

有限步传播范围期望指标判别节点传播影响力

李鑫 赵城利 刘阳洋

Distinguishing node propagation influence by expected index of finite step propagation range

Li Xin Zhao Cheng-Li Liu Yang-Yang

引用信息 Citation: *Acta Physica Sinica*, 69, 028901 (2020) DOI: 10.7498/aps.69.20191313

在线阅读 View online: <https://doi.org/10.7498/aps.69.20191313>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

您可能感兴趣的其他文章

Articles you may be interested in

一种基于社交影响力和平均场理论的信息传播动力学模型

An information diffusion dynamic model based on social influence and mean-field theory

物理学报. 2017, 66(3): 030501 <https://doi.org/10.7498/aps.66.030501>

基于区域密度曲线识别网络上的多影响力节点

Identifying multiple influential nodes based on region density curve in complex networks

物理学报. 2018, 67(19): 198901 <https://doi.org/10.7498/aps.67.20181000>

一种改进的基于信息传播率的复杂网络影响力评估算法

An improved evaluating method of node spreading influence in complex network based on information spreading probability

物理学报. 2017, 66(20): 208901 <https://doi.org/10.7498/aps.66.208901>

基于多重影响力矩阵的有向加权网络节点重要性评估方法

Evaluation method of node importance in directed-weighted complex network based on multiple influence matrix

物理学报. 2017, 66(5): 050201 <https://doi.org/10.7498/aps.66.050201>

节点中心性对复杂网络传播模式的影响分析

Analysis of the effect of node centrality on diffusion mode in complex networks

物理学报. 2017, 66(12): 120201 <https://doi.org/10.7498/aps.66.120201>

基于加权 K -阶传播数的节点重要性

Node importance based on the weighted K -order propagation number algorithm

物理学报. 2019, 68(12): 128901 <https://doi.org/10.7498/aps.68.20190087>

有限步传播范围期望指标判别节点传播影响力

李鑫[†] 赵城利 刘阳洋

(国防科技大学文理学院, 长沙 410073)

(2019年8月30日收到; 2019年10月30日收到修改稿)

在线社交网络逐渐成为人们不可或缺的重要工具, 识别网络中具有高影响力的节点作为初始传播源, 在社会感知与谣言控制等方面具有重要意义. 本文基于独立级联模型, 给出了一个描述有限步传播范围期望的指标——传播度, 并设计了一种高效的递推算法. 该指标在局部拓扑结构信息的基础上融合了传播概率对影响力进行刻画, 能够较好地反映单个节点的传播影响力. 对于多传播源影响力极大化问题, 本文提出了一种基于传播度的启发式算法——传播度折扣算法, 使得多个传播源的联合影响力最大. 最后, 将上述方法应用到三个真实网络中, 与经典指标和方法相比, 该方法不需要知道网络的全局结构信息, 而是充分地利用了网络的局部结构信息, 可以较快地筛选出高传播影响力的传播源.

关键词: 社交网络, 影响力度量, 影响力极大化, 信息传播

PACS: 89.70.Cf, 89.75.Fb

DOI: 10.7498/aps.69.20191313

1 引言

随着网络科学的快速发展, 节点的传播影响力越来越受到人们的关注. 在信息的传播过程中, 具有高传播影响力的节点可以加速信息的传播, 这对于宣传产品, 推送广告等方面具有重要的作用. 因此, 该领域的一个核心问题就是如何选择网络中的若干节点作为传播源, 使其传播影响力极大化, 又具体可分为单个传播源影响力排序问题和多个传播源影响力极大化问题 (需要考虑不同传播源之间影响力的重叠).

对于网络中节点影响力排序问题, 最简单高效的方法是度中心性^[1], 它反映节点的邻居数, 例如微博中拥有粉丝数多的大V用户就具有很高的传播影响力. 后来, Chen等^[2]又给出了一个局部中心性指标, 该指标考虑了二阶邻居数, 但这两个指标在反映节点传播影响力时没有考虑传播概率的影响, 也无法充分反映节点的局部拓扑结构. 对于

全局性的指标有介数中心性^[3]、接近度中心性^[4]、K核^[5]、特征向量中心性^[6]等. 而对于多源影响力极大化 (influence maximization) 问题, 最初由 Domingos 和 Richardson^[7]提出, 后被 Kempe 等^[8]证明该问题是 NP-Hard 难问题. 一种最简单的选取策略就是选出某指标值最大的若干节点作为传播源, 但这种算法选出的各传播者之间可能会有较大影响力重叠. 因此, Leskovec 等^[9]采用贪心算法来寻求初始传播集, 但这种算法的复杂度较大不适用于大型网络. Chen 等^[10]提出两种改进的贪心算法新贪心算法和最大贪心算法, 这两种改进算法均比贪心算法的算法复杂度要低, 但在大型网络中仍具有较高算法复杂度. 后来, 一些特殊的启发式算法被提出, 度折扣算法^[11]就是其中一种高效的启发式算法, 它是对探索式算法的一种优化策略的改进, 具有较高的实用价值, 其基本思想为当节点的邻居节点已被选为传播源时, 会对该节点的度指标产生折扣效应, 从而达到了去重叠的目的. 再后来, Kimura 等^[12]利用分解极大连通子图来寻找最优

[†] 通信作者. E-mail: 287998116@qq.com

传播源, 基于此, 又提出了利用最大影响路径的方法^[13], 这种方法虽然大大降低了了算法复杂度, 但限制了只能在最短路径上传播. Wang 等^[14]提出了一种结合动态规划的算法选择最优传播源, 提升了算法效率, Li 等^[15]考虑积极消极影响力, 提出了一种新的模型, 尽管这些算法效率都很高, 但很容易陷入局部最优的情况, 效果不及贪心算法.

在线社交网络具有规模庞大, 结构复杂的特点, 运用节点的局部信息指标描述节点的传播影响力更具现实意义. 但目前的局部性指标均不能够很好的刻画节点的局部拓扑结构, 因此本文基于独立级联模型, 提出一种描述节点有限步传播范围期望的指标——传播度. 对于网络中的任一节点作为传播源, 我们充分分析信息在局部网络中的传播路径, 设计了一种精确的递推算法计算有限步后传播范围的期望值, 之后通过真实网络数据集仿真验证了该指标与节点传播影响力具有更高一致性. 另一方面, 在考虑多传播源影响力极大化问题时, 选择的各传播源影响力会发生重叠, 因此基于传播度提出了一种启发式算法——传播度折扣算法, 同样通过真实网络数据集仿真验证了该算法相较于经典方法具有更好的效果.

本文的主要贡献在以下几个方面:

1) 提出了描述网络节点传播能力的指标——传播度, 用于描述节点在有限步后能够传播到网络节点范围的期望值. 并设计了计算该指标的迭代算法, 可以高效的计算节点的任意阶 (步数) 的传播度.

2) 相较于传统指标, 传播度指标考虑了传播概率对节点传播的影响. 对于小型网络, 可以直接利用该算法计算节点的最终传播期望; 对于大型网络, 可利用低阶传播度刻画节点的传播影响力, 充分利用了节点的局部拓扑结构信息.

3) 对于多传播源问题, 基于传播度指标提出了一种传播度折扣的算法, 能有效地去掉不同初始传播源间的影响力重叠, 相较于原有的度折扣算法有更好的选择效果.

2 有限步传播范围期望的算法

独立级联模型 (IC 模型) 是网络信息传播的一种最常用的模型, 当一个节点 v 被激活时, 它会以概率将其邻居节点 u 激活, 这种尝试仅进行一次, 在此之后, 节点 v 仍处于激活状态, 但不具有传播信息的能力. 根据该传播模型, 定义 s 阶传播度表

示传播 s 步后所有节点被激活的概率之和. 由于网络结构十分复杂, 为了保证概率的计算不重不漏, 先提出两个辅助算法, 最终给出一个高效的递推算法.

2.1 辅助算法一

为了更清楚地反映初始传播源的传播情况, 将节点的局部拓扑结构转化成树的形式, 并定义为该传播源的传播树, 该算法称为传播树算法. 根据信息在网络中的传播路线, 传播树是基于特定的单个种子节点的网络的另一种拓扑表示 (简单的例子见图 1). 图 1(b) 为图 1(a) 的传播树, 传播树的生成规则如下:

Step 1 确定一个种子节点作为 P_0 代节点;

Step 2 将种子节点的邻居作为种子节点的孩子, 记为 P_1 代节点, 令 $k = 1$;

Step 3 对于 P_k 代个体的任一个节点 i , 将节点 i 的邻居且非节点 i 的祖先的节点作为节点 i 的孩子. 则将所有 P_k 代个体的孩子记为 P_{k+1} 代个体;

Step 4 若 P_{k+1} 代个体数量大于 0, 则令 $k = k + 1$, 返回 Step 3; 否则, P_0 到 P_k 个体对应的树即为网络的传播树.

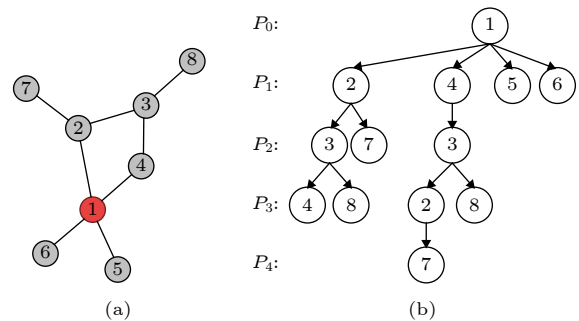


图 1 一个简单网络例子 (a) 网络图; (b) 图 (a) 对应的传播树

Fig. 1. A simple network example: (a) The network diagram; (b) the corresponding propagation tree of (a).

2.2 传播度的计算公式及辅助算法二

2.2.1 传播度的计算公式

在 2.1 节传播树概念的基础上, 先引入几个参量, 用 $p_t(u)$ 表示节点 u 在 t 时间内 (包含 t 时刻) 被感染的概率; $f_t(u_i)$ 表示 P_t 代的第 i 个 u 节点在 t 时刻被感染的独立概率; $f_t(u_{i_1}, \dots, i_n)$ 表示 P_t 代的第 i_1, \dots, i_n 个 u 节点在 t 时刻被感染的联合概率; $f_t(u)$ 表示 P_t 代的所有的 u 节点在 t 时刻

被感染的联合概率. 那么独立概率计算公式为

$$f_t(u_{\{i\}}) = f_{t-1}(v_{\{\bar{i}\}}) \times \beta \times (1 - p_{t-1}(u)), \quad (1)$$

其中 $v_{\{\bar{i}\}}$ 节点为 $u_{\{i\}}$ 节点的双亲节点, 且有

$$p_s(u) = \sum_{t=0}^s f_t(u), \quad (2)$$

那么种子节点 seed 的 s 阶传播度的计算公式为

$$d_s(\text{seed}) = \sum_u p_s(u). \quad (3)$$

2.2.2 辅助算法二

由 2.2.1 节的介绍, 关键问题在于如何计算联合概率 $f_t(u)$, 因此提出了辅助算法二——联合概率算法. 在介绍算法前, 首先分析概率的传播公式, 假设 u 节点在 t 时刻 (传播树的 P_t 代个体中) 出现 N 次, 此时节点 u 的双亲节点也有 N 个, 不失一般性, 记其双亲节点为 v_1, v_2, \dots, v_n , 其出现次数分别为 m_1, m_2, \dots, m_n 次, 因此共有 $m_1 + m_2 + \dots + m_n = N$ 条传播路径在 t 时刻可能传播到节点 u , 那么对于任意的 $j \in \{1, 2, \dots, n\}$, 节点 u 的双亲节点 v_j 有 m_j 条路径通向 u . 在 P_{t-1} 代个体中, 这 m_j 条路径经过的节点 v_j 在所有 v_j 中的序号保存在集合 X_j 中; 在 P_t 代个体中, 这 m_j 条路径经过的节点 u 在所有 u 中的序号保存在集合 Y_j 中. 在计算 $f_t(u)$ 时, 首先考虑相同双亲节点间的联合概率,

$$f_t(u_{Y_j}) = f_{t-1}(v_{jX_j}) \times \beta, \quad (4)$$

进一步计算不同双亲间的联合概率,

$$f_t(u) = \left[1 - \prod_{j=1}^n (1 - f_t(u_{Y_j})) \right] \times (1 - p_{t-1}(u)), \quad (5)$$

其中集合 X_j 只有一个元素时, 即为公式 (1) 所得的独立概率, 当 X_j 含有多个元素时, 仍用该方法计算联合概率. 进一步,

$$p_t(u) = p_{t-1}(u) + f_t(u). \quad (6)$$

事实上, 该算法在编程时是一个计算联合概率 $f_t(u_w)$ 可调用函数, 当 $t-1$ 时间内传播树中各点的独立概率已递推求得, 用 $F(t, u, W)$ 表示调用函数, 其算法的伪代码如表 1 所列.

2.3 传播度递推求解

由方程 (6) 可知, 在计算 $p_t(u)$ 时, 需要知道 $f_t(u)$ 和 $p_{t-1}(u)$. 而计算 $f_t(u)$ 时, 根据 2.2 节中的分析, 最终转化为计算 $t-1$ 时间内 (包含 $t-1$ 时刻) 传播树中各点的独立概率的计算式. 因此, 可

以用递推算法最终求得传播度. 求解 s 阶传播度的算法流程伪代码如表 2 所列. 有了表 2 所列的递推

表 1 联合概率算法

Table 1. Joint probability algorithm.

辅助算法二 联合概率算法	
输入	时间 t , 节点 u 以及对于 u 的序号集 W
输出	联合概率 $f_t(u_w)$
1.	if W 只含一个元素 then
2.	找到该节点的双亲节点 $v_{(i)}$
3.	$f_t(u_w) = f_{t-1}(v_{(i)}) \times \beta \times (1 - p_{t-1}(u))$
4.	else
5.	找到传播树种 P_t 代序号集为 W 的 u 节点, 记其双亲节点为 v_1, v_2, \dots, v_l , 出现次数为 m_1, m_2, \dots, m_l , 这些路径对于双亲节点的序号集分别为 X_1, X_2, \dots, X_l , 对于节点 u 的序号集分别为 Y_1, Y_2, \dots, Y_l .
6.	for $j = 0 \rightarrow l$ do
7.	$f_t(u_{Y_j}) = F(t-1, v_j, X_j) \times \beta$
8.	end for
9.	$f_t(u_w) = [1 - \prod_{j=1}^l (1 - f_t(u_{Y_j}))] \times (1 - p_{t-1}(u))$
10.	end if
11.	return $f_t(u_w)$

表 2 种子节点 s 阶传播度的算法流程伪代码

Table 2. Pseudocode of the algorithm flow for s order propagation of the seed node.

递推算法 求种子节点 s 阶传播度	
输入	seed, β , s
输出	d_s
1.	$\text{Tree}_s = \text{FTree}(\text{seed})$ % 获取节点 seed 的 s 阶传播树
2.	赋初值 $p_0(\text{seed}) = f_0(\text{seed}) = 1$, 其他情况 $p_t(u) = f_t(u) = 0$
3.	for $t = 1 \rightarrow s$ do
4.	for $u_y \in P_t$ do % 遍历传播树中 P_t 代的个体
5.	获得 u_y 的双亲节点 v_x % 表示节点 u 所在传播树层中第 x 个 u 节点
6.	$f_t(u_y) = f_{t-1}(v_x) \times \beta \times (1 - p_{t-1}(u))$
7.	end for
8.	for $u \in P_t$ do
9.	获得 u 的序号集 W % 指 P_t 代节点 u 出现次序构成的集合
10.	计算联合概率 $f_t(u) = F(t, u, W)$
11.	end for
12.	for $u \in \text{Tree}_s$ do
13.	$p_t(u) = p_{t-1}(u) + f_t(u)$
14.	end for
15.	end for
16.	$d_s(\text{seed}) = \sum_u p_s(u)$
17.	return d_s

算法,便可以编程高效的计算节点的任意阶的传播度.

2.4 仿真验证

综上所述,对于一个网络,确定初始种子节点,先用传播树算法得到相应的传播树,在利用递推算得到任何时刻任意节点的期望值 $p_t(i)$,进而得到任意时刻的传播度值.为了进一步验证算法的精确性,我们基于独立级联模型对图1所示网络进行传播仿真.选择不同的仿真次数,比较情况见图2.

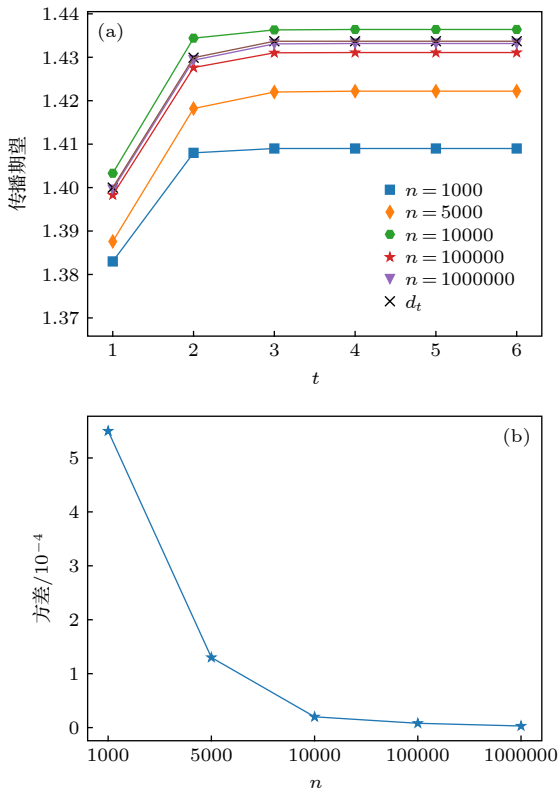


图2 (a) 不同仿真次数下的近似传播期望值和传播度随时间的变化; (b) 不同仿真次数的近似传播期望值与传播度的方差变化

Fig. 2. (a) Approximate propagation expectation and the degree of propagation over time for different simulation times; (b) variance of the approximate propagation expectation and the degree of propagation of different simulation times.

由图2可知,模型计算的期望值是精确的.有了传播度的指标后,便可以用该指标来反应节点的传播影响力.

对于小型网络,可以直接计算最高阶的传播度来反应节点的传播影响力,该方法有两方面优势.一方面,在现实中,出于商业考虑,可能不会选取

明显较优的核心节点,如图1小型网络中的1号节点,因此,进一步比较2,3号,甚至5,6,7,8号节点的传播影响力优劣具有重要意义.另一方面,网络中节点的传播影响力是与传播概率相关的,而传播度指标可以很好地将传播概率融入到刻画节点传播影响力中去.

如图1所示,节点最多传递四步,因此利用上述方法计算所有节点的四阶传播度来反映节点的传播影响力.为了比较各节点传播影响力的大小以及验证节点的传播影响力与传播概率相关,图3给出了不同传播概率下各节点四阶传播度的变化情况.

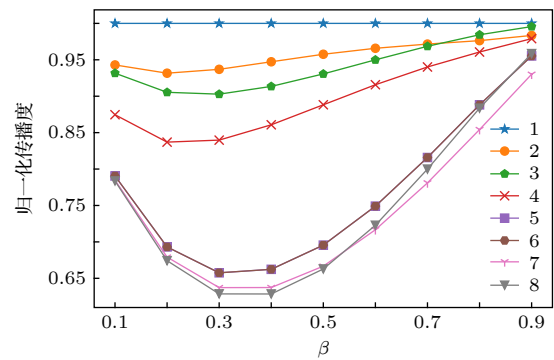


图3 图1中的网络各节点4阶传播度(归一化)随传播概率变化

Fig. 3. Changes of the 4th degree of propagation (normalized) of network node in Fig. 1 with the propagation probability.

由图3可知,随着传播概率不断增加,各节点的传播影响力排序会发生明显变化,且可以量化比较各节点传播影响力大小.当传播概率较大时,可以选择普通的节点作为传播源,兼顾成本和传播能力两个方面.后文将主要关注传播度在大型网络中识别高传播影响力传播源的作用展开详细讨论.

3 基于传播度的节点影响力最大化方法

初始种子节点的传播影响力最大化问题分为单源和多源两种情况,下面分别给出判别算法.

3.1 单个节点的影响力排序

由传播度的定义知道,一阶传播度为 $d_1 = 1 + \beta d_{seed}$,其中 d_{seed} 种子节点的度,因此一阶传播度和节点度是等价的.而二阶传播度事实上不仅反

映了种子节点的一阶邻居和二阶邻居情况, 还反映了一阶邻居之间的连边拓扑关系, 例如, 种子节点的邻居节点很有可能实在第二步通过另一个邻居节点被传播的. 类似地, 考虑更高阶传播度时, 不仅反映了高阶邻居的情况, 还反映了较低阶邻居之间的拓扑关系.

3.2 传播度折扣解决多源影响力极大化问题

关于多源传播影响力极大化问题, 一般采用启发式算法, 然而不同传播源的传播影响力会有较多“重叠”部分, 因此在算法进行的过程中需要采用去重叠过程. 由于定义的传播度是基于概率期望的, 基于这一特性, 去重叠过程是十分简便的, 下面进行理论分析.

考虑 s 阶传播度, 计算出所有节点的 s 阶传播度, 选择最大的一个节点加入种子节点集 R 的第一个节点, 其各节点的传播期望为 $p_s(i), i \in \{1, 2, \dots, N\}$. 记 ptotal 为 N 阶向量, 表示考虑种子节点集 R 中所有节点的 s 阶传播度, 网络中所有节点未患病的概率. 不失一般性, 假设 R 中已有 k 个节点, 这些节点到网络中所有节点的传播期望值为 $p_s^{(j)}(i), i \in \{1, 2, \dots, N\}, j \in \{1, 2, \dots, k\}$, 表示 R 中的第 j 号种子节点到节点 i 的传播概率 (仅考虑节点的 s 阶邻居内的节点). 则相应的 ptotal 值变为

$$\text{ptotal}(i) = \prod_{t=1}^k (1 - p_s^{(t)}(i)), \quad (7)$$

下面需要选择第 $k+1$ 个节点, 任选一个节点 v , 需要考虑节点 v 与已选种子节点集 R 的重叠情况, 进而求得节点 v 的有效贡献值. 一方面, 节点 v 是有效的, 说明已选节点没有传播到节点 v , ptotal 向量记录了已选节点未传播到其他节点的概率, 因此可以得到节点集 R 未传播到节点 v 的概率, 记为 $\text{ptotal}(v)$. 另一方面, 由于节点 v 可能传播到的节点可能已被传播, 因此需要引入折扣因子, 若节点 v 到节点 u 的传播期望值为 $p_s(u)$, 则引入折扣因子后变为 $p_s(u) \times \text{ptotal}(u)$, 因此在选择第 $k+1$ 个节点前, 先计算节点 v 的传播折扣度为

$$\tilde{d}_s = \text{ptotal}(v) \times \sum_{u \in G.\text{neis}(v)} (p_s(u) \times \text{ptotal}(u)), \quad (8)$$

其中 $G.\text{neis}(v)$ 表示节点 v 的 s 阶邻居内的节点 (不包括 v 节点本身). 将候选者按传播折扣度排序, 选出最优者加入种子节点集合. 多次执行上述操作, 即可得到近似最优的传播影响力节点集.

3.3 模型总述

综上所述, 算法的总流程图如图 4 所示. 从新定义的传播度指标出发, 充分考虑节点所有可能就近的传播路径和传播概率的大小, 分别对单源节点影响力排序问题和多源影响力极大化问题给出了解决方案. 对于小型网络, 我们在 2.4 节中指出, 可以计算最高阶的传播度从而可以精确的找出最优传播源. 对于大型网络, 将在第 4 节详细验证该模型的优势.

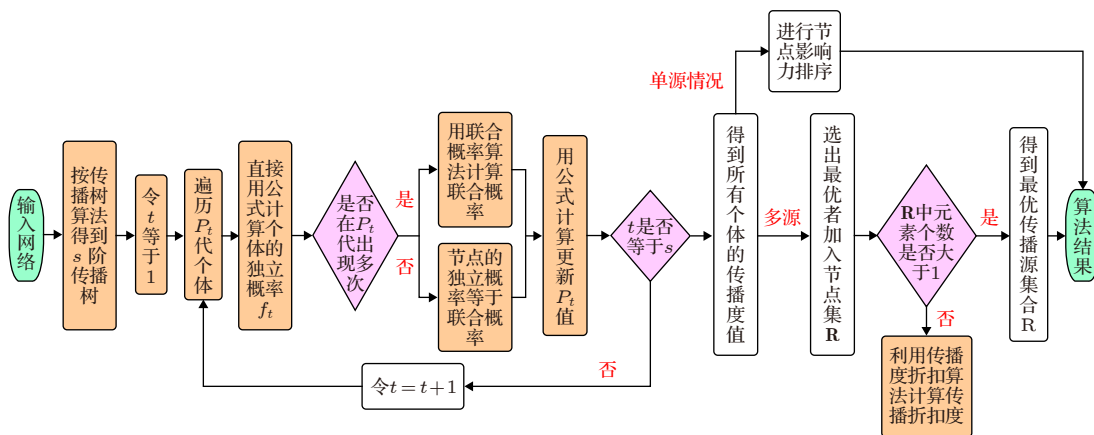


图 4 基于传播度的节点影响力极大化算法流程图

Fig. 4. Flow chart of node influence maximization algorithm based on propagation degree.

4 效果分析

由渗流理论可知,当传播概率较大且高于渗流阈值时,单个节点就具有传播到全局的能力,此时可以用全局传播概率(参见文献[16])反映节点的传播能力.因此,在检验算法效果时,应分两种情况考虑:一种是传播概率较小且低于渗流阈值,此时用平均传播范围表示节点传播能力;当传播概率较大时,用全局传播概率来反映.

4.1 数据集

为了评价模型,采用如下3个不同领域的真实网络数据集(无向)进行仿真实验.

1) Deezer^[17]: 数据来源于音乐流服务网站 Deezer 提供的数据,表示罗马尼亚用户朋友网络.数据下载于: <http://snap.stanford.edu/data/>(简称 SLNDC),共有 41773 个节点,平均度为 6.024.

2) Email-Enron^[18]: 数据来源于安然电子邮件网络,下载于 SLNDC,共有 36692 个节点,平均度为 10.020.

3) Facebook^[17]: 数据来源 Facebook 朋友网络,下载于 SLNDC,共有 13866 个节点,平均度为 12.528.

4.2 单源影响力排序效果

采用节点的仿真影响力值和传播度值的 kendall's tau 系数^[19]来反映单源影响力的排序效果.该指数反映了两个序参量的一致性,大小介于-1和1之间,当该系数越大时,两个序参量越趋于一致.该算法排序效果见图5和图6.

由图5和图6可知,传播度相较于度,高阶传播度相较于低阶传播度与节点的传播能力之间具有更高的一致性,特别是在传播概率偏大时,这种效果是明显的,这种现象同样适用其他的真实网络.

目前最常用的节点重要度排序指标有 k 核^[20](coreness),特征向量中心性^[6](eigenvector),考虑节点最近邻居核次近邻居的多级邻居信息指标^[2](local centrality),而介数中心性和接近度中心性由于其算法复杂度较高,不适用于大型网络.不同方法下的比较情况见图7和图8.

由图7和图8可以看出,对于大型网络,二阶传播度相较于其他指标在不同的传播概率和网络下具有更好的效果和更高的稳定性.因此,我们给出的传播度能更好地利用节点的局部拓扑结构来描述单源影响力排序.

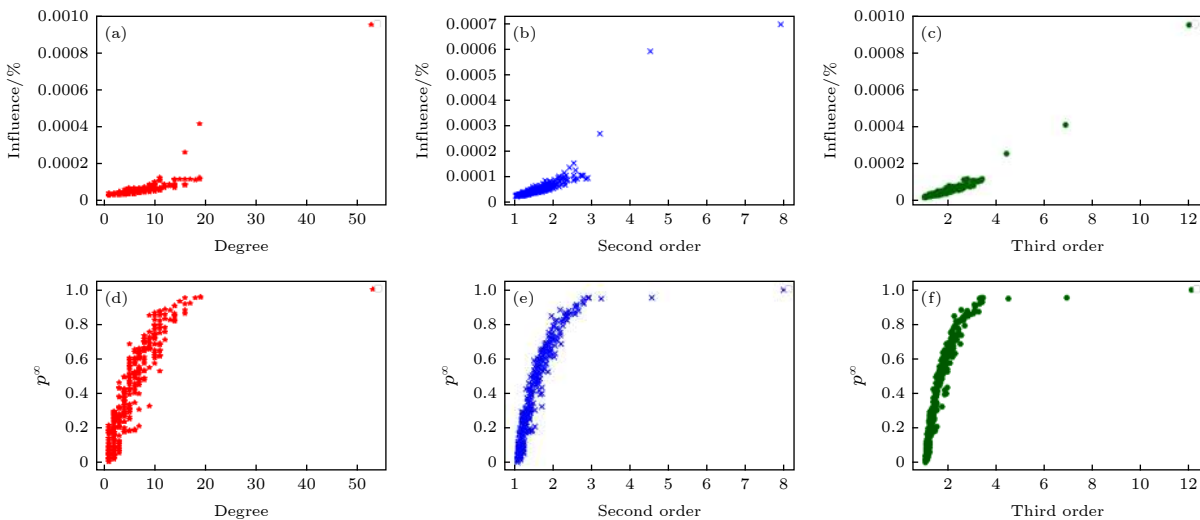


图5 (a)–(c) 分别反映了 Deezer 网络随机挑选若干节点的传播能力与度,二阶传播度,三阶传播度的关系,其中 $\beta = 0.06$; (d)–(f) 分别反映了 Deezer 网络随机挑选若干节点全局传播概率(参见文献[12])与度,二阶传播度,三阶传播度的对应情况,其中 $\beta = 0.12$ (这里通过仿真实验不难发现该网络的渗流阈值介于 0.06 和 0.12 之间)

Fig. 5. (a)–(c): Respectively reflects the relationship between the propagation capacity and degree, second-order propagation, and third-order propagation of several nodes randomly selected by the Deezer network, where $\beta = 0.06$; (d)–(f): Respectively reflects the global propagation probability of randomly selected nodes by the Deezer network (see reference[12]) Correspondence with degrees, second-order propagation, and third-order propagation, where $\beta = 0.12$ (here, it is not difficult to find through the simulation experiments that the percolation threshold of the network is between 0.06 and 0.12).

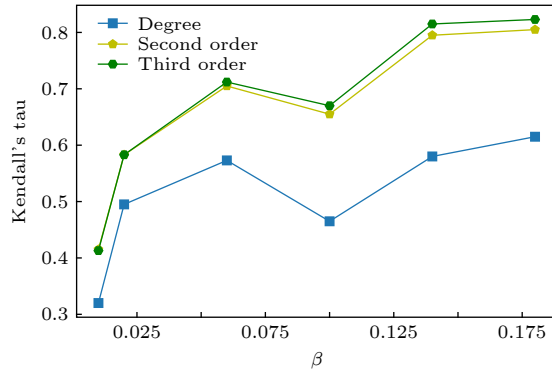


图 6 不同传播概率下度, 二阶、三阶传播度与仿真传播能力的 kendall's tau 系数

Fig. 6. Kendall's tau coefficient for different propagation probabilities, second-order, third-order propagation and simulation propagation ability.

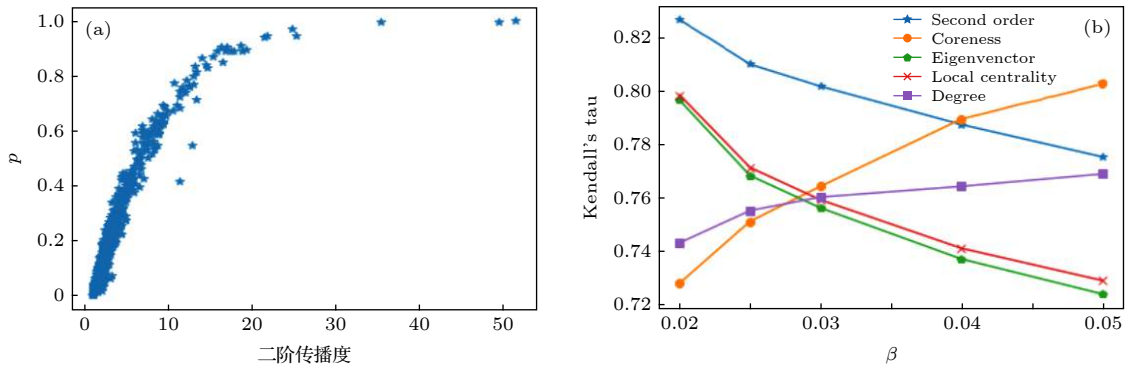


图 7 (a) 在 Email-Enron 网络中, 传播概率为 0.02 的情况下二阶传播度和全局传播概率 p 的关系; (b) 各指标在不同概率下与传播能力的 kendall's tau 系数

Fig. 7. (a) The relationship between the second-order propagation degree and the global propagation probability p in the case of the Email-Enron network with a propagation probability of 0.02; (b) the kendall's tau coefficient of each indicator with different propagation rates and propagation capabilities.

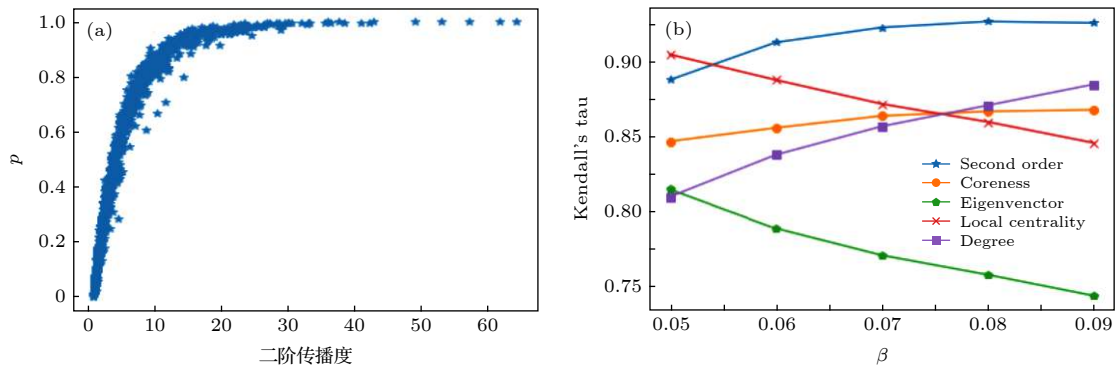


图 8 (a) 在 Facebook 网络中, 传播概率为 0.09 的情况下二阶传播度和全局传播概率 p 的关系; (b) 各指标在不同概率下与传播能力的 kendall's tau 系数

Fig. 8. (a) The relationship between the second-order propagation degree and the global propagation probability p in the case of a Facebook network with a propagation probability of 0.09; (b) the kendall's tau coefficient of each indicator with different propagation rates and propagation capabilities.

4.3 多源影响力极大化效果

通过实验发现, 分解极大连通子图等方法虽然算法复杂度低, 但由于做了过多近似, 效果不如算法复杂度较高的贪心算法, 因此这些方法不是本文

研究的重点, 而是致力于贪心算法的一种改进算法. 为了验证传播度折扣算法的效果, 将该算法与最大度 [1] (degree), 核 [20] (coreness), 特征向量 [6] (eigenvector), 介数 [3] (betweenness), 折扣度算法 [11]

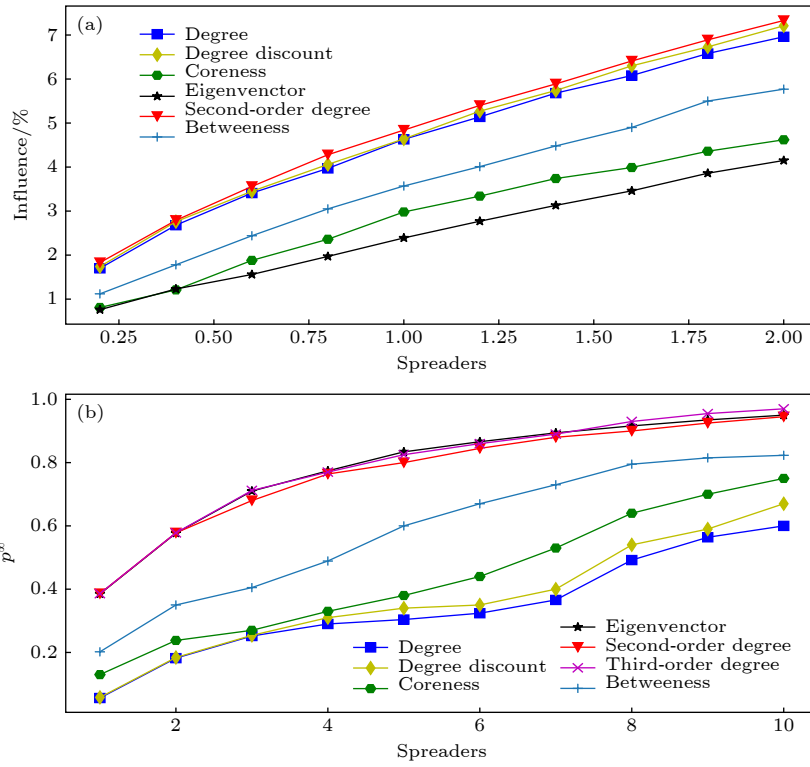


图 9 在 Deezer 网络中 (a) 不同算法下传播范围随所选种子节点数量的变化和 (b) 不同算法下全局传播概率随种子节点数量的变化, 其中候选节点为度为 6, 7, 8 的 9792 个节点

Fig. 9. In the Deezer network, (a) the variation of the propagation range with the number of selected seed nodes under different algorithms; (b) the variation of the global propagation probability with the number of seed nodes under different algorithms. The candidate nodes are 9792 nodes with a degree of 6, 7 and 8.

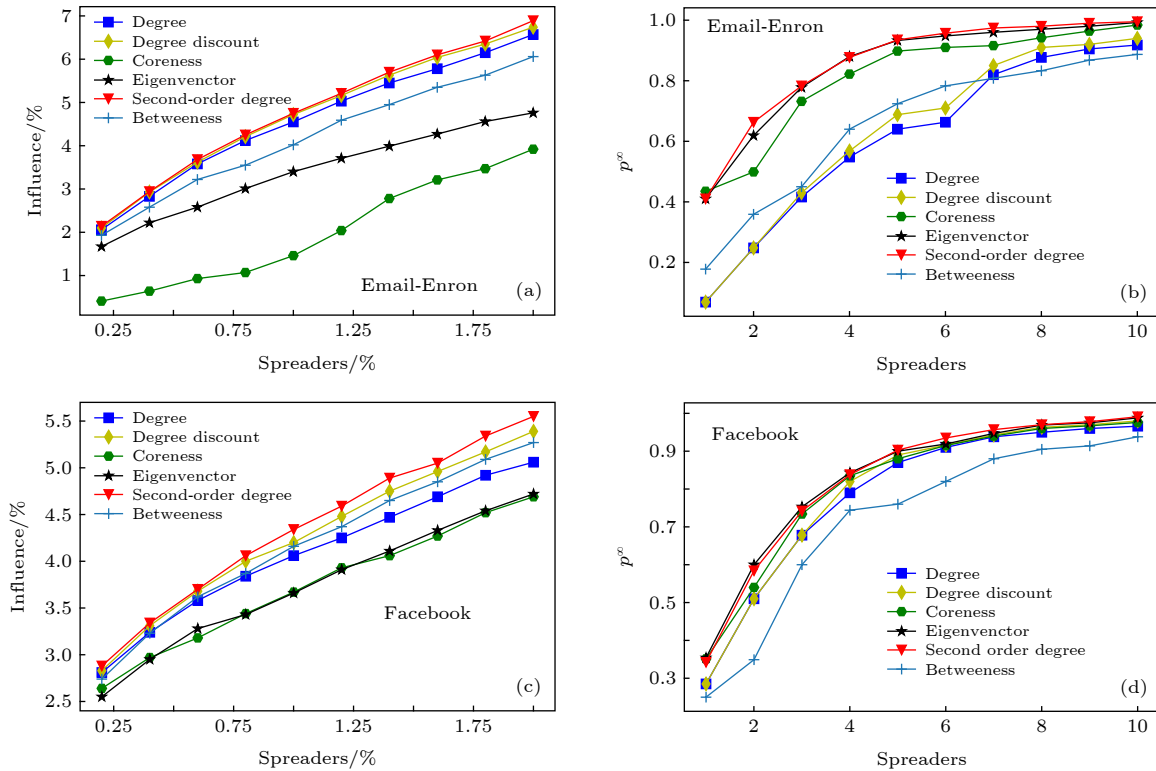


图 10 (a), (b) 和 (c), (d) 分别比较了 Email-Enron, Facebook 网络在不同算法下所选种子节点的传播性能
Fig. 10. (a), (b), and (c), (d), respectively, compare the propagation performance of Email-Enron, the selected seed node of the Facebook network under different algorithms.

(degreediscount) 进行仿真比较. 同样, 也分两种情况来比较: 一种取传播概率较小且低于渗流阈值, 另一种取传播概率较大且高于渗流阈值. 算法效果见图 9 和图 10 所示.

由图 9 和图 10 可以看出, 传播度折扣算法能够较好的解决多源影响力极大化问题, 因此在利用局部信息去除影响力重叠时, 传播度折扣算法具有不错的效果.

5 总 结

在充分考虑节点局部拓扑结构信息的基础上, 提出了有限步传播范围期望的指标——传播度, 并设计了高效精确的递推计算方法. 在此基础上, 将该指标用于单源影响力排序问题, 实验证明, 该指标与节点的传播能力具有更高的一致性. 另外, 对于多源影响力极大化问题, 本文基于传播度的概念, 设计了一种启发式算法——传播度折扣算法, 相较于经典的算法, 具有更好的筛选效果.

给出了一种精确计算有限步传播范围期望的递推算法, 事实上, 当传播到达一定步数后, 即当传播步数较大时, 节点被传播的概率往往很小, 因此, 如何设置局部的递推终止条件, 引入较小的误差, 可以进一步优化算法, 计算更高阶的传播度, 这一问题有待进一步研究.

参考文献

[1] Fershtman M 1997 *Soc. Networks* **19** 193
 [2] Chen D B, Lu L Y, Shang M S, Zhang Y C, Zhou T 2012

Physica A **391** 1777
 [3] Freeman L C 1977 *Sociometry* **40** 35
 [4] Bavelas A 1950 *J. Acoust. Soc. Am.* **22** 725
 [5] Dabbousi B O, Rodriguez-Viejo J, Mikulec F V, Heine J R, Mattoussi H, Ober R, Jensen K F, Bawendi M G 1997 *J. Phys. Chem. B* **101** 9463
 [6] Min L, Liu Z, Tang X Y 2015 *Acta Phys. Sin.* **64** 088901 (in Chinese) [闵磊, 刘智, 唐向阳 2015 物理学报 **64** 088901]
 [7] Domingos P, Richardson M 2001 *Proceedings of KDD'01 ACM SIG KDD International Conference on Knowledge Discovery and Data Mining* San Francisco, CA, USA, August 26–29, 2001 p57
 [8] Kempe D, Kleinberg J, Tardos E 2003 *Proc. 9th ACM SIGKDD Intl. Conf. On Knowledge Discovery and Data Mining* New Washington, DC, USA, August 24–27, 2003 p137
 [9] Leskovec J, Krause A, Guestrin C, Faloutsos C, VanBriesen J, Glance N 2007 *Proc. 13th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining* USA, August 24–27, 2007 p420
 [10] Chen W, Wang Y, Yang S 2009 *Proc. 15th ACM SIGKDD Intl. Conf. on Knowledge discovery and data mining* Paris, France, June 28–July 1, 2009 p199
 [11] Sheikahmadi A, Nematbakhsh M A, Shokrollahi A 2015 *Physica A* **436** 833
 [12] Kimura M, Saito K 2006 *Knowledge-Based Intelligent Information and Engineering Systems: 10th international conference* Bournemouth, UK, October 9–11, 2006 p259
 [13] Chen W, Wang C, Wang Y 2010 *Proc. 16th ACM SIGKDD Intl. Conf. on Knowledge discovery and data mining* WashingtonDC, USA, July 25–28, 2010 p1029
 [14] Wang Y, Cong G, Song G 2010 *Proc. 16th ACM SIGKDD Intl. Conf. on Knowledge discovery and mining* Washington DC, USA, July 25–28, 2010 p1039
 [15] Li D, Xu Z M, Chakraborty N 2014 *PloS One* **9** e102199
 [16] Hu Y Q, Ji S G, Jin Y L, Feng L, Stanley H E, Havlin S 2018 *PNAS* **115** 7468
 [17] Lancichinetti A, Fortunato S 2009 *Phys. Rev. E* **80** 056117
 [18] Leskovec J, Lang K, Dasgupta A, Mahoney M W 2009 *Internet Mathematics* **6** 29
 [19] Lapata M 2006 *Comput. Linguist.* **32** 471
 [20] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E, Makse H A 2010 *Nat. Phys.* **6** 888

Distinguishing node propagation influence by expected index of finite step propagation range

Li Xin[†] Zhao Cheng-Li Liu Yang-Yang

(College of Liberal Arts and Sciences, National University of Defense Technology, Changsha 410073, China)

(Received 30 August 2019; revised manuscript received 30 October 2019)

Abstract

On-line social networks have gradually become an indispensable tool for people. Identifying nodes with high influence in the network as an initial source of communication is of great significance in social perception and rumor control. According to the independent cascade model, in this paper we present an index describing the finite step propagation range expectation as the degree of propagation, and design an efficient recursive algorithm. Based on the local topology information, the index combines the propagation probability to characterize the influence, which can better reflect the propagation influence of a single node. For a single propagation source influence ordering problem, the node degree of propagation and propagation capability are better consistent with each other. And the propagation degree can well describe the propagation influence of nodes under different networks and propagation probabilities. For maximizing the multi-propagation source influence, in this paper we propose a propagation-based heuristic algorithm which is called propagation discount algorithm. This algorithm makes the joint influence of multiple propagation sources maximized. Finally, in this paper we apply the above method to three real networks, showing better effects than the classic indicators and methods. The algorithm has three advantages. First, the expected value of the final propagation range of each node in the small network can be accurately calculated. Second, the degree of propagation fully considers the local topology of the node and belongs to a locality indicator. Third, the indicator combines the effect of propagation probability and yields good outcomes under different networks and propagation probabilities.

Keywords: social network, influence measurement, influence maximization, information dissemination

PACS: 89.70.Cf, 89.75.Fb

DOI: [10.7498/aps.69.20191313](https://doi.org/10.7498/aps.69.20191313)

[†] Corresponding author. E-mail: 287998116@qq.com