越大说明网络越密集。

表1 数据集属性

Table 1 The attributes of datasets

数据集	节点数	边数	最大 度数	平均 度数	同配 系数	聚类 系数	网络 密度
KarateClub	34	78	17	4.59	-0.48	0.57	0.14
Football	115	613	12	10.66	0.16	0.40	0.09
HighSchool	134	406	17	6.06	0.29	0.54	0.05
SocDolphins	62	159	12	5.13	-0.04	0.26	0.08

在现有影响力最大化问题的研究中, 大多数研究人员主要关注如何在网络中选取具有最佳影响力的种子节点集合, 也就是通过研究创造出更先进的影响力最大化算法来近似选取种子节点集合, 并不关注种子节点集合大小的问题, 即网络传播影响力的固有能力的的问题。本文主要研究网络传播影响力的能力, 提出网络传播影响力的能力是有限的, 也就是在网络中选择种子节点的时候并不是越多越好, 每个网络存在一个最优的种子集合大小, 一旦种子集合大小超过了最优值, 其多出的种子节点所带来的影响力几乎不能起到积极的作用, 反而会增加实验的成本。因为本文主要研究种子节点集合的大小, 所以需要获得种子节点的算法作为载体来求得种子节点, 本文使用 4 种经典的算法来选取种子节点, 4 种算法分别为: 贪婪算法、度中心性、点介数中心性和基于  $k$  核过滤核覆盖算法。使用上述方法分别选出 4 个数据集具有最佳影响力的 10 种子节点, 并对其影响力做出分析。

#### 3.1 渗流实验

我们提出网络传播影响力的固有能力与相变值  $p_c$  有关, 所以在网络  $G$  上进行渗流实验。通过改变传播概率  $p$ , 对网络  $G$  进行多次渗流实验, 建立传播概率  $p$  与渗流后网络  $\text{GP}$  的最大连通子图  $\text{GP}'$  平均大小  $s$  的函数关系。在具体实验中, 设定传播概率  $p$  为 0.001~1 的数, 且为 0.001 的倍数, 也就是说传播概率  $p$  有 1 000 种情况。然后根据

不同的传播概率  $p$  对网络进行渗流实验, 并且每个传播概率  $p$  进行  $R$  次独立的渗流实验, 因为渗流实验具有随机性, 本文中通过设置较高的  $R$  值来获取足够的实验结果, 本文设置  $R=1\,000$ 。渗流实验后, 计算得到渗流后网络 GP 的最大连通子图 GP' 的平均大小  $s$ , 通过多项式拟合的方法对  $p$  和  $s$  进行拟合, 形成  $p$  和  $s$  的拟合函数  $F(x)$ 。渗流模拟实验具体结果如图 2 所示。在图 2 中,  $p$  表示网络的传播概率,  $s$  表示每个传播概率  $p$  下

$R$  次渗流模拟后所得的最大连通子图大小均值。图 2 中蓝色部分是由 1 000 个点构成的散点图, 每个点对应了一个传播概率  $p$  以及一个最大连通子图大小均值  $s$ 。图中红色曲线是对  $p$  和  $s$  进行拟合得到的拟合函数  $F(x)$  的曲线。由图 2 发现随着  $p$  值的增大, 曲线逐渐平缓, 在  $p$  值较小时,  $s$  的增长速率较大。即  $p$  值较小时, GP 由零散的小团块组成, 当  $p$  值越来越大时, GP 趋向由主团块组成。

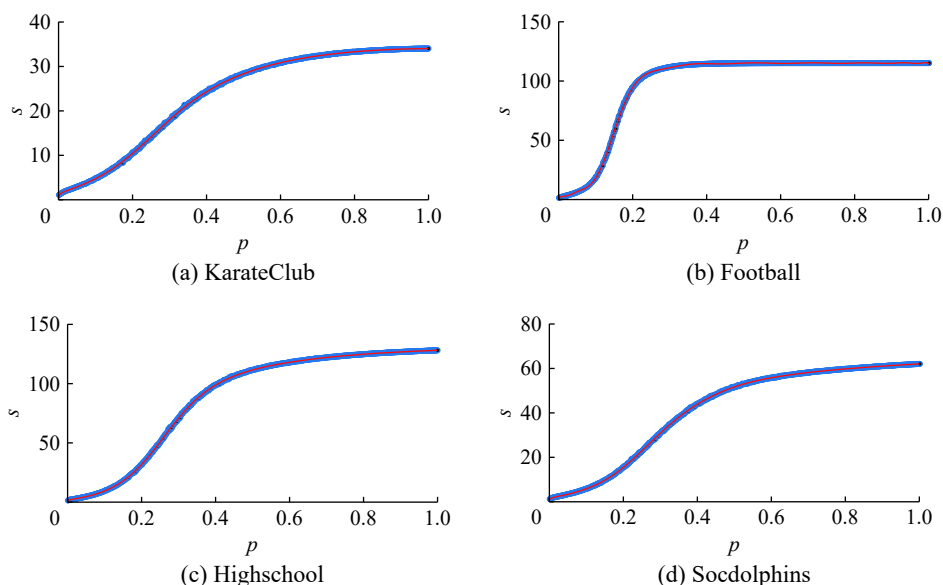


图 2 渗流实验

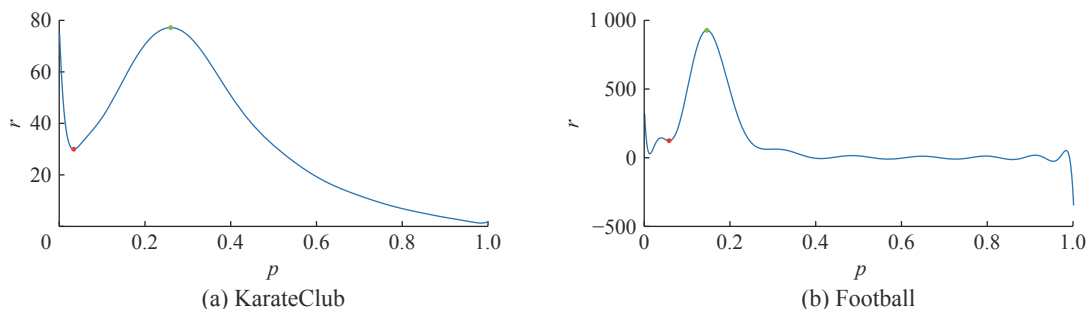
Fig. 2 The percolation experiment

### 3.2 相变值

为了计算网络  $G$  的相变值  $p_c$ , 需要计算函数  $F(x)$  的变化速率。在图 3 中,  $p$  为传播概率,  $r$  为函数  $F(x)$  的变化速率, 即函数  $dF(x)$  的值。我们可以得到函数  $dF(x)$  的最大值  $dF(p_m)$ , 点  $p_m$  为图 3 中绿色的点, 也就是函数  $F(x)$  变化最快的时候。所有找的相变点  $p_c$ , 在  $p_m$  的左邻域中, 也就是图 3 中红色的点, 其中  $dF(p_c)$  为距离  $dF(p_m)$  最近的最小值, 也就是函数  $dF(x)$  变化增长到最快时的起点位置。当网络的传播概率  $p$  小于相变值  $p_c$  时, 变化速率  $r$  还处于较低水平, GP 由零散的小团块组成, 当传播概率  $p$  大于  $p_c$  时, GP 趋向

由主团块组成, GP 逐渐呈现出以最大连通子图为主的图结构。本文提出相变值  $p_c$  反应了网络  $G$  传播影响力的固有能力, 也就是相变值反应网络  $G$  中边被激活的能力, 即在影响力传播模型下, 被激活边占总边数的比例。因此可以得到网络  $G$  最优的种子节点集合的大小  $k'$ 。

因此可以计算出 4 个数据集的相变值与最优的种子节点集合的大小, 其中 KarateClub、Football、HighSchool 和 SocDolphins 的相变值分别为 0.034、0.059、0.022 和 0.029。KarateClub、Football、HighSchool 和 SocDolphins 的最优的种子节点集合的大小分别为 2、7、3 和 2。



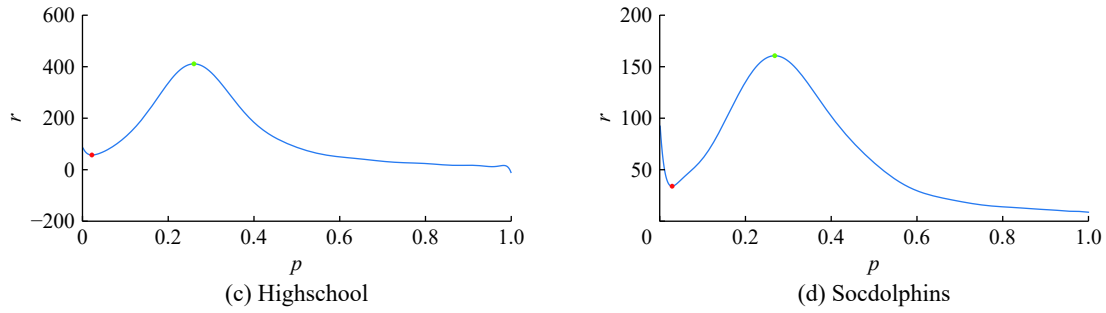


图3 拟合函数变化速率

Fig. 3 The changing rate of fitting function

### 3.3 影响力模拟

本文使用4种影响力最大化算法来选取种子节点集合,4种算法分别是:简单贪婪算法<sup>[11]</sup>、度中心性、点介数中心性<sup>[33]</sup>以及基于 $k$ 核过滤核覆盖算法<sup>[34]</sup>。其中,简单贪婪算法通过迭代的方式逐节点计算 $I(S \cup \{v\})$ ,并在每轮迭代中将使函数值最大的节点 $v$ 加入到种子节点集合 $S$ 中,直到选满 $k$ 个种子节点,迭代结束。度中心性和点介数中心性则选择度最大的 $k$ 个节点作为种子节点。基于 $k$ 核过滤核覆盖算法则是通过预先计算出最优的核数 $k_{opt}$ ,通过 $k$ 核分解出最小核数为 $k_{opt}$ 的子图,在子图中选择核数最大的节点作为种子节点。

在影响力模拟实验中,我们使用独立级联模型作为影响力模拟算法以及使用 $I(x)$ 计算影响力,并且使用上述4种算法选取影响力最大的10个种子节点,分别对1~10大小种子节点集合

进行影响力模拟。4个数据集种子节点集合影响力实验结果如图4所示。在图4中, $k$ 为种子节点个数, $I(k)$ 为种子节点集合的影响力。图4(a)为KarateClub数据集种子节点集合影响力的实验结果,在图中我们可以发现4种算法选出的种子节点集合在大小为3时,影响力大小几乎趋于平衡,说明当前数据集适合3个以下种子节点作为种子节点集合。图4(b)为Football数据集种子节点集合影响力的实验结果,4种算法的影响力呈逐渐上升趋势,并且在种子节点个数为6时,增长趋势逐渐变缓。图4(c)为HighSchool数据集种子节点集合影响力的实验结果,4种算法的影响力呈逐渐上升趋势,图像在种子节点个数 $k$ 分别为6时上升趋势逐渐放缓。图4(d)为SocDolphins数据集种子节点集合影响力的实验结果,4种算法的影响力波动较大,总体呈上升趋势,但我们可以观察到当 $k=5$ 开始,影响力已经开始小于种子节点的个数。

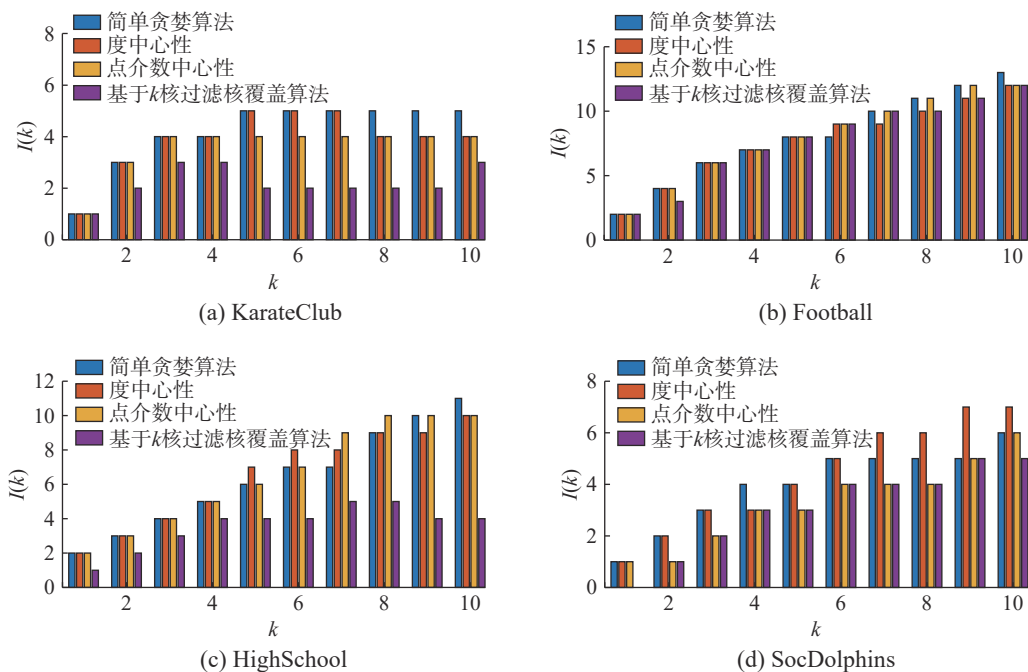


图4 影响力模拟

Fig. 4 The influence simulation

### 3.4 平均影响力分析

在图5中,  $k$  为种子节点个数, 纵坐标为当前种子节点集合单个节点的平均影响力。从图5(a)中可以发现, 当  $k=2$  时单个种子节点的平均影响力是最高的, 1个种子节点平均影响力1.5个节点左右。当  $k>4$  时, 单个种子节点的影响力不足于1个节点。所以 KarateClub 数据集适合选择2个种子节点作为种子节点集合较合适。从图5(b)中可以发现, 当  $k$  为1~3时, 种子节点的平均影响力最大, 1个种子节点平均影响2个节点, 当  $k>6$  时, 发现种子节点的平均影响力已经不足1.5个, 所以我们认为 Football 数据集适合选取7个种子节点作为种子节点集合比较合适。从图5(c)

中可以发现, 当  $k=1$  时单个种子节点的平均影响力是最高的, 1个种子节点平均影响力为2个节点左右, 当  $k>3$  时, 我们发现种子节点的平均影响力已经不足1.5个, 所以我们认为 HighSchool 数据集适合选取3个种子节点作为种子节点集合比较合适。从图5(d)中可以发现, 1个种子节点平均影响力最高为1个节点左右。点介数算法选取的种子节点在  $k>1$  时就出现了平均影响力的下降, 贪婪算法和度中心性算法选出的种子节点的影响力分别在  $k$  为4和3时出现下降, 因此 SocDolphins 数据集适合选择2个种子节点作为种子节点集合比较合适。

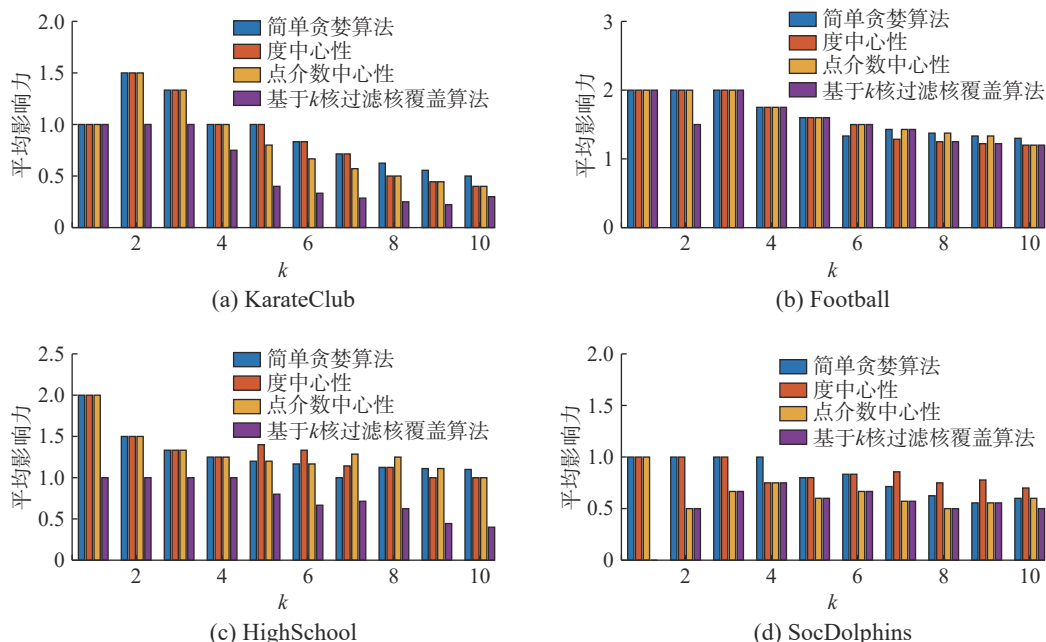


图5 平均影响力

Fig. 5 The average influence

通过实验可以看出, 一个网络的种子节点集合大小并不是越大越好, 而是存在一个上限, 当种子节点集合的大小超出了这个上限, 多出的种子节点并不能带来很好的边际收益。根据我们所提出的算法计算出的最优种子节点集合大小  $k'$ , 基本反映了一个网络传播影响力能力的上限, 因此为种子节点个数的选取提供了很好的参考, 并且可以用于一些选取最优种子节点集合算法中, 减少额外的时间开支。

## 4 结束语

本文主要对无向网络传播影响力的固有能力和进行的研究, 通过对网络进行渗流模拟得到网络的相变值, 发现相变值可以反应网络传播影响力的

能力, 并提出一种基于渗流模型的影响力最大化算法来选取网络所适合的种子节点集合的大小。在算法中, 我们建立传播概率与渗流模拟后网络最大连通子图大小的关系, 得到网络相变值  $p_c$ 。当传播概率  $p$  大于  $p_c$  时, 渗流后网络倾向于由一个主团块组成, 当传播概率  $p$  等于  $p_c$  时, 网络由多个大型团块组成。算法通过相变值与种子节点集合大小的换算, 得到当前网络最优的种子节点集合大小。实验结果表明该临界点对影响力最大化种子节点集合的大小选取起着重要的指导性作用。

## 参考文献:

- [1] BOND R M, FARISS C J, JONES J J, et al. A 61-million-

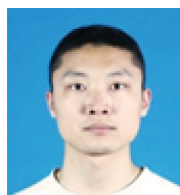


- person experiment in social influence and political mobilization[J]. *Nature*, 2012, 489(7415): 295–298.
- [2] 蒋庆丰, 门朝光, 田泽宇, 等. 社会感知的延迟容忍网络节点合作机制[J]. *哈尔滨工程大学学报*, 2017, 38(6): 921–930.
- JIANG Qingfeng, MEN Chaoguang, TIAN Zeyu, et al. A social-aware node cooperation mechanism for DTN[J]. *Journal of Harbin Engineering University*, 2017, 38(6): 921–930.
- [3] LU Zaixin, FAN Lidan, WU Weili, et al. Efficient influence spread estimation for influence maximization under the linear threshold model[J]. *Computational social networks*, 2014, 1: 2.
- [4] MORONE F, MAKSE H A. Corrigendum: influence maximization in complex networks through optimal percolation[J]. *Nature*, 2015, 527(7579): 544.
- [5] ZHU Zhiguo. Discovering the influential users oriented to viral marketing based on online social networks[J]. *Physica A: statistical mechanics and its applications*, 2013, 392(16): 3459–3469.
- [6] SADOVYKH V, SUNDARAM D, PIRAMUTHU S. Do online social networks support decision-making?[J]. *Decision support systems*, 2015, 70: 15–30.
- [7] SCHMITT P, SKIERA B, VAN DEN BULTE C. Referral programs and customer value[J]. *Journal of marketing*, 2011, 75(1): 46–59.
- [8] VERBRAKEN T, GOETHALS F, VERBEKE W, et al. Predicting online channel acceptance with social network data[J]. *Decision support systems*, 2014, 63: 104–114.
- [9] GUILLE A, HACID H, FAVRE C, et al. Information diffusion in online social networks: a survey[J]. *ACM SIGMOD record*, 2013, 42(1): 17–28.
- [10] GOLDENBERG J, LIBAI B, MULLER E. Talk of the network: a complex systems look at the underlying process of word-of-mouth[J]. *Marketing letters*, 2001, 12(3): 211–223.
- [11] 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, 2003: 137–146.
- [12] NEMHAUSER G L, WOLSEY L A, FISHER M L. An analysis of approximations for maximizing submodular set functions-I[J]. *Mathematical programming*, 1978, 14(1): 265–294.
- [13] FISHER M L, NEMHAUSER G L, WOLSEY L A. An analysis of approximations for maximizing submodular set functions-II[M]//BALINSKI M L, HOFFMAN A J. *Polyhedral Combinatorics: Dedicated to the Memory of D. R. Fulkerson*. Berlin, Heidelberg: Springer, 1978: 73–87.
- [14] CORNUEJOLS G, FISHER M L, NEMHAUSER G L. Location of bank accounts to optimize float: an analytic study of exact and approximate algorithms[J]. *Management science*, 1977, 23(8): 789–810.
- [15] WILLIAMS H P. Integer and combinatorial optimization[J]. *Journal of the operational research society*, 1990, 41(2): 177–178.
- [16] CHEN Wei, WANG Chi, WANG Yajun. Scalable influence maximization for prevalent viral marketing in large-scale social networks[C]//Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Washington, USA, 2010: 1029–1038.
- [17] LEE J R, CHUNG C W. A fast approximation for influence maximization in large social networks[C]//Proceedings of the 23rd International Conference on World Wide Web. Seoul, Korea, 2014: 1157–1162.
- [18] JUNG K, HEO W, CHEN Wei. IRIE: scalable and robust influence maximization in social networks[C]//Proceedings of IEEE 12th International Conference on Data Mining. Brussels, Belgium, 2012: 918–923.
- [19] KITSACK M, GALLOS L K, HAVLIN S, et al. Identification of influential spreaders in complex networks[J]. *Nature physics*, 2010, 6(11): 888–893.
- [20] CARMi S, HAVLIN S, KIRKPATRICK S, et al. A model of Internet topology using  $k$ -shell decomposition[J]. *Proceedings of the national academy of sciences of the United States of America*, 2007, 104(27): 11150–11154.
- [21] GAO Shuai, MA Jun, CHEN Zhumin, 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] 王皓, 高立群, 欧阳海滨. 多种群随机差分粒子群优化算法及其应用[J]. *哈尔滨工程大学学报*, 2017, 38(4): 652–660.
- WANG Hao, GAO Liqun, OUYANG Haibin. Multi-population random differential particle swarm optimization and its application[J]. *Journal of Harbin Engineering University*, 2017, 38(4): 652–660.
- [23] 刘畅, 刘利强, 张丽娜, 等. 改进萤火虫算法及其在全局优化问题中的应用[J]. *哈尔滨工程大学学报*, 2017, 38(4): 569–577.
- LIU CHANG, LIU Liqiang, ZHANG Lina, et al. An improved firefly algorithm and its application in global optimization[J]. *Journal of Harbin Engineering University*,



- 2017, 38(4): 569–577.
- [24] JI Shenggong, LÜ Linyuan, YEUNG C H, et al. Effective spreading from multiple leaders identified by percolation in the susceptible-infected-recovered (SIR) model[J]. *New journal of physics*, 2017, 19(7): 073020.
- [25] NEWMAN M E J. Spread of epidemic disease on networks[J]. *Physical review E*, 2002, 66(1): 016128.
- [26] MORONE F, MAKSE H A. Influence maximization in complex networks through optimal percolation[J]. *Nature*, 2015, 524(7563): 65–68.
- [27] CHEN Shuo, FAN Ju, LI Guoliang, et al. Online topic-aware influence maximization[J]. *Proceedings of the VLDB endowment*, 2015, 8(6): 666–677.
- [28] LIU Bo, CONG Gao, XU Dong, et al. Time constrained influence maximization in social networks[C]//*Proceedings of IEEE 12th International Conference on Data Mining*. Brussels, Belgium, 2012: 439–448.
- [29] ZACHARY W W. An information flow model for conflict and fission in small groups[J]. *Journal of anthropological research*, 1977, 33(4): 452–473.
- [30] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks[J]. *Proceedings of the national academy of sciences of the United States of America*, 2001, 99(12): 7821–7826.
- [31] ROSSANA M, JULIE F, ALAIN B, et al. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys[J]. *PLoS one*, 2015, 10(9): e0136497.
- [32] LUSSEAU D, SCHNEIDER K, BOISSEAU O J, et al. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations: can geographic isolation explain this unique trait?[J]. *Behavioral ecology and sociobiology*, 2003, 54(4): 396–405.
- [33] FREEMAN L C. A set of measures of centrality based on betweenness[J]. *Sociometry*, 1977, 40(1): 35–41.
- [34] 李阅志, 祝园园, 钟鸣. 基于  $k$ -核过滤的社交网络影响最大化算法 [J]. *计算机应用*, 2018, 38(2): 464–470.
- LI Yuezhi, ZHU Yuanyuan, ZHONG Ming.  $k$ -core filtered influence maximization algorithms in social networks[J]. *Journal of computer applications*, 2018, 38(2): 464–470.

### 作者简介:



花勇, 男, 1994 年生, 硕士研究生, 主要研究方向为影响力最大化与复杂网络。



陈伯伦, 男, 1986 年生, 讲师, 博士, 主要研究方向为链接预测、推荐系统和数据挖掘。主持国家自然科学基金青年基金、江苏省自然科学基金青年基金多项。发表学术论文 30 余篇。



朱国畅, 男, 1996 年生, 硕士研究生, 主要研究方向为人工智能、数据挖掘与机器学习。