



基于库仑力模型的动态社会网络积极影响力最大化算法

卢敏¹, 陈光鲁¹, 杨晓慧¹, 黄淳岚¹, 乐光学^{1,2}

(1. 江西理工大学理学院, 江西 赣州 341000;

2. 嘉兴学院数理与信息工程学院, 浙江 嘉兴 314001)

摘要: 影响力最大化问题已经成为社会网络中重要的研究内容, 其影响力传播模型和求解算法是关键的核心问题。为了提高预测传播结果的准确度, 引入传播过程中激活节点数量动态变化与节点间信任关系对 IC 模型进行改进, 结合社会影响力与库仑力之间的相似性, 提出一种基于信任关系的动态社会库仑力 (dynamic social Coulomb forces based on trust relationship, DSC-TR) 模型, 构建一种优化的随机贪心 (random greedy, RG-DPIM) 算法求解影响最大化问题。仿真实验结果表明, DSC-TR 模型的预测准确度明显优于 SC-B、IC 模型; RG-DPIM 算法性能优于 G-DPIM、IPA、TDIA 算法。

关键词: 社会网络; 影响最大化; 库仑力; 传播模型; 信任关系

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2020162

Dynamic social network active influence maximization algorithm based on Coulomb force model

LU Min¹, CHEN Guanglu¹, YANG Xiaohui¹, HUANG Chunlan¹, YUE Guangxue^{1,2}

1. College of Science, Jiangxi University of Science and Technology, Ganzhou 341000, China

2. College of Mathematical Information and Engineering, Jiaxing University, Jiaxing 314001, China

Abstract: The problem of maximizing influence has become an important research content in social networks, and its influence propagation model and solving algorithm are the key core issues. In order to improve the accuracy of predicting the propagation results, the dynamic change of the number of activated nodes and the trust relationship between the nodes during the propagation process were introduced to improve the IC model. Combining the similarity between social influence and Coulomb force, a dynamic based on trust relationship was proposed, a dynamic social coulomb forces based on trust relationships (DSC-TR) model was proposed, and an optimized random greedy (RG-DPIM) algorithm was constructed to solve the problem of maximum impact. Simulation results show that the prediction accuracy of the DSC-TR model is obviously better than that of SC-B and IC models. The performance of RG-DPIM algorithm is obviously better than that of G-DPIM, IPA and TDIA algorithms.

Key words: social network, influence maximization, Coulomb force, diffusion model, trust relationship

收稿日期: 2019-07-24; 修回日期: 2020-05-18

通信作者: 乐光学, ygx_9018@163.com

基金项目: 国家自然科学基金资助项目 (No.11704163); 江西省教育厅重点研究项目 (No.GJJ160594)

Foundation Items: The National Natural Science Foundation of China (No.11704163), Key Research Project of Education Commission of Jiangxi Province (No.GJJ160594)



1 引言

社会网络影响力最大化表示在特定的社会网络内利用某种影响力传播模型模拟^[1], 寻找出使得社会网络中传播效果达到最大的若干节点的集合^[2]。社会网络影响力最大化问题成为当前的研究热点, 在广告投放、谣言遏制、网络营销等方面^[3]具有很大的应用价值。

实际社会网络容易受到诸多因素的影响, 导致影响力传播概率并不是固定的。人与人之间的信任与不信任关系将促进或抑制影响力的传播^[4], 而影响力传播的变化则会影响到影响力传播概率, 这与库仑定律中的“同名相斥, 异名相吸”极为相似, 本文利用库仑定律与社会网络的相似关联性来模拟影响传播。

最早 Domingos 等^[5]提出了影响力最大化问题, 后来此问题引起了较大的关注, 有很多研究人员进行了相关的研究, 例如 Kempe 等^[6]首次从离散化的角度解决了如何选择出有影响力的种子节点, 证明了该问题是个 NP 难问题, 提出了独立级联 (independent cascade, IC) 模型^[7]与线性阈值 (linear threshold, LT) 模型^[8], 但真实社会网络信息传播与上述模型的影响力传播机制不完全一样。针对这一不足, 基于 IC 模型和 LT 模型的扩展模型被提出, 比如 Barbieri 等^[9]设计的 TIC 模型, 可以将其应用到对用户间传播概率的计算中。刘勇等^[10]进一步提出了基于主题-兴趣的独立级联 (topic-interest independent cascade, TI-IC) 模型, 针对用户的兴趣在不同主题上的分布与扩散信息的主题分布来计算传播概率, 提高了预测传播结果的准确度。Wang 等^[11]提出了多级姿态的线性阈值 (linear threshold model with multi-level attitude, LT-MLA) 模型, 与传统的线性阈值模型相比, 增加了捕捉用户交互态度状态和能力。Chen 等^[12]提出了新的 IC-N 模型, 该模型与负面意见结合起来, 并引入一个参数来建模人们对产品产生

负面情绪的影响。Hosseini-Pozveh 等^[13]提出了符号感知级联 (sign-aware cascade, SC) 模型, 该模型利用真实社会网络中的不信任关系规定为主动节点对节点执行相反动作或采纳相反意见进行建模。随后, Hosseini-Pozveh 等^[14]将真实社会网络中的不信任关系规定为主动节点对节点不执行相应动作或采纳相应意见, 通过实验证明了真实社会网络中存在不信任关系时, 主动节点对节点往往不再执行相应的动作或采纳相应的意见。Shang 等^[14]依据社区结构构建一种基于多邻居潜力的社区网络影响力最大化 (influence maximization based on multi-neighbor potential in community network, IMPC) 模型, 此模型的应用优势在于较高的全面性, 例如将影响力传播与邻居节点之间的关系考虑在内。Li 等^[15]提出一种新的从众性感知级联 (conformity-aware cascade, C²) 模型, 该模型利用用户从众性之间的相互作用, 从基础数据中获取节点的影响概率, 以此来估计影响力扩散, 在此基础上, 他们还提出了该模型的变体模型 (C³), 在 C² 模型的基础上充分考虑了特定环境中的周围因素。Habiba 等^[16]首次构建动态变化模型, 将时刻进行切片, 将上一时刻的激活节点作为下一时刻的初始节点, 但未考虑节点间的影响力。Wang 等^[17]提出了 LT 模型的扩展模型 LT-A, 针对社会中存在的信任与不信任因素, 将其规定为影响节点为积极节点与消极节点, 根据消极偏移的原则规定被影响节点受消极影响权重偏大, 同时考虑传播过程中影响阈值的变化程度, 并针对该模型提出积极影响力最大化 (positive influence maximization, PIM) 问题。Wang 等^[18]提出了流体扩散 (fluid spread) 模型, 利用流体动力学理论揭示了影响扩散的时间演化过程。谭振华等^[19]提出一种分析谣言在人群中扩散的在线社交网络空间谣言传播分析模型 (GRP model), 将万有引力类比为节点间谣言扩散紧密程度。此外, 传染病模型^[20]以及博弈论模型^[21]等模型也得到了

较多的应用。

Kempe 等^[6]在求解过程中引入了贪心算法,虽然这种方法能够得到近似的结果,但是在计算大型网络时难以达到较高的效率要求,针对此问题,学术界提出了多种改进方法。Leskove 等^[22]依据问题的子模性提出了 CELF 算法,使算法的运行时间缩短为原来的 1/700。Goyal 等^[23]进一步优化了 CELF 算法,提出了 CELF++ 算法,与 CELF 算法相比, CELF++ 算法将效率提高了 35%~55%。马茜等^[2]提出了预选式贪心 (greedy in advance, GIA) 算法和全静态 (all static, AS) 算法。GIA 算法需要先通过一定的方式获取具有较大影响力的节点,然后进一步从中提取特定的种子节点,通过这种方式改善了计算效率; AS 算法通过快照的方式省去重复计算节点影响力的时间,提高了算法效率。另外, Rui 等^[24]提出了 RNR 算法,利用邻居节点影响和反向排名信息来估计当前节点的影响力,当排序达到稳定后,依次选择种子节点。Li 等^[25]提出了分类邻居算法,利用 K-Shell 方法将邻居节点进行分类后计算节点影响力,根据节点影响力选取种子节点。

综上所述,关于社会网络影响力扩散模型^[8-14]的研究成果主要是基于静态模型、以兴趣相似度为主阐述其影响效用,解决促销产品与市场营销场景下的应用问题。Habiba 等^[16]在研究中发现实际的社会网络影响力传播并不是一成不变的,这使得学术界对动态社会网络影响力传播的关注不断增加,与之相关的研究持续增多,并取得了一些成果,参考文献[16-19]分别从时刻切片、阈值动态变化、流体扩散、万有引力角度揭示了影响力传播过程中的动态变化。另外还有很多学者重点关注动态影响力最大化问题^[26],并展开了大量的研究工作。

针对该问题,本文通过对真实社会网络中影响力传播过程的动态变化分析,揭示信任关系在传播过程中所影响激活节点数量发生动态变化的

原因,引入传播过程中激活节点数量动态变化与节点间的信任关系,基于库仑定律与社会网络的相似关联性,提出基于信任关系的动态社会库仑力 (dynamic social Coulomb forces based on trust relationship, DSC-TR) 模型、随机贪心 (random greedy, RG-DPIM) 算法,其核心思想是:结合社会网络中影响力传播的动态变化特征设计了 DSC-TR 模型,基于 DSC-TR 模型提出动态积极影响力最大化 (DPIM) 问题,然后结合 RG-DPIM 算法进行求解。

2 DSC-TR 建模

2.1 节点自身影响力计算

在建模过程中首先需要计算节点自身影响力,其主要表示未激活时节点自身的价值,具体的特性主要体现在如下方面。

- 节点自身影响力大小不受其他节点的影响。
- 节点自身影响力需要达到一定条件,即节点影响力随着其影响范围的增大而变大。
- 各个节点的影响力与节点之间的距离直接相关,如果距离较小,则当前节点会对其他节点产生更大的影响。

节点 i 的自身影响力 α_i 模型为:

$$\alpha_i = \frac{\sum_{h=1}^m \frac{d_i^h}{h}}{\sum_{v \in V} \sum_{h=1}^m d_v^h} \times (\lg |N_i^{\text{out}}| + 1) \quad (1)$$

其中, m 代表传播跳数, d_i^h 代表 i 中 h 层节点的邻居节点数目, N_i^{out} 代表对应的出度。

对于现实社会网络影响力传播来说,影响力在传播两步后基本不会形成明显的影响,所以在本次研究中主要对节点的第二、三层邻居节点进行计算,此时 $m=2$ 。

2.2 节点间社会库仑力计算

基于节点间存在的信任与不信任关系,当节点间为信任关系时,促进影响力进行传播,



称为积极社会库仑力；否则被称作堵塞社会库仑力。

将社会网络节点抽象为电荷，电荷量代表节点自身影响力，电荷之间存在的库仑力代表节点间的社会库仑力，其库仑力方向代表影响传播扩散或抑制。类比结果见表1。

表1 类比结果

库仑力模型	DSC-TR 模型
电荷	节点
电荷量	节点自身影响力
库仑力	社会库仑力
库仑力内外方向	影响传播扩散或抑制

根据以上特征可知，两个节点 i 、 j 的社会库仑力如下所示：

$$F_{ij} = \begin{cases} k \frac{\alpha_i \alpha_j}{r^2} > 0, \text{节点} i \text{与节点} j \text{之间为信任关系} \\ -k \frac{\alpha_i \alpha_j}{r^2} < 0, \text{节点} i \text{与节点} j \text{之间为不信任关系} \end{cases} \quad (2)$$

在式(2)中， F_{ij} 代表两个节点 i 、 j 的社会库仑力， r 表示两节点之间的最短路径， k 表示社会库仑系数。

根据库仑定律可知：如果处于真空内的两点电荷的库仑力是独立存在的，则不会受到其他点电荷的影响。由此可以得到计算邻居节点对当前

节点影响力传播的方法，库仑力叠加计算示例如图1所示，社会库仑力叠加计算示例如图2所示，具体过程如下。

(1) 假设 q_1 、 q_2 、 q_3 都为正电荷，其库仑力 F_{13} 与 F_{23} 叠加后得到的库仑力 F_3 满足 $F_3 = F_{13} + F_{23}$ ，即 v_1 与 v_2 、 v_2 与 v_3 均为信任关系，则 v_1 、 v_2 将促进 v_3 的影响力传播，其社会库仑力满足 $F_3 = F_{13} + F_{23}$ 。

(2) 假设 q_2 、 q_3 都为正电荷， q_1 为负电荷，其库仑力 $-F_{13}$ 与 F_{23} 叠加后得到的库仑力 F_3 满足 $F_3 = -F_{13} + F_{23}$ ，即 v_1 与 v_3 为不信任关系、 v_2 与 v_3 为信任关系，那么此时 v_1 、 v_2 将会对 v_3 的影响力传播分别形成促进、抑制的效果，对应的社会库仑力表示为 $F_3 = -F_{13} + F_{23}$ 。

(3) 假设 q_1 、 q_2 都为负电荷， q_3 为正电荷，其库仑力 $-F_{13}$ 与 $-F_{23}$ 叠加后得到的库仑力 $-F_3$ 满足 $-F_3 = -F_{13} - F_{23}$ ，即 v_1 与 v_3 、 v_2 与 v_3 均为不信任关系，则 v_1 、 v_2 将抑制 v_3 的影响力传播，其社会库仑力满足 $-F_3 = -F_{13} - F_{23}$ 。

2.3 动态影响力传播概率计算

随着影响力传播时间不断增长，激活节点的数量增加，未激活节点所受的社会库仑力叠加之和随之发生动态变化。针对这一现象，结合社会库仑力的叠加计算方法，提出叠加社会库仑力。叠加社会库仑力充分考虑前驱节点的影响力，当

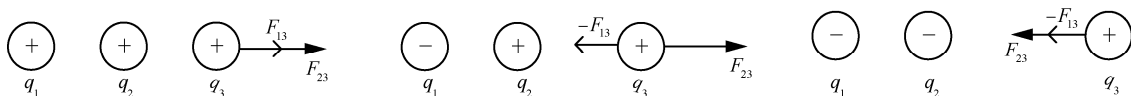


图1 库仑力叠加计算示例

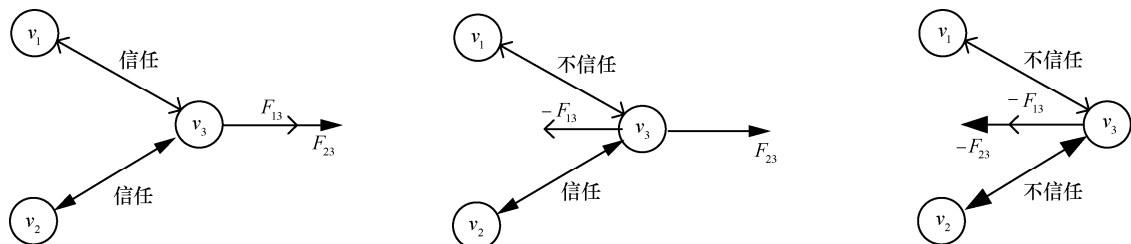


图2 社会库仑力叠加计算示例

前驱节点数量增加或者减少时, 叠加社会库仑力会发生相应的动态变化, 故节点 i 的叠加社会库仑力 β_i 表示为:

$$\beta_i = \sum_{l \in L} F_{li} \quad (3)$$

在式 (3) 中, l 为 i 的前驱节点。

节点 i 的传播影响力表示为 γ_i , 具体形式如下所示:

$$\gamma_i = \alpha_i + \beta_i \quad (4)$$

根据 γ_i 的大小可以将传播影响力划分为积极传播影响力、堵塞传播影响力两种类型, 其中前者需要满足 $\gamma_i > 0$, 后者则满足 $\gamma_i < 0$ 。根据上述分析可知, 节点传播影响力与影响力传播概率必然存在一定的关联, 二者之间会相互影响, 如果积极、堵塞两种传播影响力较大, 那么对应节点激活或者堵塞的概率同样较大。动态影响力初始传播概率如下所示:

$$P_{ij}^t = \frac{\gamma_i}{\sum_{v \in I^t} \gamma_v} \quad (5)$$

在社会网络影响力传播过程中存在激活节点重复叠加的现象, 针对该现象, 将初始时刻后的传播概率改为:

$$P_{ij}^{t+1} = P_{ij}^t + \frac{N_I^{\text{active}}}{N_I} P_{ij}^t \quad (6)$$

在式 (6) 中, I 代表 i 的邻居节点集, N_I^{active} 、 N_I 分别表示 I 内处于激活状态的节点数目与 I 的总节点数目。

2.4 DSC-TR 模型特征分析

社会网络影响力最大化问题的特性体现在传播模型影响力函数的子模性、单调性等方面, 具体的内容如下所示。

定理 1 DSC-TR 模型的影响力最大化问题属于 NP 难问题。

证明 Kempe 等^[6]针对 IC、LT 模型的影响力最大化问题进行了大量的研究, 通过多种方法证

明其属于一个典型的 NP 难问题; 在 IC 模型的基础上进行扩展, 得到了 DSC-TR 模型, 因此可以证明其同样属于 NP 难问题^[4]。

推论 如果某个基础模型下的影响力最大化问题属于一个 NP 难问题, 则改进后的模型依然保持相同的特性, 即仍然属于 NP 难问题。

定理 2 DSC-TR 模型下积极影响力的影响函数不具有单调性与子模性。

证明 要想证明信息传播模型不具有单调性与子模性, 仅需举一反三例即可, 如图 3 所示。

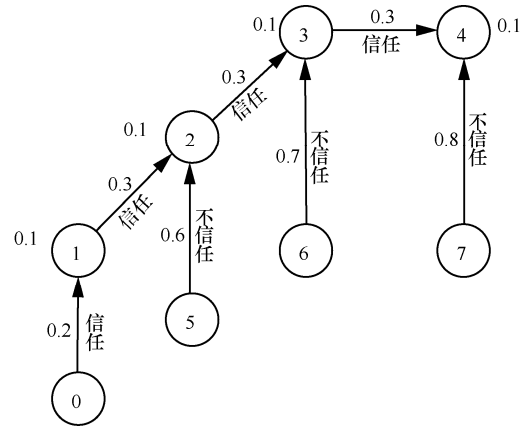


图 3 DSC-TR 模型单调性与子模性证明反例

单调性 初始节点集 S 只包含节点 0, 此时节点 1、2、3、4 被激活, 即 $f_{\text{DSC-TR}}(\{0\}) = 4$, 当节点 7 加入节点集 S 后, 此时节点 1、2、3 被激活, 由于节点 4 被堵塞, 即 $f_{\text{DSC-TR}}(\{0, 7\}) = 3$, 因此, $f_{\text{DSC-TR}}(\{0, 7\}) < f_{\text{DSC-TR}}(\{0\})$, DSC-TR 模型下影响力的积极影响函数不具有单调性。

子模性 集合 T 含有两个节点 0、5, 初始节点集 S 仅含有节点 0, 对于 T 来说, 节点 2 处于堵塞状态, 故只有节点 1 被激活, 即 $f_{\text{DSC-TR}}(\{0, 5\}) = 1$, 当节点 i 为节点 7 时, 则 $f_{\text{DSC-TR}}(\{0, 7\}) - f_{\text{DSC-TR}}(\{0\}) = -1$, $f_{\text{DSC-TR}}(\{0, 5, 7\}) - f_{\text{DSC-TR}}(\{0, 5\}) = 0$ 。由此可以证明 DSC-TR 模型下积极影响力的影响函数没有子模性。

2.5 影响力传播机制

在 DSC-TR 模型传播中, 仅仅是初始节点集



合处于激活状态, 如果将其他的节点激活则能加入后续的传播过程中。但是对其进行堵塞后, 则会结束传播的过程。在算法 1 内, A_1 、 B_1 分别代表初始迭代后的激活节点与堵塞节点集合。具体的描述如下所示。

步骤 1 首先需要对初始节点集 $i \in A_0$ 、节点阈值 θ_v 进行设置, 在此基础上对初始传播概率 P_{ij}^t 进行计算。

步骤 2 检索全图 $G(V, E)$, 满足 $|P_{ij}^t|$ 大于 θ_v 时激活或堵塞, 即 $|P_{ij}^t| > \theta_v$, 将激活节点或堵塞节点分别加入激活或堵塞节点集合, 然后输出得到的激活节点集 A_1 、堵塞节点集 B_1 。

步骤 3 由于步骤 2 的激活行为产生级联的重叠面积较多, 故根据初始迭代激活节点集 A_1 、初始迭代堵塞节点集 B_1 、节点阈值 θ_v , 计算下一时刻传播概率为 P_{ij}^{t+1} 。

步骤 4 检索全图 $G(V, E)$, 依据激活节点生成级联, 需满足 $|P_{ij}^{t+1}| > \theta_v$, 输出激活节点集 A 与堵塞节点集 B 。

步骤 5 每次更新激活或堵塞节点, 重复步骤 4, 直至网络中所有节点不再进行激活或堵塞行为。

算法 1 DSC-TR 模型的影响力传播机制

输入 社会网络图 $G(V, E)$, θ_v , 初始节点集合 A_0

输出 最终激活节点集合 A , 最终堵塞节点集合 B

$A_1 \leftarrow A_0$

$B_1 \leftarrow \emptyset$

initialize θ_v randomly with a value in $[0, 1]$

while (A_0 is not empty)

for $i \in A_0$

compute $P_{ij}^t = \frac{\gamma_i}{\sum_{v \in V} \gamma_v}$

compute $\text{Influ}_{\text{positive}}(A_0)$

solve A_1 、 B_1

end for

for $i \in A_1$

compute $P_{ij}^{t+1} = P_{ij}^t + \frac{N_I^{\text{active}}}{N_I} P_{ij}^t$

compute $\text{Influ}_{\text{positive}}(A_1)$

solve A 、 B

end for

end while

3 DPIM 问题求解算法

3.1 DPIM 问题

DPIM 问题的内容为: 根据 DSC-TR 模型传播机制以及节点的信任关系, 在社会网络图 $G(V, E)$ 内获取节点数目为 k 的种子节点集合 S , 并保证达到积极影响力下激活种子节点数目的最大值, 具体计算式表示为:

$$\text{Influ}_{\text{positive}}(S) = \arg \max_{S \in V', |S|=k} \sigma(S) \quad (7)$$

其中, $\sigma(\cdot)$ 表示影响扩散函数, $\sigma(S)$ 表示在 DSC-TR 模型下被积极影响力激活的节点数量。

针对该问题给出整体求解框架, 如图 4 所示。

3.2 RG-DPIM 算法

RG-DPIM 算法的优势主要体现在效率较高等方面, 由此可以弥补贪心算法的不足, 同时还能够改善计算结果的精度, 这使其得到了更多的关注。在算法 2 中, S' 代表初始节点集合, S'' 代表替换后的节点集合, $\text{Influ}_{\text{positive}}(S')$ 、 $\text{Influ}_{\text{positive}}(S'')$ 均为节点集合, 具体描述如下。

步骤 1 任意选择 k 个种子节点的 S' , 然后对 $\text{Influ}_{\text{positive}}(S')$ 进行计算。

步骤 2 任意选择一个其他的节点 j 取代 S' 内的 i , 然后继续对 $\text{Influ}_{\text{positive}}(S'')$ 进行计算。

步骤 3 将 $\text{Influ}_{\text{positive}}(S')$ 与 $\text{Influ}_{\text{positive}}(S'')$ 进行比较, 如果 $\text{Influ}_{\text{positive}}(S') > \text{Influ}_{\text{positive}}(S'')$, 则保留节点 i ; 如果 $\text{Influ}_{\text{positive}}(S') < \text{Influ}_{\text{positive}}(S'')$, 则保留节点 j 。

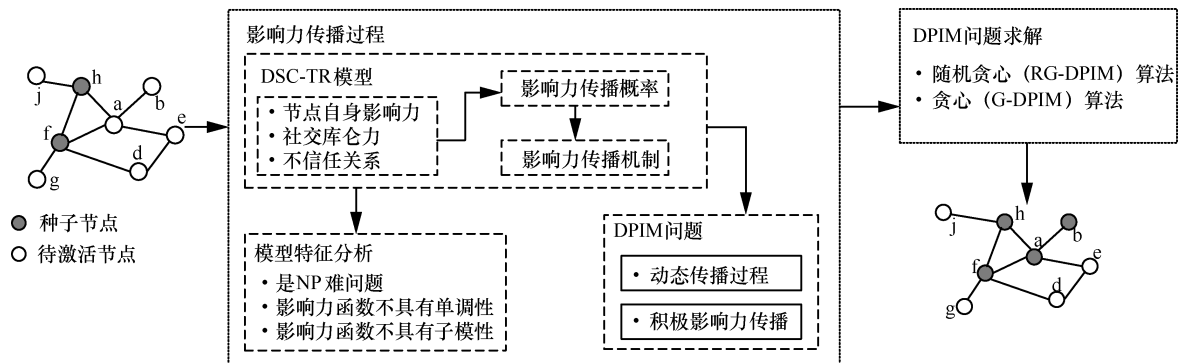


图 4 整体求解框架

步骤 4 重复迭代 N 次, 以增加准确性。

算法 2 随机贪心 (RG-DPIM) 算法

输入 社会网络图 $G(V, E)$, 种子节点数目 k ,

随机次数 N

输出 种子节点集合 S

$S \leftarrow \emptyset$

random select of k nodes

$S' = \{\text{node}i | \text{random select of } k \text{ nodes}\}$

Compute $\text{Influ}_{\text{positive}}(S')$

for Random_{number} = 1 to N

random select a node j in $\bar{V} = V/S'$

random select a node i in S'

node j replaces node i

$S'' = j \cup (S'/i)$

Compute $\text{Influ}_{\text{positive}}(S'')$

if $\text{Influ}_{\text{positive}}(S') > \text{Influ}_{\text{positive}}(S'')$

$S = S'$

else if $\text{Influ}_{\text{positive}}(S') < \text{Influ}_{\text{positive}}(S'')$

$S = S''$

end if

end for

4 实验分析

4.1 数据处理

(1) 数据预处理

进行对比实验, 其中选用的数据集主要有 Bitcoin OTC^[4]、Epinions^[4]。根据信任评分与信任阈值之间的大小关系可以判断是否为信任关系, 如果前者高于后者则属于信任关系, 否则属于不信任关系。其中两种类型的数据集分别见表 2、表 3。

表 2 真实数据集

数据集	节点数/个	边数/条	信任度范围	信任阈值
Bitcoin OTC 数据集	5 881	35 592	-10~10	0
Epinions 数据集	131 828	841 372	1~5	3

(2) 实验环境

实验中所有算法均在 Pycharm Anaconda 2 环境下用 Python 2.7 编写, 所有实验都在 Intel Core i7-7700HQ 2.8 GHz CPU、8 GB 的计算机上运行。

4.2 模型对比分析

本节将 DSC-TR 模型与 SC-B 模型、IC 模型

表 3 实验对比数据集

数据集	组数/组	抽样节点数/个	抽样边/条
Epinions 数据集	1	131 828	841 372
	10	542~834	5 000
Bitcoin OTC 数据集	10	205~436	1 000



进行实验对比。模型对比分析指标分别为精确率 (precision, P)、召回率 (recall, R)、 $F1$ 分数 ($F1\text{-score}$) [10], SC-B 模型、IC 模型传播机制如下。

(1) 包含堵塞节点的信号感知级联 (SC-B) 模型

模型中含有较多的节点, 各个节点的状态只能是堵塞、激活或未激活中的一种, 处于激活状态的节点能够激活或者堵塞其他状态的节点, 由此改变其状态, 但是此过程是单向的。

(2) 独立级联 (IC) 模型

此模型相对于 SC-B 模型缺少了堵塞状态, 每个激活节点能够对未激活节点进行激活操作, 但是此过程是单向的。在这两种模型中, 如果没有节点堵塞或者激活, 则意味着传播过程结束。

已知目标集 A_0 含有 k 个种子节点, 然后在 DSC-TR、SC-B、IC 模型上分别进行了 1 000 次模拟, τ 代表设定的阈值, 然后对各个节点被激活的概率 p_v 进行计算, 如果满足条件 $p_v > \tau$, 则将该节点判定为预测激活节点; 否则属于预测未激活节点。对于真实激活节点或真实未激活节点的判定, 将目标集 A_0 输入 DSC-TR、SC-B、IC 模型上运行得到。本实验中, 阈值 τ 设置为 0.003, 模拟方法是蒙特卡洛方法。

目标集 A_0 的种子节点数目分别设置为 5、10、15、20、25, 最终得到的指标结果如图 5

所示。

可以看出: DSC-TR 模型对比 SC-B 模型精确率、召回率和 $F1$ 分数均上升 10%~15%, DSC-TR 模型与 IC 模型的精确率、召回率和 $F1$ 分数均上升 20%~25%。上述所列举的 3 种模型中, 预测效果最好的是 DSC-TR 模型。

4.3 算法对比分析

算法对比分析前需要确定 DSC-TR 模型阈值 θ_v , 首先将 RG-DPIM 算法与 G-DPIM 算法在不同阈值下进行对比, 其次将 RG-DPIM 算法与 G-DPIM 算法、独立影响路径 (IPA) 算法、基于三度理论的启发式 (TDIA) 算法进行实验对比。G-DPIM 算法、IPA 算法、TDIA 算法的求解过程如下。

(1) 贪心 (G-DPIM) 算法

在传播过程中需要得到激活节点数目最大的节点集合。

(2) 独立影响路径 (IPA) 算法

将影响路径作为影响评价单元, 通过传播概率大于设定阈值 θ 得到传播路径, 将此路径近似为影响函数, 影响路径的传播概率为沿该路径的传播概率的乘积, 其中阈值 θ 的设定值与参考文献[24]中的设定值相等, 即 $\theta = \frac{1}{320}$ 。

(3) 基于三度理论的启发式 (TDIA) 算法

将三度理论与网络拓扑结构相结合构造节

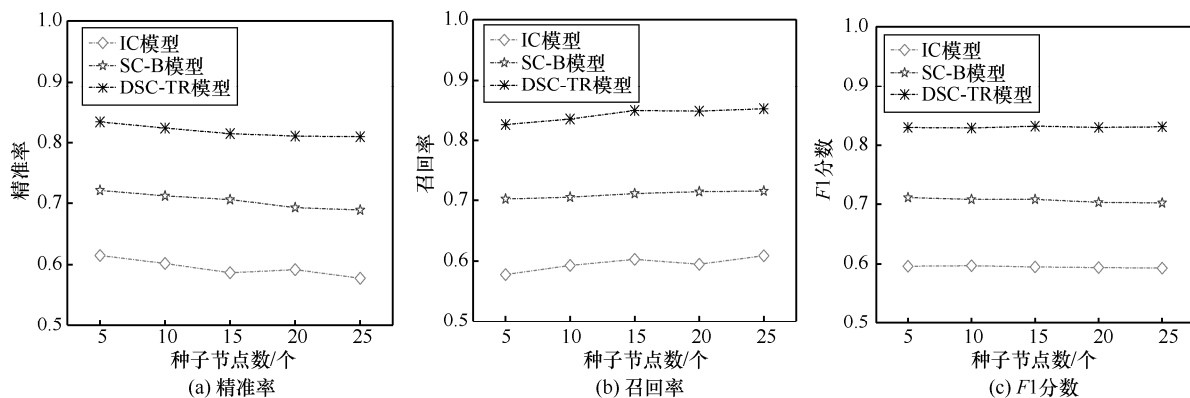


图5 对比模型指标

点影响力函数,模拟传播过程,计算节点影响力后进行排序,选取前 k 个节点组成种子节点集合。

在不同阈值算法的对比实验过程中, $k=5$, 阈值范围取 0.01~0.1。取平均值,以此提升实验结果的准确性与可靠性。实验结果如图 6、图 7 所示。

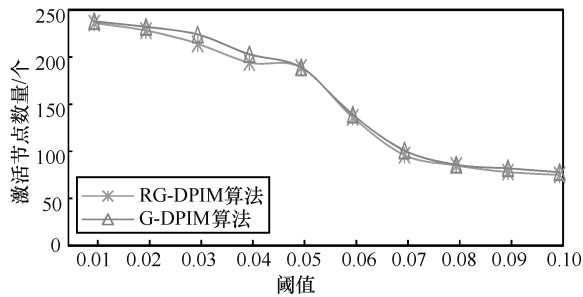


图6 Bitcoin OTC 数据集中不同阈值算法对比

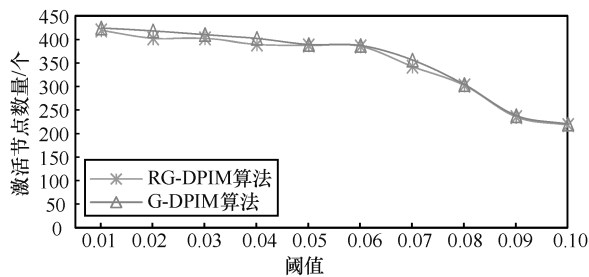


图7 Epinions 数据集中不同阈值算法对比

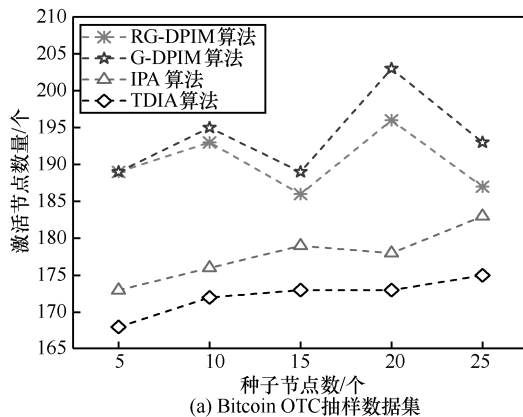
可以看出: RG-DPIM 算法与 G-DPIM 算法在解决 DPIM 问题时,激活节点数量相差不大。图 6 中,当阈值 $0 \leq \theta_v \leq 0.05$ 时,下降较慢;当 $0.05 \leq \theta_v \leq 0.08$ 时,下降较快;当 $0.08 \leq \theta_v \leq 0.10$

时,趋于稳定。在图 7 中,当阈值 $0 \leq \theta_v \leq 0.06$ 时,下降较慢;当 $0.06 \leq \theta_v \leq 0.08$ 时,下降较快;当 $0.08 \leq \theta_v \leq 0.10$ 时,趋于稳定。结果表明:阈值 θ_v 过大会降低影响力传播的敏感度,过小则难以得到显著的传播效果。为方便下述算法对比实验,将阈值 θ_v 确定为 0.05。

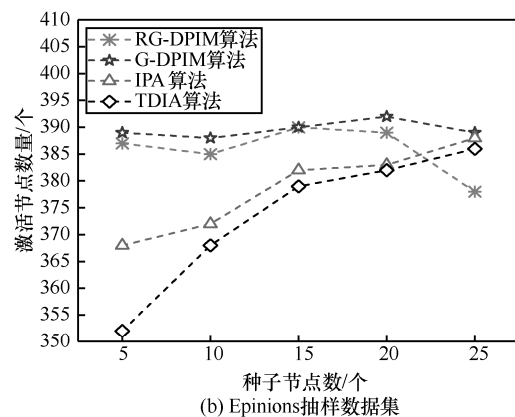
算法对比实验中, RG-DPIM、G-DPIM、IPA、TDIA 算法选取不同数量的种子节点进行模拟仿真,对比指标为:激活节点数量和运行时间。在实验过程中计算各组运行结果的平均值,以此提升实验结果的准确性与可靠性。实验过程中,分别设置 $k=5$ 、10、15、20、25, 阈值 $\theta_v=0.05$ 。实验结果如图 8、图 9 所示。

可以看出:图 8 (a) 中, RG-DPIM 算法与 G-DPIM 算法相差 5 个激活节点左右,与 IPA、TDIA 算法相比,多 20~30 个激活节点。图 8 (b) 中, RG-DPIM 算法与 G-DPIM 算法相差 5 个激活节点左右,与 IPA、TDIA 算法相比,当 $k=5$ 或 10 时, RG-DPIM 算法比 IPA、TDIA 算法多 20~35 个激活节点;当 $k=15$ 、20、25 时, RG-DPIM 算法比 IPA、TDIA 算法多 10 个激活节点左右。结果表明, RG-DPIM 算法与 G-DPIM 算法的激活节点数量相差较小; RG-DPIM 算法的激活节点数量高于 IPA、TDIA 算法。

图 9(a)中, $k=5$ 时, RG-DPIM 算法与 G-DPIM、



(a) Bitcoin OTC抽样数据集



(b) Epinions抽样数据集

图8 激活节点数量实验对比

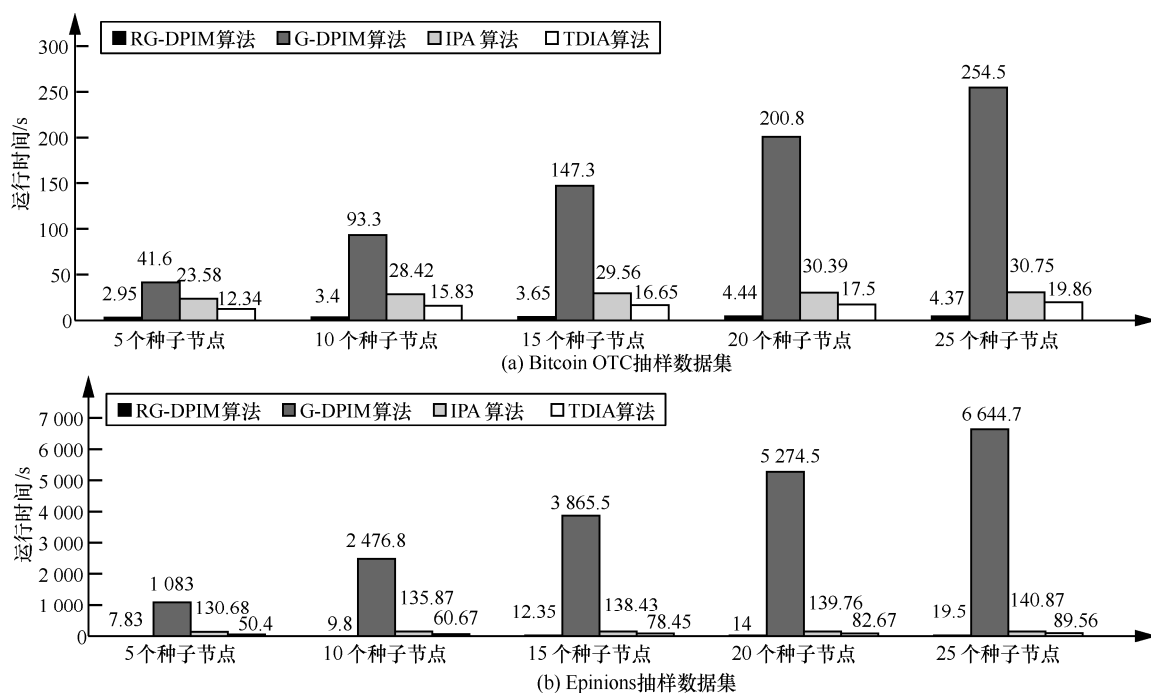


图9 运行时间实验对比

IPA、TDIA 算法相比,运行时间减少了 38.65 s、20.65 s、9.39 s; $k=25$ 时,运行时间分别减少了 250.13 s、26.38 s、15.49 s。图 9 (b) 中, $k=5$ 时, RG-DPIM 算法与 G-DPIM、IPA、TDIA 算法相比,运行时间分别减少了 1 075.17 s、122.85 s、42.57 s; $k=25$ 时,运行时间分别减少了 6 625.2 s、121.37 s、70.06 s。结果表明, RG-DPIM 算法比 G-DPIM、IPA、TDIA 算法运行时间短, RG-DPIM 算法在计算时间复杂度方面上具有明显优势。

5 结束语

本文针对真实社会网络中存在的信任因素与影响力传播过程中的动态变化现象,结合库仑力与社会影响力的相似性构建 DSC-TR 模型,对该模型采用 RG-DPIM 算法求解 DPIM 问题。实验结果表明: DSC-TR 模型预测准确度明显优于 SC-B、IC 模型; RG-DPIM 算法可以在保证影响范围与 G-DPIM、IPA、TDIA 算法相差不多的情况下,更高效地解决 DPIM 问题。在未来的研究工作中,将结合用户自身属性和用户间综合影响因素构建

模型,提出更有效的适用社会网络影响力的传播模型。

参考文献:

- [1] 韩忠明, 陈炎, 刘雯, 等. 社会网络节点影响力分析研究[J]. 软件学报, 2017, 28(1): 84-104.
HAN Z M, CHEN Y, LIU W, et al. Research on node influence analysis in social networks[J]. Journal of Software, 2017, 28(1): 84-104.
- [2] 马茜, 马军. 在影响力最大化问题中寻找种子节点的替补节点[J]. 计算机学报, 2017, 40(3): 674-686.
MA Q, MA J. Discovering the substitutes for the seeds in influence maximization problem[J]. Chinese Journal of Computers, 2017, 40(3): 674-686.
- [3] 据春华, 鲍福光, 许翀寰. 基于社会网络协同过滤的社会化电子商务推荐研究[J]. 电信科学, 2014, 30(9): 82-86.
JU C H, BAO F G, XU C H. Research on social network collaborative filtering based e-commerce recommending[J]. Telecommunications Science, 2014, 30(9): 82-86.
- [4] HOSSEINI-POZVEH M, ZAMANIFAR K, NAGHSH-NILCHI A R. Assessing information diffusion models for influence maximization in signed social networks[J]. Expert Systems With Applications, 2019, 119(4): 476-490.
- [5] DOMINGOS P, RICHARDSON M. Mining the network value of customers[C]//Proceedings of the 7th ACM SIGKDD Inter-

- national Conference on Knowledge Discovery and Mining. New York: ACM Press, 2001: 57-66.
- [6] KEMPE D, KLEINBERG J, TARDOS E. Maximizing the spread of influence in a social network[C]//Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2003: 137-146.
 - [7] GOLDENBERG J, LIBAI B, MULLER E. Using complex systems analysis to advance marketing theory development: modeling heterogeneity effects on new product growth through stochastic cellular automata[J]. *Academy of Marketing Science Review*, 2001, 9(3): 1-18.
 - [8] GRANOVETTER M. Threshold models of collective behavior[J]. *American Journal of Sociology*, 1987, 83(6): 1420-1443.
 - [9] BARBIERI N, BONCHI F, MANCO G. Topic-aware social influence propagation models[J]. *Knowledge and Information Systems*, 2013, 37(3): 555-584.
 - [10] 刘勇, 谢胜男, 仲志伟, 等. 社会网中基于主题兴趣的影响最大化算法[J]. *计算机研究发展*, 2018, 55(11): 2406-2418
LIU Y, XIE S N, ZHONG Z W, et al. Topic-interest based influence maximization algorithm in social networks[J]. *Journal of Computer Research and Development*, 2018, 55(11): 2406-2418.
 - [11] WANG F, WANG G J, XIE D Q. Maximizing the spread of positive influence under LT-MLA model[J]. *Advance in Services Computing*, 2016, 10(11): 450-463.
 - [12] CHEN W, COLLINS A, CUMMINGS R, et al. Influence maximization in social networks when negative opinions may emerge and propagate[C]//Proceedings of 2010 SIAM International Conference on Data Mining. Columbus: SIAM Press, 2010: 379-390.
 - [13] HOSSEINI-POZVEH M, ZAMANIFAR K, NAGHSH-NILCHI A R, et al. Maximizing the spread of positive influence in signed social networks[J]. *Intelligent Data Analysis*, 2016, 20(1): 199-218.
 - [14] SHANG J X, WU H C, ZHOU S B, et al. IMPC: influence maximization based on multi-neighbor potential in community networks[J]. *Physica A: Statistical Mechanics and Its Applications*, 2018, 512(15): 1085-1103.
 - [15] LI H, BHOWMICK S S, SUN A, et al. Conformity-aware influence maximization in online social networks[J]. *The VLDB Journal*, 2014, 5(8): 117-141.
 - [16] HABIBA, BERGER-WOLF T Y. Influence maximizing the extent of spread in a dynamic network[C]//Proceedings of International Conference on Knowledge Discovery and Data Mining. Berlin: Springer-Verlag, 2007: 542-550.
 - [17] WANG H H, YANG Q, LEI F, et al. Maximizing positive influence in signed social networks[C]//Proceedings of International Conference on Cloud Computing and Security. Berlin: Springer International Publishing, 2015: 356-357.
 - [18] WANG F, JIANG W J, LI X L, et al. Maximizing positive influence spread in online social networks via fluid dynamics[J]. *Future Generation Computer Systems*, 2018, 86(9): 1491-1502.
 - [19] 谭振华, 时迎成, 石楠翔, 等. 基于引力学的在线社会网络空间谣言传播分析模型[J]. *计算机研究与发展*, 2017, 54(11): 2586-2599.
TAN Z H, SHI Y C, SHI N X, et al. Rumor propagation analysis model inspired by gravity theory for online social networks[J]. *Journal of Computer Research and Development*, 2017, 54(11): 2586-2599.
 - [20] GUIZANI N, ELGHARIANI A, KOBES J, et al. Effects of social network structure on epidemic disease spread dynamics with application to ad hoc networks[J]. *IEEE Network*, 2019, 33(3): 139-145.
 - [21] GRIFFIN C, RAJTMAYER S, SQUICCIARINI A, et al. Consensus and information cascades in game-theoretic imitation dynamics with static and dynamic network topologies[J]. *SIAM Journal on Applied Dynamical Systems*, 2019, 18(2): 597-628.
 - [22] LESKOVE C J, KRAUSE A, GUESTRIN C, et al. Cost-effective outbreak detection in networks[C]//Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2007: 420-429.
 - [23] 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 Conference Companion on World Wide Web. Hyderabad: Association for Computing Machinery, 2011: 47-48.
 - [24] RUI X, MENG F, WANG Z, et al. A reversed node ranking approach for influence maximization in social networks[J]. *Applied Intelligence*, 2019, 49(7): 2684-2698.
 - [25] LI C, WANG L, SUN S, et al. Identification of influential spreaders based on classified neighbors in real-world complex networks[J]. *Applied Mathematics and Computation*, 2018(320): 512-523.
 - [26] 潘剑飞, 徐丽丽, 董一鸿. 动态社区演化研究进展[J]. *电信科学*, 2017, 33(1): 24-33.
PAN J F, XU L L, DONG Y H. Research progress of dynamic community evolution[J]. *Telecommunications Science*, 2017, 33(1): 24-33.
 - [27] KIM J, KIM S K, YU H. Scalable and parallelizable processing of influence maximization for large-scale social networks[C]//Proceedings of the 29th International Conference on Data Engineering. Piscataway: IEEE Press, 2013: 266-277.



[28] 王欢欢. 社交网络影响力最大化的研究[D]. 南京: 南京航空航天大学, 2016.

WANG H H. Research on influence maximization in social networks[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2016.

[作者简介]



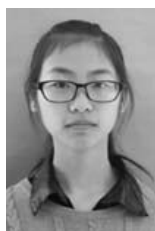
卢敏 (1964—), 男, 江西理工大学理学院教授, 主要研究方向为网络通信及电子材料器件等。



陈光鲁 (1995—), 男, 江西理工大学理学院硕士生, 主要研究方向为社会网络和数据挖掘。



杨晓慧 (1996—), 女, 江西理工大学理学院硕士生, 主要研究方向为边缘计算。



黄淳岚 (1997—), 女, 江西理工大学理学院硕士生, 主要研究方向为移动云计算。



乐光学 (1963—), 男, 嘉兴学院数理与信息工程学院教授, 主要研究方向为多云融合与协同服务、无线 mesh 网络与移动云计算、混成与嵌入式系统。