

# 社交网络中基于贝叶斯和半环代数模型的 节点影响力计算机理\*

赵佳 喻莉<sup>†</sup> 李静茹

(华中科技大学电子信息与工程系, 武汉 430074)

(武汉光电国家实验室, 武汉 430074)

(2013年1月5日收到; 2013年3月19日收到修改稿)

本文综合考虑网络结构及节点间的互动等关键因素, 提出了一种节点影响力分布式计算机理. 首先根据节点交互行为在时域上的自相似特性, 运用带折扣因子的贝叶斯模型计算节点间的直接影响力; 然后运用半环模型来分析节点间接影响力的聚合; 最后根据社交网络的小世界性质及传播门限, 综上所述计算出节点的综合影响力. 仿真结果表明, 本文给出的模型能有效抑制虚假粉丝导致的节点影响力波动, 消除了虚假粉丝的出现对节点影响力计算带来的干扰, 从中选择影响力高的若干节点作为传播源节点, 可以将信息传播到更多数目的节点, 促进了信息在社交网络中的传播.

**关键词:** 社交网络, 影响力, 贝叶斯, 半环代数

**PACS:** 02.50.Fz, 02.10.H, 89.20.Ff

**DOI:** 10.7498/aps.62.130201

## 1 引言

社交网络改变了以往传播纯粹的客户端/服务器结构, 人人都可以成为媒体, 传播更类似一种对等 (Peer to peer, P2P) 结构, 包含的社会链接关系极大地促进了病毒式营销传播, 即用户可以把应用、产品或信息传递给他们的好友, 然后他们的好友又把这种信息或产品推荐给他们好友的好友. 由于社交网络中的好友关系一般是比较可信的, 这种病毒式营销<sup>[1]</sup> 无论是能将应用、产品或信息推送到广大潜在用户, 拓展应用、产品或信息的覆盖范围, 或是在舆情控制中, 遏制谣言传播等方面, 有着重大意义. 例如, 在社交网络中的产品推广中, 利用部分影响力很高的节点进行宣传, 在较短期限内吸收大量人气<sup>[2]</sup>; 针对谣言传播<sup>[3,4]</sup>, 利用部分影响力很高的节点引导公众认识, 达到有效辟谣的目的等. 因此, 社交网络中病毒式营销传播依赖于选择部分影响力很高的节点作为传播源节点, 使得合理计算

节点的影响力以选择合适的传播源节点显得十分重要, 在社交网络研究中获得了广泛的关注.

社交网络的结构或拓扑通常被用来挖掘计算网络中节点的影响力: Zhou 等<sup>[5]</sup> 通过社团结构划分使得影响力最大化从而获取若干个关键节点; Newman<sup>[6]</sup> 的研究表明, 社交网络中度数较高的节点受到攻击失效, 网络的连通性就被破坏, 因此节点的度是衡量节点影响力的重要参数, 最为直观的一种方法是直接选择度最高的若干个节点作为初始营销节点; PageRank<sup>[7]</sup> 方法, 根据邻接矩阵和设定初始值不停循环直到收敛, 从而得出节点的影响力; 吕琳媛等<sup>[8]</sup> 针对 PageRank 方法在非连通网络中排序不唯一的缺陷, 提出了一种 LeaderRank 方法, 网络中增加一个与所有节点双向连通的节点, 使得整个网络连通且排序唯一; 对于 PageRank 方法本质上只应用了主特征向量, Ilyas<sup>[9]</sup> 指出仅仅依靠主特征向量不足与表征整个网络, 采用主分量中心势 (PCC) 方法, 选取若干个主特征向量来确定网络的前若干个影响力最大的节点.

\* 国家自然科学基金重点项目 (批准号: 61231010, 60972016) 和湖北省杰出青年科学家基金 (批准号: 2009CDA150) 资助的课题.

<sup>†</sup> 通讯作者. E-mail: hustlyu@mail.hust.edu.cn

上述方法均为依据网络的拓扑结构特性, 确定网络中较为重要的一些节点以促进病毒式宣传. 然而目前社交网络中部分节点为获取较高的影响力, 采用了增加虚假关注的手段, 例如在微博网络中虚假粉丝的出现, 该类粉丝不会评论或转发博主的信息, 只是通过提高博主的关注人数以提升博主的关注度, 即通过虚假关注提升节点的度. 依据基于网络结构的节点影响力判断方法, 该类节点的影响力得到了虚假提升. 虚假关注改变了节点的度以及网络的结构, 成为了仅依靠结构衡量节点影响力方法的一个障碍, 严重误导社交网络中的病毒式营销宣传. 因此, 将网络结构作为节点影响力的衡量标准不足以应对虚假粉丝造成的此类障碍.

鉴于此, 本文综合考虑节点的度和节点间的互动等关键因素, 设计了一种节点影响力的分布式计算机理. 依据节点互动行为在时域上的自相似特性, 本文使用带折扣因子的贝叶斯模型计算节点直接影响力, 运用半环模型<sup>[10]</sup>来分析节点间接影响力的聚合, 然后根据小世界性质和传播门限, 联合两者计算出节点的综合影响力. 该机理可以根据具体的传播门限要求, 计算出相应的影响力, 即该节点在具体的传播门限下的扩散范围. 仿真表明: 该机理能较为全面地计算节点影响力, 因而能有效对抗虚假粉丝攻击, 抑制虚假粉丝导致的节点影响力的波动, 选择其中影响力较高的若干节点作为信息的初始传播节点, 能增加获取信息的节点数目, 有效促进了消息传播.

## 2 基于贝叶斯和半环代数模型的节点影响力计算机理

节点的影响力是指网络中节点对于其他节点行为的支配能力, 本文主要研究节点在传播行为方面的影响力. 影响力有以下三个特征: 首先, 影响力本质上是一种不确定性, 是节点能将信息传播到一定范围内的可能性; 其次, 鉴于互惠性质并非广泛存在于社交网络节点之间, 影响力可以是不对等的; 再次, 节点间的关系是逐渐演化的, 故影响力是时变的. 据此, 本节首先提出了直接影响力和间接影响力的概念, 并定义相关计算规则.

### 2.1 直接影响力及其计算规则

文中将直接影响力定义为某节点对直接相连的节点的影响力, 如图 1 所示, 节点 A 对节点 B 的

影响力即为直接影响力.

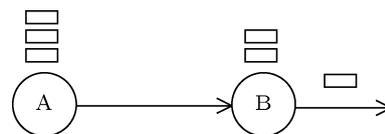


图 1 源节点 A 对于传播节点 B 的直接影响力

考虑到节点互动行为 (对信息的接受转发行为) 在时域上具有相似性, 本文首先应用贝叶斯模型计算节点的影响力, 根据节点之间传播行为的历史记录来预期下一跳节点进行传播的可能性.

在直接影响力的贝叶斯计算模型中, 定义互动行为的随机变量如下:  $a(i)$  表示第  $i$  次互动中节点 B 的行为,  $a(i) = 1$  表示第  $i$  次互动中节点 B 接受或转发节点 A 的信息,  $a(i) = 0$  则表示第  $i$  次互动中传播节点既不接受也不转发节点 A 的信息; 总的传播信息的次数  $k = \sum_N a(i)$ , 记录在  $N$  次互动中, 节点 B 接受或转发节点 A 的信息次数.

根据最小无偏估计可得, 源节点对传播节点的直接影响力即传播节点下一次是否接受或转发信息的概率为

$$I_{AB} = p(a(i+1) = 1) = k/N, \quad (1)$$

其次根据社交网络话题热度时变的性质以及节点的上下线行为, 很久之前节点的互动行为相比于近期互动行为所占的比重更小, 因此引入折扣因子度量节点近期参与传播的可能性.

若用户在近期内很活跃, 则近期参与传播的可能性较高; 而若用户曾经参与传播但是近期内不活跃, 则其传播可能性降低. 假设一个话题持续热度的时间为  $T$ , 则据现在时刻  $T$  以内的行为是积极有效的, 能较准确的用来计算节点的影响力, 那么

$$I_{AB} = p = \frac{\sum_{t-t_i > T} a(i) \beta_i^{t-t_i-T} + \sum_{t-t_i < T} a(i)}{\sum_{t-t_i > T} s(i) \alpha^{t-t_i-T} + \sum_{t-t_i < T} s(i)}, \quad (2)$$

其中  $t$  为当前时刻,  $t_i$  为传播发生时刻,  $\alpha, \beta \in (0, 1)$ , 保证了距现在时刻大于  $T$  的节点行为对于影响力的作用弱于较近时刻的行为.

### 2.2 间接影响力及其计算规则

近来社交网络挖掘的关键是其中节点间的信任关系, 其主要传播方式是源传播给好友, 好友再传播给他的好友, 依此串联传播给各个目的节点, 因此考察社交网络影响力的传播时, 除去直接相连

的节点, 还不可避免地遇到间接相连节点的影响力计算问题. 据此, 文中将间接影响力定义为某节点通过中间节点对另一节点的影响力.

间接影响力分为两类: 一类是计算一条路径上影响力的串联传播 (如图 2 所示), 即节点 A 对节点 B 存在一定影响力, 节点 B 对节点 C 存在一定影响力, 则可以得到节点 A 对节点 C 的间接影响力; 另一类是多条路径影响力的聚合 (如图 3 所示), 如假设节点 A 与节点 B 之间存在两条路径:  $p_1$  和  $p_2$ , 则聚合节点 A 通过路径  $p_1$  对节点 B 的影响力与节点 A 通过路径  $p_2$  对节点 B 的影响力, 得到节点 A 对节点 B 的间接影响力.

为计算以上两类间接影响力, 本文引入半环代数结构<sup>[10]</sup>  $(R, \otimes, \oplus)$ .

### 2.2.1 半环代数应用及其合理性

半环代数结构  $(R, \otimes, \oplus)$  中,  $R$  为两两节点之间的影响力的取值集合  $[0, 1]$ ,  $\otimes, \oplus$  是定义在  $R$  上的二元运算, 具有如下性质.

$\otimes$  算子服从结合律, 且存在一个中性元素  $1 \in R$  (即  $\forall I \in R, I \otimes 1 = 1 \otimes I = I$ ) 和一个吸收元素  $0 \in R$  (即  $\forall I \in R, I \otimes 0 = 0 \otimes I = 0$ ).

$\oplus$  算子服从交换律、结合律, 且存在一个中性元素  $0 \in R$  (即  $\forall I \in R, I \oplus 0 = 0 \oplus I = I$ ).

在半环  $(R, \otimes, \oplus)$  上定义偏序关系  $\leq$ , 如果  $\leq$  关于两个算子单调, 那么该半环结构称为有序半环  $(R, \oplus, \otimes, \leq)$ .

基于半环代数的以上性质, 本文中用  $\otimes$  算子来计算一条路径上的间接影响力, 用  $\oplus$  算子来聚合多条路径源节点对于目的节点的影响力. 以下分析应用半环代数来计算间接影响力的合理性.

$\otimes$  算子计算一条路径上的间接影响力时, 如图 2 所示, 则节点 A 对 B 节点的影响力为  $I_{AB} = I_{AM} \otimes I_{MB}$ , 若是节点 A 确认节点 M 完全或完全不接受或转发自己的信息, 则  $I_{AM}$  等于 1 或 0, 那么  $I_{AB} = I_{AM} \otimes I_{MB} = 1 \otimes I_{MB} = I_{MB}$  或  $I_{AB} = I_{AM} \otimes I_{MB} = 0 \otimes I_{MB} = 0$ , 验证了中性元素和吸收元素的存在. 选取合适的运算操作即可保证结合律. 同时, 影响力的传播中一定要保证源节点对于中间节点的影响力为正; 在实际应用中, 应保证该影响力大于一定门限, 否则该路径传播效果微乎其微, 徒增计算复杂度.

$\oplus$  算子用来聚合多条路径上源节点对于目的节点的影响力. 假设图 3 中节点 A 通过路径  $p_1$  对节点 B 的影响力为  $I_1$ , 节点 A 通过路径  $p_2$  对节点

B 的影响力为 0, 则联合得到节点 A 对节点 B 的聚合影响力  $I = I_1 \oplus 0 = 0 \oplus I_1 = I_1$ , 表明存在中性元素. 选取合适的运算操作即可保证结合律和交换律.

影响力的本质——不确定性——保证了偏序关系. 随着单条路径中, 随传播路径增长, 不确定性会单调增大. 根据半环代数的性质, 若满足传播下降和偏序的关系, 则不会形成环路传播.

本文对于社交网络中单条路径或多条路径的间接影响力的计算规则如下.

### 2.2.2 单径计算规则

如下图 2 所示, 其中 A 表示源节点, M 表示中间节点, B 表示目的节点.

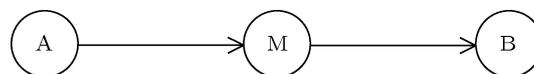


图 2 源节点对于目的节点的影响力串联传播

当节点 A 通过中间节点 M 传播时, 根据半环代数的性质, 可得一条路径上影响力的计算为

$$I_{AB} = R_{AM} \otimes I_{MB}, \quad (3)$$

其中源节点与中间节点的影响力  $R_{AM} \geq I_{\Delta}$ , 其中  $I_{\Delta} > 0$  为影响力传播门限.

具体地, 在社交网络中, 节点对于信息的处理有评论或转发等, 其中评论表示节点接受这则信息, 转发行为则可以促进信息的进一步传播.

### 2.2.3 多径计算规则

与单径传播相比, 多径传播提升同一源节点对于目的节点的影响力. 从直观上理解, 目的节点可以通过多条路径得到消息, 首先是易于得到消息, 其次增加了消息的接受度.

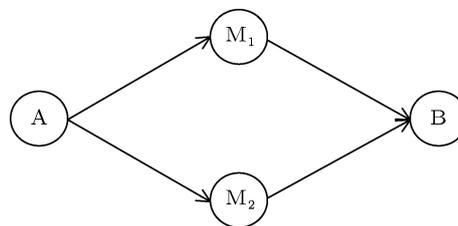


图 3 影响力的多径传播

如图 3 为两条路径的传播影响力, 其中 A 表示源节点,  $M_1$  和  $M_2$  表示中间节点, B 表示目的节点. 多径传播数学表示为

$$I = \bigoplus_i I_{pi} = I_{p1} \oplus I_{p2}. \quad (4)$$

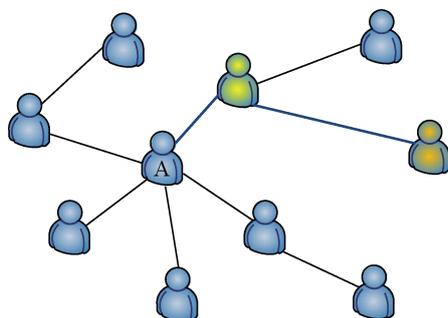
$\oplus$  算子中零元为中性元,  $I = I_1 \oplus 0 = 0 \oplus I_1 = I_1$ , 则增加多条路径传播, 影响力不会减小.

为保证  $\oplus$  算子的交换律和结合律以及简化计算, 取  $\oplus$  算子为  $\max$ , 既能够保证多径传播下的影响力不会降低.

### 3 社交网络中节点综合影响力的计算

基于对影响力的上述理解和计算传播规则, 本节结合社交网络的性质计算源节点的综合影响力.

依据小世界性质<sup>[11,12]</sup>, 社交网络的拓扑结构



使得节点在很大概率上只需要几跳便可连接到网络中的其他节点, 据此本文提出的是一种分布式影响力计算方法, 无需集中管理, 记录节点通过逐跳传播方式与邻近的几跳节点的交互行为, 依此计算出该源节点的影响力.

依据带折扣因子的贝叶斯模型和偏序性质保证无环的半环代数计算源节点与其他节点之间的直接或间接影响力, 然后与影响力传播门限  $I_\Delta$  判断, 保留大于  $I_\Delta$  的连接, 从而得到了有向无环的贝叶斯影响力网络, 如图 4 所示.

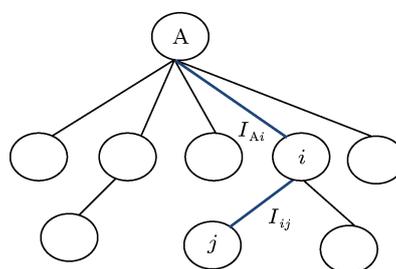


图 4 构建源节点 A 的影响力贝叶斯网络

综合影响力的计算. 根据影响力的性质和计算规则, 可以得出源节点 A 到  $n$  跳范围 ( $n$  的取值根据具体的网络确定, 根据影响力随传播降低的性质, 只需找出有限跳数内的范围即可, 如在微博网络中因为二跳通信的普遍存在选取  $n = 2$ ) 内节点的贝叶斯影响力网络, 然后通过累加即可得到节点的综合影响力

$$I = \sum_{i \in R_A} I_{Ai}, \quad (5)$$

其中  $R_A$  表示在节点 A 根据小世界性质划定的传播范围, 且  $I_{Ai} > I_\Delta$ . 其中根据具体情况选择合适的传播门限  $I_\Delta$  即可计算相应情况下节点的综合影响力, 如网络中个体对于某好电影信息均比较感兴趣时, 可适当降低该电影信息的传播门限; 又比如设定不同的传播门限, 可以估计在不同尺度下该节点的传播扩散能力.

影响力更新. 在节点发出信息后, 观察其下一跳节点的行为, 首先进行相邻节点之间的传播概率更新, 若  $|p_{\text{new}} - p_{\text{old}}| > \Delta p$  ( $\Delta p$  为更新门限) 时, 进行相关路径上的更新.

该更新方法的好处在于只有节点之间发生了信息交互的变化大于一定更新门限时才影响到该

节点对所在的路径, 而不必对每次交互行为进行全局更新, 极大的降低了更新代价.

复杂度分析. 本文算法结合半环代数分布式计算节点的影响力, 每个节点的计算量与节点的平均度数和跳数范围相关, 假设整个网络节点数为  $N$ , 平均度数为  $k_{\text{mean}}$ , 最大度数为  $k_{\text{max}}$ , 跳数为  $l$ , 则一般情况下的复杂度  $O(Nk_{\text{mean}}^l)$ , 最坏情况下的复杂度为  $O(Nk_{\text{max}}^l)$ .

基于结构的算法中, 度中心势算法复杂度较低, 为  $O(E)$ , 其中  $E$  为网络中的边数,  $E = N \times k_{\text{mean}}$ ; PageRank 算法采用迭代法时, 其复杂度为  $O(EL)$ , 其中  $E$  为网络中的边数,  $L$  为迭代次数; LeaderRank 算法复杂度为  $O(E'L)$ , 其中  $E' = E + N$ , PCC 算法首先计算特征向量, 选取其中  $p$  个向量计算到原点的距离, 使得复杂度较高, 为  $O(pN^4)$ .

根据基于半环代数的影响力传播衰减的规则, 依据传播门限可以忽略过小间接影响力, 使得跳数  $l$  可以取值较小. 若取  $l = 1$  则本文算法与度中心势算法复杂度相当; 若取  $1 < l < \log_{k_{\text{mean}}} L + 1$ , 则本文算法复杂度比与度中心势算法高, 而低于 PageRank 和 LeaderRank 算法; 若  $l = \log_{k_{\text{mean}}} L + 1$  则与 PageRank 和 LeaderRank 相当.

## 4 仿真与分析

本节中本文算法与基于结构的节点影响力算法进行了比较. 首先介绍了对比算法与使用的数据集; 然后实验分析虚假粉丝导致的节点影响力波动情况, 并选取一组在线社会网络数据集讨论跳范围和不同门限下的传播范围取值; 最后是依据各方法计算出的影响力降序选择若干数目节点作为传播的初始节点, 比较最终传播到的节点数目.

### 4.1 实验设置

本节中与本文算法进行对比的算法有度中心势方法、PageRank 方法、LeaderRank 方法和 PCC 方法.

本节中使用的数据集包含如下两类: 首先是具有显著结构特性的生成合成数据集——BA 模型<sup>[13,14]</sup>生成的无标度网络, 根据真实网络具备的网络增长和偏好连接显著特征, 生成的网络中节点

的度服从幂率分布的无向虚拟社交网络, 该网络幂指数为 3<sup>[15]</sup>, 与较多实际网络相近; 其次, 考虑到社交网络的多样性, 本文选择了三个典型的实际数据集进行了仿真实验:

**Facebook-like social network 数据集**<sup>[16]</sup>: 该数据集描述加州大学欧文分校学生的在线社会网络. 它是一个有权有向图, 共 1899 个节点, 节点表示学生, 边表示学生之间的信息接受和发送.

**Netscience 数据集**<sup>[17]</sup>: 这是研究网络理论的科学家合作网. 网络中的节点代表论文的作者, 边是每篇文章中作者的合作关系. 该网络共有研究网络理论不同领域的 1589 位科学家.

**Email 数据集**<sup>[18]</sup>: 该数据集描述了西班牙洛维拉·依维尔基里大学员工之间的电子邮件交互网络, 节点表示员工, 边表示员工之间存在电子邮件交互. 通过计算, 给出了各个实验网络数据集的相关信息, 如表 1 所示.

表 1 网络数据集相关参数

网络	节点数	边数	聚类系数	平均路径长度
无标度网络	1000	8000	0.0320	3.1706
Facebook-like social network 数据集	1899	20296	0.0588	Inf
NetScience 网络	1589	2742	0.1892	Inf
Email 网络	1133	10902	0.2200	3.6060

### 4.2 仿真结果分析

#### 4.2.1 虚假粉丝加入前后节点影响力的变化

以微博网络为例, 虚假粉丝共有如下几类: 第一类为无头像、无微博以及无粉丝, 关注对象一半是名人, 一半是同类账户; 第二类虚假粉丝: 有头像、有微博、有粉丝, 但微博数量低于 5 篇, 粉丝数量在 100 个以下, 关注有真实用户, 有虚假粉丝用户; 第三类虚假粉丝: 有头像、有微博、有粉丝, 每天定时更新微博, 粉丝数量 100 到 300 之间, 关注多为真实用户.

虚假粉丝的共性是几乎不参与其他节点的互动, 因此在仿真中假设用户节点之间的互动关系不变, 增加虚假粉丝节点. 这些虚假粉丝节点改变了网络结构, 同时增加虚假粉丝节点与其他节点之间微弱的互动关系. 节点影响力的变化为如图 5 所示:

X 轴表示节点编号, Y 轴表示虚假粉丝情况加入前后节点影响力变化差值.

图 5(a) 和 (d) 分别表示无标度网络和 Email 网络中虚假粉丝加入前后节点影响力的变化, 可以看到度中心势、PageRank、LeaderRank 和 PCC 方法受到了虚假粉丝加入的干扰, 从图 5(a) 的局部图可以清楚的看到这四种方法变化趋势类似, 虚假粉丝的加入引起了产生了节点影响力的较大波动; 图 5(b) 和 (c) 分别表示的 Facebook like 在线社交网络和 NetScience 网络, 其非连通特性使得导致排序不唯一的 PageRank 方法不再适用, 而且 NetScience 网络极度稀疏, 导致网络对应的特征值和特征向量接近于 0, 亦无法应用 PCC 方法; 图 5(b) 中度中心势和 LeaderRank 方法导致部分节点的影响力发生大幅波动, 图 5(c) NetScience 网络节点多而且边极度稀疏, 虚假粉丝对于节点的度的影响较小.

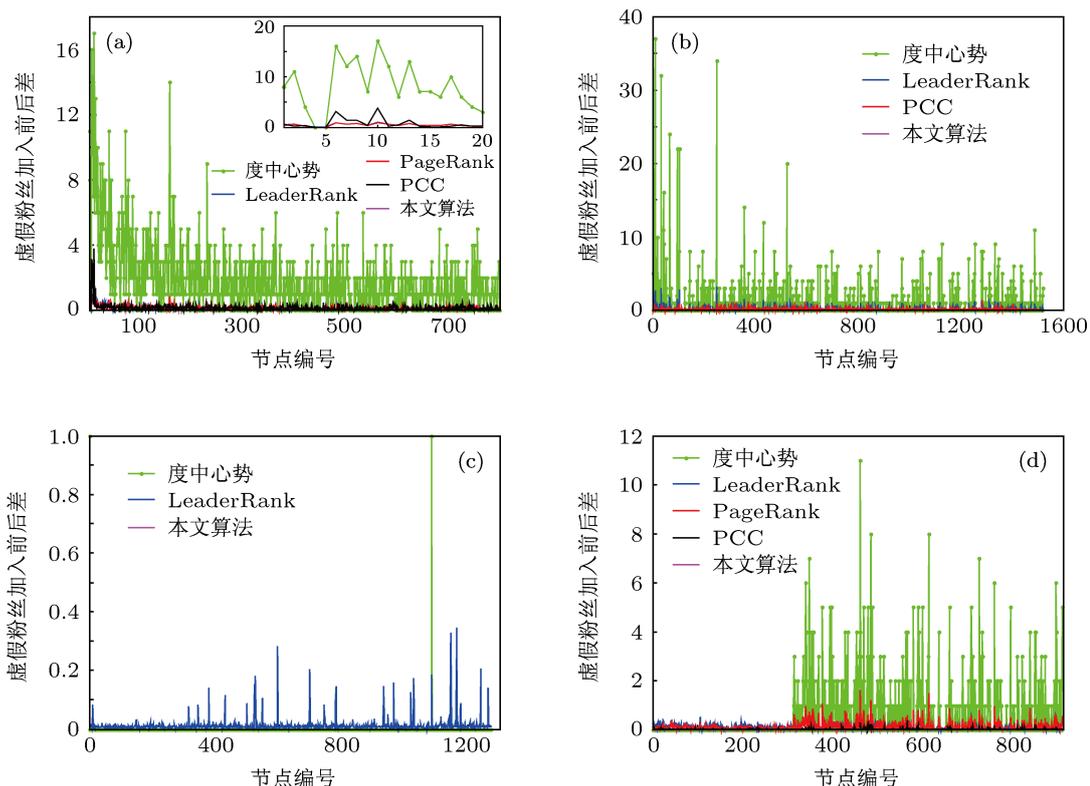


图5 在虚假粉丝加入前后节点影响力的变化 (a) 无标度网络; (b) Face-book like social network 数据集; (c) NetScience 网络; (d) Email 网络

可以看到, 本文的方法在各种类型网络中节点影响力波动小于其他方法. 这是因为与基于结构的方法相比, 本文算法建立在互动的基础上, 由于虚假粉丝缺乏与其他节点之间的互动, 本文虚假粉丝加入前后节点影响力的变化差聚集于  $y = 0$  附近, 表明虚假粉丝加入前后节点的影响力波动很小, 表明本文算法能有效应对虚假粉丝对于节点影响力的干扰.

以下选取 Facebook-like social network 数据集讨论跳范围和不同门限下节点的传播范围.

从图6中可以看到, 该在线社交网络中, 节点的二跳影响力和三跳影响力选取出的前十位 90%一致, 一跳影响力选取出的前十位与两者 70%一致. 在计算影响力时选取二跳, 用较低的计算复杂度就可以获取较高的精度.

图7表示不同门限下各个节点的传播范围. 图7(a)和(b)分别说明随着节点的传播范围随着跳数的增加而增加; 随着传播门限的增加, 节点的传播范围在降低: 图7(a)中单个节点在传播门限为  $T = 0.3$  时的最大传播范围接近 70 个节点, 而图7(b)中单个节点在传播门限为  $T = 0.5$  时的最大传播范围不到 25 个节点.

12	12	12
323	323	323
9	1312	1312
605	9	9
105	105	431
1624	605	105
679	431	605
398	398	1624
297	1624	398
103	341	679

图6 Facebook-like social network 数据集节点的一、二、三跳影响力前十位节点编号

#### 4.2.2 虚假粉丝存在下传播性能分析

在本节中, 我们用串联传播模型<sup>[19]</sup>为信息传播建模. 在该模型中, 节点  $u$  被它的邻居节点  $v$  以权值  $b(u, v)$  进行影响, 同时, 每个节点都有预先设定的阈值  $\theta_u \in [0, 1]$ , 该阈值表示节点  $u$  被它的活跃节点影响的难度, 通常阈值被随机选择为均匀分布. 初始设置一个活跃节点集合, 传播过程以离散方式进行, 当节点在时刻  $t$  成为活跃节点, 它会一直保持活跃. 在每个时刻, 若节点  $u$  与邻居节点之间的权值大于或等于  $\theta_u$ , 那么节点  $u$  被激活.

对于本组仿真, 本文进行了 1000 次独立实验,

并将其结果取平均,如图 6 所示. 其中, X 轴表示初始目标集合的数目, Y 轴表示获得消息的最终节点数目,即在扩散过程末所有激活节点数目.

同 4.2.1 所述, PageRank 不适用于图 5(b) 和 (c) 表示的 Facebook-like social network 数据集和

NetScience 网络, PCC 不适用图 5(c) NetScience 网络. 实验表明, 本文方法相比其他三种方法, 在虚假粉丝存在的情况下, 在上述各种网络中都能获得较大的最终激活节点数目, 较为有效地促进了信息的传播扩散.

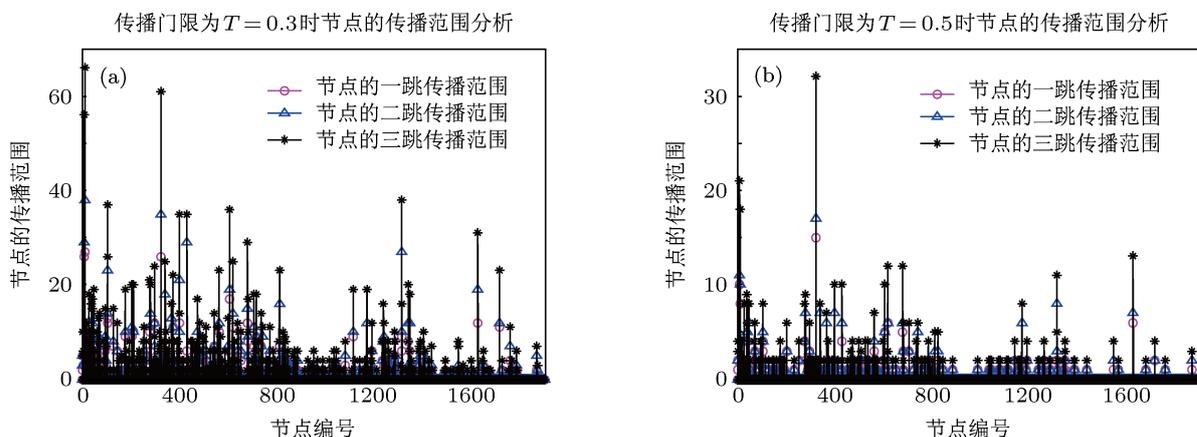


图 7 Facebook-like social network 数据集不同传播门下节点的传播范围

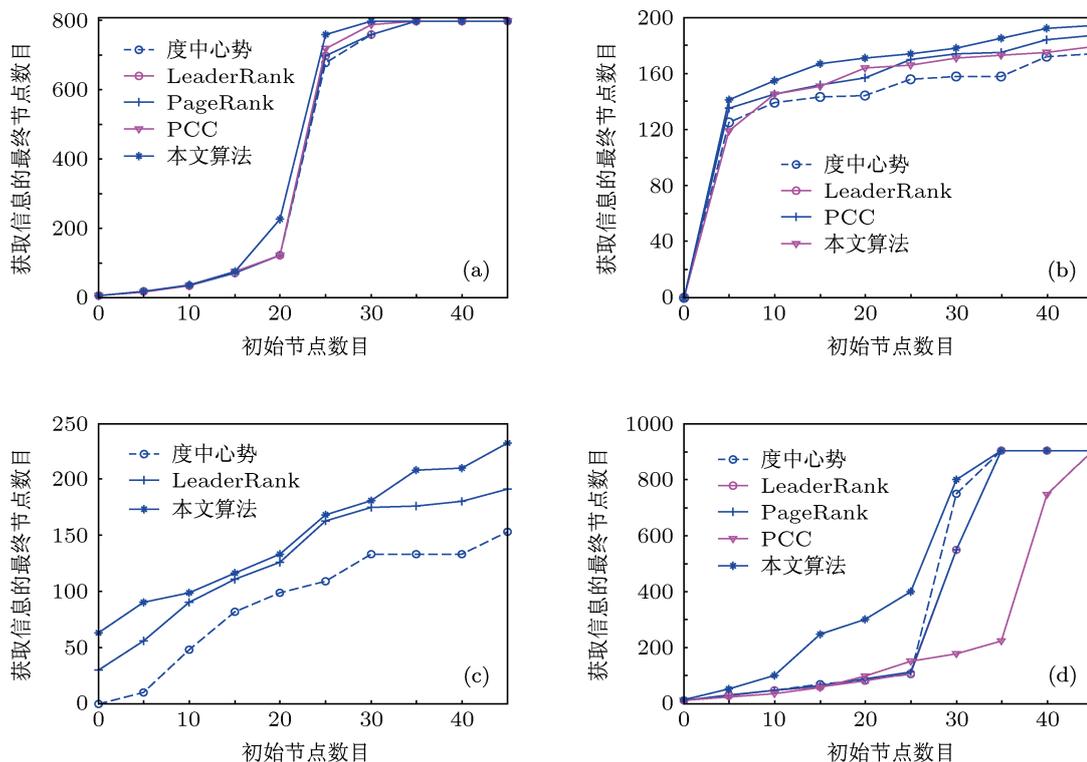


图 8 网络中虚假粉丝存在的传播情况 (a) 无标度网络; (b) Facebook-like social network 数据集; (c) NetScience 网络; (d) Email 网络

在连通拓扑图 8(a) 和 (d) 中, PageRank 和 LeaderRank 算法性能相当, 是因为在连通性良好的网络中, 两种方法排序一致; PCC 的传播性能受到网络拓扑本身的影响较大, 在不同的网络中有较大变化;

度中心势的方法在较稠密的网络中能激活较多节点, 而在连通性差的稀疏网络(如图 8(c))中传播性能严重降低; 本文方法在各种网络中均能以相同数目的初始节点激活较多的节点, 在传播节点数目一

定的情况下可以选择较少的初始节点,表明其传播性能良好.

综上,本文算法能有效地应对虚假粉丝对于节点自身的影响力的干扰,保持影响力的稳定,从而选择合适的初始节点来激活更多节点,促进了信息在虚假粉丝存在情况下的传播扩散.

## 5 结论

社交网络中提升虚假关注的手段如虚假粉丝的出现一定程度上改变了网络结构,严重干扰了基于网络结构的传统节点影响力计算方法,阻碍了对

节点影响力的正确判断.基于此,本文提出了一种基于贝叶斯模型和半环代数的分布式节点影响力计算机理.本文算法仅需通过与相邻几跳节点的通信分布式计算节点影响力;可以根据具体网络状况调整传播门限,计算出相应的节点传播范围.仿真结果表明:本文算法相比现有算法如 PageRank, LeaderRank, PCC 等,能有效地抑制虚假粉丝干扰,保持网络中各节点影响力的稳定,从而选择合适的初始传播节点,获得了更多的最终激活节点,促进了虚假粉丝存在的情况下节点在各种类型网络中的信息传播.

- 
- [1] Carmen Camarero, Rebeca San José 2011 *Computers in Human Behavior* **27** 2292
- [2] Eytan Bakshy, Jake M Hofman, Winter A Mason, Duncan J Watts 2011 *Proceedings of the fourth ACM international conference on Web search and data mining* p65
- [3] Mursel Tasgin, Haluk O Bingol 2012 *Advs. Complex Syst.* **15** 1250061
- [4] Gu Y R, Xia L L 2012 *Acta Phys. Sin.* **61** 238701 (in Chinese) [顾亦然, 夏玲玲 2012 物理学报 **61** 238701]
- [5] Zhou J Y, Zhang Y L, Cheng J 2011 *the 6th International Conference on Frontier of Computer Science and Technology (FCST-11)* p1512
- [6] Newman M E J 2003 *SIAM Review* **45** 167
- [7] Page L, Brin S, Motwani R, Winograd T 1999 *Technical Report Stanford InfoLab* 1999—66
- [8] Lü L Y, Zhang Y C, Yeung C H, Zhou T 2011 *PLoS ONE* **6** e21202
- [9] Ilyas M U, Shafiq M Z, Liu A X, Radha H 2011 *Infocom* p561
- [10] Theodorakopoulos G, Baras J S 2006 *IEEE Journal on Selected Areas in Communications* **24** 318
- [11] Lars Backstrom, Paolo Boldi, Marco Rosay, Johan Ugander, Sebastiano Vigna, 2011 arXiv: 1111.4570
- [12] Watts D J, Strogatz S H 1998 *Nature* **393** 440
- [13] Albert-László Barabási, Réka Albert 1999 *Science* **286** 509
- [14] Wang X F, Li X, Chen G R 2012 *Network Science: An Introduction (1st Edn.)* (Beijing: Higher Education Press) (in Chinese) [汪小帆, 李翔, 陈关荣 2012 网络科学导论 (北京: 高等教育出版社)]
- [15] Guo J L, Wang L N 2007 *Acta Phys. Sin.* **56** 5635 (in Chinese) [郭进利, 汪丽娜 2007 物理学报 **56** 5635]
- [16] Opsahl T, Panzarasa P 2009 *Social Networks* **31** 155
- [17] Newman M E J 2006 *Phys. Rev. E* **74** 036104
- [18] Guimera R, Danon L, Diaz-Guilera A, Giralt F, Arenas A 2003 *Phys. Rev. E* **68** 065103
- [19] Watts D J 2001 *Proceedings of the National Academy of Science of the United States of America* **99** 5766

# Node influence calculation mechanism based on Bayesian and semiring algebraic model in social networks\*

Zhao Jia Yu Li<sup>†</sup> Li Jing-Ru

(Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China)

(Wuhan National Lab of Optoelectronics, Wuhan 430074, China)

(Received 5 January 2013; revised manuscript received 19 March 2013)

## Abstract

In social networks, many applications and spreading depend on the nodes with high influence to do viral marketing, which indicates that nodes' influence should be measured in a comprehensive and reasonable way. The appearance of fake fans results in change of network topology and brings new challenge to topology-based traditional methods. This paper incorporates both the network topology and interactions among nodes into our new distribution mechanism of node influence calculation in social networks. Considering the similarity of node behaviors in time domain and several key factors, this paper presents by a discounted Bayesian model for direct influence between nodes at first. Then a semi-ring-based aggregation implements for indirect influence and the composite influence are obtained by the combination of both direct and indirect influences. Simulation shows that this mechanism not only performs well against fake fans attack and restrains the fluctuation of nodes' influence, but also spreads to more nodes when we choose several nodes with high influence under our method to be source nodes.

**Keywords:** social networks, influence, Bayesian, semiring algebraic

**PACS:** 02.50.Fz, 02.10.H, 89.20.Ff

**DOI:** 10.7498/aps.62.130201

---

\* Project supported by the National Natural Science Foundation of China (Grant Nos. 61231010, 60972016), and the Funds of Distinguished Young Scientists (Grant No. 2009CDA150).

<sup>†</sup> Corresponding author. E-mail: hustlyu@mail.hust.edu.cn