

# 多轮次影响传播下的增益节点成本最小化动态策略

张龙姣<sup>1</sup>, 付冰洋<sup>1</sup>, 史麒豪<sup>1,2</sup>, 宋明黎<sup>1</sup>, 王 灿<sup>1</sup>, 章 悦<sup>3\*</sup>

(1. 浙江大学 计算机科学与技术学院, 浙江 杭州 310027; 2. 浙大城市学院 计算机与计算科学学院, 浙江 杭州 310015;  
3. 浙江省平安建设大数据重点实验室, 浙江 杭州 310016)

**摘要:**为了减少商家在社交网络上进行多轮次商品推广的营销成本,针对多轮次影响力传播过程中的增益节点选择问题展开研究。基于多轮次影响增益传播模型,提出了自适应的增益节点选择策略,该策略在已知种子节点的前提下,能够在近似线性的算法复杂度下,找到最小化达到传播影响阈值所需的营销轮次的近似策略。实验结果表明,相较于现有启发式算法和非自适应算法,所设计的自适应策略能够减少7.3%~18.3%达到指定阈值所需的传播轮次,有效减少推广成本。

**关键词:** 社交网络; 影响力传播增益; 影响力最大化; 增益节点; 成本最小化

中图分类号: TP301 文献标志码: A 文章编号: 1001-2486(2025)03-021-11



论文  
拓展

## Adaptive strategy for boosting node costs minimization in multi-round influence

ZHANG Longjiao<sup>1</sup>, FU Bingyang<sup>1</sup>, SHI Qihao<sup>1,2</sup>, SONG Mingli<sup>1</sup>, WANG Can<sup>1</sup>, ZHANG Yue<sup>3\*</sup>

(1. College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China;

2. School of Computer and Computing Science, Hangzhou City University, Hangzhou 310015, China;

3. Zhejiang Provincial Key Laboratory of Social Security Governance Big Data, Hangzhou 310016, China)

**Abstract:** In order to reduce the marketing costs of merchants promoting products over multiple rounds on social networks, this study made an exploration on the selection of boosting nodes during the process of multi-round influence propagation. Based on the model of multi-round influence boosting propagation mode, an adaptive strategy for choosing boosting nodes was designed. Given known seed nodes, this strategy could find an efficient method to minimize the number of marketing rounds needed to reach a certain threshold of social influence, with nearly linear algorithmic complexity. Experimental results show that compared to existing heuristic algorithms and non-adaptive algorithms, the designed adaptive strategy can reduce the promotion rounds required to reach a specified threshold by 7.3%~18.3%, effectively reducing the promotion cost.

**Keywords:** social network; boosting influence spread; influence maximization; boosting node; costs minimization

近年来,以 Facebook、Twitter、微博为代表的社交网络平台发展迅速。社交网络通过互联网连接用户关系,为信息的传播提供了快速便捷的途径。在社交网络上,信息可以通过好友间“口耳相传”的效应进行传播。用户可以通过社交网络平台进行消息的发布和转发。具体来说,发帖人发布新的消息后,他的粉丝在一定概率下会进行转发,新的消息会沿着社交网络上的好友关系传播以获得更多用户的关注,从而产生一定的社会

影响力。

舆论信息传播和营销产品推广都能够利用社交网络的传播效应。因此,在线社交网络中的信息传播过程已经吸引了多个领域的广泛研究,包括计算机科学、物理学、流行病学<sup>[1]</sup>。以广告投放为例,商家或广告商在社交网络上进行广告投放时,通常选择具有大量粉丝的账号作为广告的初始发布人,这些账号发布的消息能够在短时间内被大量用户看到并进一步转发,这种连锁传播

收稿日期:2023-12-06

基金项目:国家自然科学基金资助项目(62372399);浙江大学上海高等研究院繁星科学基金资助项目(SN-ZJU-SIAS-001)

第一作者:张龙姣(2000—),女,河北保定人,博士研究生,E-mail:zhlljoan@zju.edu.cn

\*通信作者:章悦(1984—),男,浙江杭州人,高级工程师,硕士,E-mail:232436610@qq.com

引用格式:张龙姣,付冰洋,史麒豪,等.多轮次影响传播下的增益节点成本最小化动态策略[J].国防科技大学学报,2025,47(3):21-31.

Citation:ZHANG L J, FU B Y, SHI Q H, et al. Adaptive strategy for boosting node costs minimization in multi-round influence[J]. Journal of National University of Defense Technology, 2025, 47(3): 21-31.

效应使得广告信息能够在短时间内达到更多的用户,从而提高产品或服务的知名度和销售量。以微博为代表的社交平台上,广告的转发量可以反映产品在社会上的影响力。选择有“影响力”的初始发布人、提升中间用户的转发积极性都可以扩大广告的影响力。

传播影响最大化和传播成本最小化是信息传播研究中的两个关键问题,它们互为对偶问题。Domingos 和 Richardson 在市场病毒式营销策略的启发下,提出了传播影响最大化问题<sup>[2]</sup>。假设已知社交网络中的数据,并能够估计用户之间互相影响的概率,商家可以向网络中少数具有“影响力”的用户提供免费试用服务,由他们开始向朋友推荐产品,触发影响级联效应,从而使产品或服务能够被网络中最大期望数量用户所采纳。影响力的传播过程基于特定的传播模型,最为主流的模式包括线性阈值模型、独立级联模型和触发模型等<sup>[3]</sup>。在这些模型中,最初被选中进行信息发布的用户通常被称为种子节点。在信息传播过程中,受到种子节点影响并进行转发信息的用户则被称为激活节点。在传播的开始阶段只有种子节点处于激活状态。信息从这些种子节点出发根据特定的传播模型规则进行扩散。在这个过程中,成功被激活的节点数量被视为此次信息传播所产生的影响力大小。影响力最大化问题旨在设计种子节点的选择策略,使得传播结束后,激活节点数量的期望值最大化。在大部分传播模型下,影响力最大化问题是一个非确定性多项式时间(nondeterministic polynomial time, NP)难问题,Kempe 等<sup>[3]</sup>在线性阈值模型和独立级联模型上,利用影响函数的单调性和子模性为贪心策略给出了近似性证明。Borgs 等<sup>[4]</sup>通过采样反向可达集优化蒙特卡罗算法模拟传播实例的耗时。Tang 等<sup>[5-7]</sup>在采样反向可达集的基础上进一步缩小了所需采样数量的下界。

但这些工作主要针对单独节点的影响力,没有考虑节点间的助推作用和社群结构的影响。近年来,一些工作基于图上的社群结构检测方法对传统近似算法进行了模型优化和效率优化<sup>[8-9]</sup>。一部分工作改进了传统的影响力传播模型,对于实际营销过程中可能发生的产品助推<sup>[10]</sup>、产品竞争<sup>[11]</sup>、节点助推<sup>[12-13]</sup>进行了建模。Lin 等<sup>[12]</sup>提出了增益节点的概念。除了与网络中具有“影响力”的用户合作外,商家可以采用发放优惠券或定向推送广告的方式提升一部分用户的消费意愿。这部分用户相比于受到激励前,在收到好友

的消费推荐后会更愿意购买并向其他好友推荐这个产品。如何选择激励目标用户成为一个值得研究的问题。可惜的是目前的工作仅考虑了单轮次传播过程中的增益节点选择问题。

在实际的营销活动中,商家会进行多轮次的营销推广。每轮次的推广由不同“有影响力”的用户群发起,商家会向一部分用户发放优惠券,提高他们的消费意愿。当不再有新消费的用户群体增长时,当前轮次的传播过程结束,商家则会开启下一轮的消费激励政策。Sun 等<sup>[14]</sup>对多轮次的病毒式营销过程进行建模,为每轮次的传播选择种子节点集合,以最大化多轮次传播结束后激活节点的期望数量。设计了多轮次触发模型,并采用跨轮次贪心策略和自适应轮次内贪心策略计算多轮次影响最大化的近似解。Tang 等<sup>[15]</sup>观察已选择的种子节点对其他用户的影响,自适应地选择下一轮传播的起始种子节点,解决了自适应种子最小化问题。Niknami 等<sup>[16]</sup>针对多轮次影响竞争问题进行探讨。

由于社交网络的动态性和社交网络传播的不确定性,仅采用单轮次的营销策略可能会浪费资源,重复激励已经购买过商品的非用户群,错过未发生消费的潜在用户群。多轮次的营销过程可以根据传播现状动态分配资源选择下一轮次的增益目标,从而更有效地提升传播影响力、节省营销成本。目前针对增益节点的研究主要集中在单轮次的营销活动中,没有充分考虑到在多轮次的营销活动中如何选择增益节点。为了能够最小化多轮次传播过程中所需要的推广成本,对于多轮次增益传播过程进行建模,并采用自适应选择算法根据每次的传播结果,动态选择下一轮的增益节点,从而获得更精确的结果,以避免静态策略产生的重复目标推广成本。这一问题属于 NP 难问题,研究提出了具有近似下界保证的优化选择策略,并且只需要近似线性的时间复杂度。通过在真实数据集<sup>[17-19]</sup>上的实验,相较于现有启发式算法<sup>[12,20]</sup>和非自适应算法<sup>[12]</sup>,所设计的自适应策略能够减少 7.3% ~ 18.3% 达到指定阈值所需的传播轮次。这种基于动态策略的增益节点成本最小化方法不仅有理论价值,也具有实际应用价值。

## 1 问题模型与定义

本节针对多轮次影响增益模型(multi-round influence boosting model, MIBM)和传播影响阈值限制下的增益节点成本最小化问题的定义进行说明。

## 1.1 信息传播模型

多轮次影响增益模型在传统影响增益模型<sup>[12]</sup>的基础上引入多轮次传播<sup>[14]</sup>的性质,即每轮从不同种子节点开始传播,提升不同的节点作为增益节点,所有轮次传播结束后的激活节点的数量即为多轮次影响增益模型的传播影响力。

### 1.1.1 影响增益模型

影响增益模型<sup>[12]</sup>是在独立级联模型的基础上提出的,该模型在影响增益最大化问题中被广泛应用。

在影响增益最大化的问题定义中,当一个节点被提升为增益节点后,更容易受到邻居节点的影响。给定社交网络  $G = (V, F)$ , 其中  $V$  表示社交网络中的节点集合,  $F$  表示社交网络中的边集合, 对于网络中的任一边  $e_{uv} \in F$  存在两个权值  $p_{uv}$  和  $p'_{uv}$  ( $p'_{uv} > p_{uv}$ ), 其中  $p_{uv}$  表示节点  $v$  受到节点  $u$  传播影响的原始概率。当  $v$  被提升为增益节点后, 受到邻居节点影响的概率会提升。 $p'_{uv}$  表示  $v$  被提升为增益节点后受到邻居节点  $u$  的传播影响的概率。

传播开始前, 选定种子集合  $S$  和增益节点集合  $B$ 。初始时刻只有种子节点处于被传播激活的状态, 其他节点均未被激活。在每个离散时刻  $t$ , 如果节点  $v$  不属于激活节点集合, 且与  $t-1$  时刻被激活的节点  $u$  存在连接关系  $e_{uv}$ , 那么节点  $u$  会尝试将信息传播给节点  $v$ , 若节点  $v$  不属于增益节点集合  $B$ , 那么被节点  $u$  激活的概率是  $p_{uv}$ , 如果  $v$  属于增益节点  $B$ , 节点  $v$  被节点  $u$  激活的概率提高为  $p'_{uv}$ 。每个被激活的节点仅有一次激活邻居节点的机会, 节点被激活后会保持激活的状态, 当没有更多新节点可以被激活时, 传播过程停止。信息传播的影响力可以表示为最终处于激活状态的节点数量。

### 1.1.2 多轮次影响增益模型

多轮次影响增益模型会进行多轮次的传播, 每轮次信息会从不同的种子集合开始扩散。在传播开始前给定一个备选种子集合  $\hat{S}$ , 其中每个节点  $s \in \hat{S}$ , 都存在一个概率  $p(s)$  表示每轮次节点  $s$  被选作种子节点传播信息的概率。在第  $t$  传播轮次开始前, 给定本轮次增益节点集合  $B_t$ , 按照种子概率分布生成本轮传播种子集合  $S_t \in \hat{S}$ , 按照影响增益模型在社交网络上进行传播, 定义第  $t$  轮次激活节点的集合为  $\rho_{\phi(t)}(B_t)$ ,  $\phi_t$  表示在第  $t$  轮次时的传播过程。不同轮次中可能激活相同的节点, 这些节点可以作为传播介质, 再次将信息扩散

到邻居节点, 但不重复计入传播影响力。传播过程是随机的, 对于一个多轮传播过程  $\phi$ , 经过  $T_n$  轮次传播后, 最终激活节点的集合可以表示为

$$\rho_{\phi}(U) = \rho_{\phi}(B_1, B_2, \dots, B_{T_n}) = \left| \bigcup_{t=1}^{T_n} \rho_{\phi_t}(B_t) \right| \quad (1)$$

式中,  $U$  表示  $T_n$  轮次中总共选择的增益节点的集合,  $B_t$  表示第  $t$  传播轮次中选择的增益节点集合。

**定义 1** 多轮次影响增益模型传播影响力。

在已知备选种子集合  $\hat{S}$  的情况下, 增益节点集合  $U$  在传播实例  $\phi$  上助推传播影响力的定义为

$$I_{\phi}(U) = I_{\phi}(B_1, B_2, \dots, B_T) = \left| \rho_{\phi}(B_1, B_2, \dots, B_T) \right| \quad (2)$$

## 1.2 MIBM 模型下的增益节点成本最小化问题

给定社交网络  $G = (V, F)$ , 同时给定一个初始传播用户备选种子集合  $\hat{S}$  和传播影响阈值  $\eta \in [1, N]$ , 其中  $N$  表示从备选种子集合出发可能影响到的节点数量。每轮次传播开始前, 选择  $b$  个节点作为这一轮次的增益节点, 目标是设计一种多轮次增益节点选择策略, 这种策略能够应对任意随机出现的信息传播实例, 并在尽可能少的时间轮次内达到预设的传播影响阈值, 从而使总体期望的传播轮次最小。

通过公式表示这一过程, 增益节点成本最小化期望找到一个策略  $\pi$ , 对任一出现的传播实例  $\phi$ , 均有  $I_{\phi}[U(\pi, \phi)] \geq \eta$  且使得期望传播轮次最小化, 即

$$\min_{\pi} \mathbb{E}[T(\pi, \phi)] \quad \text{s. t.} \quad I_{\phi}[U(\pi, \phi)] \geq \eta \quad (3)$$

$$\mathbb{E}[T(\pi, \phi)] = \sum_{\phi \in \Phi} T(\pi, \phi) \cdot p(\phi) \quad (4)$$

其中,  $U(\pi, \phi)$  表示按照  $\pi$  策略在实际传播实例  $\phi$  下选择的增益节点总集合,  $T(\pi, \phi)$  表示在实例  $\phi$  下按照策略  $\pi$  选择增益节点达到传播影响阈值时所需要的传播时间轮次。

## 2 算法设计

针对增益节点成本最小化问题设计, 本节提出了自适应贪心增益节点选择策略 (adaptive greedy multiply boosting nodes selection strategy, AdaMBTS) 和对应的下界选择策略 (lower bound of AdaMBTS, AdaMBTS-LB)。

### 2.1 自适应增益节点集合选择策略

解决增益节点成本最小化问题的一种简单的策略是: 在传播开始前, 估计每个节点的传播增

益,贪心地选择传播增益最大的节点。然而,固定的增益节点选择可能会导致助推的用户群重复,造成不必要的成本浪费。因此,采用自适应增益节点集合选择策略。在每轮传播前,先观察前几轮的传播结果,然后根据这些结果来选择下一轮次的增益节点集合。

### 2.1.1 符号与定义

由于传播过程是随机的,采用期望传播影响作为衡量增益节点集合传播效果的重要参数。在有阈值限制的问题中,仅要求最终的传播影响大于  $\eta$  即可,首先定义阈值传播影响函数

$$\Gamma_{\phi}(U) = \min [I_{\phi}(U), \eta] \quad (5)$$

为了方便表示,用  $B_t$  表示第  $t$  轮次选择的增益节点集合,并将  $(B_t, t)$  看作一个元组元素,用  $x_t$  简化表示。用  $U_t$  表示前  $t$  轮次选择的元组集合,即  $U_t = \{(B_1, 1), (B_2, 2), \dots, (B_t, t)\}$ 。

定义传播函数的边际增益函数和阈值传播函数的阈值边际增益函数。

$$I_{\phi}(x_t | U_{t-1}) = I_{\phi}(x_t \cup U_{t-1}) - I_{\phi}(U_{t-1}) \quad (6)$$

$$\Gamma_{\phi}(x_t | U_{t-1}) = \Gamma_{\phi}(x_t \cup U_{t-1}) - \Gamma_{\phi}(U_{t-1}) \quad (7)$$

用  $\Omega$  表示所有可能传播的实例组成的集合。每轮次选择一个元组后,可以观察得到反馈状态,  $\Omega_t \in \Omega$  表示与  $t$  轮次传播前观察到的传播状态一致的传播实例所组成的传播子集。定义  $n_t$  表示在第  $t$  轮次传播开始时还未被激活的节点个数,  $\eta_t$  表示第  $t$  轮次传播开始时与要求阈值的差值,即  $\eta_t = \eta - (n - n_t)$ 。那么易知

$$\Gamma_{\phi}(x_t | U_{t-1}) = \min \{I_{\phi}(x_t | U_{t-1}), \eta_t\} \quad (8)$$

则期望阈值边际增益函数可以表示为

$$\Delta(x_t | U_{t-1}) = E_{\phi \sim \Omega_t} [\Gamma_{\phi}(x_t | U_{t-1})] \quad (9)$$

即在目前传播状态下,选择元组  $x_t$  在下一轮次可以产生的阈值边际增益的期望值。

### 2.1.2 自适应贪心算法框架

参考种子节点最小化的自适应处理范式<sup>[21]</sup>,当自适应算法的优化目标函数满足强自适应子模性<sup>[21]</sup>和强自适应单调性<sup>[21]</sup>,每轮次直接利用贪心策略,选择使目标优化函数增益最大化的元组,存在理论近似保证,可以最小化所选择的元组数量以达到阈值。

可以证明,增益节点自适应选择问题的目标函数  $\Gamma(U)$  满足强自适应子模性和强自适应单调性,基于自适应成本最小化近似定理<sup>[15]</sup>,在第  $t$  轮次传播选择增益节点集合时,选择的元组  $(B^{\circ}, t)$  需要满足  $B^{\circ} = \arg \max_B \Delta[(B, t) | U_{t-1}]$ 。但计算最大期望阈值边际增益函数是一个 NP 难问题,如果可以找到近似最优解,即选择的元组需要

满足  $\Delta[(B_t, t) | U_{t-1}] \geq \alpha \Delta[(B^{\circ}, t) | U_{t-1}]$ ,达到阈值时所经过的期望传播轮次不会超过最优策略下期最少轮次的  $(\ln \eta + 1)^2 / \alpha$ <sup>[15]</sup>。算法 1 描述了自适应贪心增益节点选择策略的流程。

### 算法 1 自适应贪心增益节点选择策略

Alg. 1 Adaptive greedy multi-round boost node selection strategy

输入: 社交网络  $G$ , 备选种子集合  $\hat{S}$ , 传播影响阈值  $\eta$ , 每轮次节点数量  $b$

输出: 各轮次选择增益节点集合  $U$ , 满足  $\Gamma(U) = \eta$

初始化  $U \leftarrow \emptyset, \Gamma(U) \leftarrow 0, t \leftarrow 1$

初始化激活节点集合  $A \leftarrow \emptyset$

当  $\Gamma(U) < \eta$  时:

    选择当前轮次的增益节点集合  $(B_t, t)$ , 其中  $B_t$  满足  $B_t \subseteq V/\hat{S}, |B_t| = b$  且  $\Delta[(B_t, t) | U] \geq \alpha \Delta[(B^{\circ}, t) | U]$

    生成新一轮次传播实例, 统计新激活节点集合  $A_t$

$\Gamma(U) \leftarrow \Gamma(U) + |A_t|, A \leftarrow A \cup A_t, U \leftarrow U \cup \{(B_t, t)\}$

$t \leftarrow t + 1$

## 2.2 期望阈值边际增益函数最大化

要得到有下界保证的自适应算法近似解,就要计算最大期望阈值边际增益函数。这一问题虽然证明为 NP 难问题。但一些工作<sup>[3-7, 12]</sup>已经为这一问题提供了有近似保证的处理范式。解决这一问题分为两步: 采样足够数量的传播实例, 利用传播实例估计给定节点的传播影响; 通过贪心策略选择估计值最大节点。估计给定增益节点集合所产生的期望影响增益, 通常通过采样潜在反向可达图<sup>[12]</sup>, 但这种方法并不能简单地用来估计阈值下的期望边际增益。在本节中, 采用多根潜在反向可达图 ( $k$ -root potential reverse reachable graph,  $k$ PRR-Graph) 的采样和截断影响传播最大化 (truncated influence maximizing, TRIM) 算法<sup>[15]</sup> 选择增益节点使期望阈值边际增益函数最大化。

### 2.2.1 增益节点的传播影响估计

文献[12]提出了潜在反向可达图的概念, 利用潜在反向可达图估计期望影响增益。

对于社交网络  $G = (V, F)$  上的边  $e_{uv} \in F$ , 其有  $p_{uv}$  的概率被激活, 有  $1 - p'_{uv}$  的概率阻塞传播, 除此之外还存在  $p'_{uv} - p_{uv}$  的概率, 当且仅当  $v$  被提升为增益节点后, 它才能变成激活边, 否则作为阻塞边存在。定义这类边为待增益边。

一个随机潜在反向可达图对应一个传播实例

的子图。生成一个随机潜在反向可达图,从备选种子集中均匀选择本轮次的种子节点,然后在其他节点中均匀随机选择一个节点  $r$  作为根,并从  $r$  开始进行反向广度优先搜索。搜索过程中访问到的每条边都按照上述概率保留为激活边、待增益边或阻塞边。由  $r$  到达所有种子节点的非阻塞边组成的最小子图就构成了一个随机潜在反向可达图  $R$ 。

在随机潜在反向可达图  $R$  上定义一个二值函数  $y_R(B)$ 。当且仅当从种子节点出发存在激活路径可以到达根节点,或者提升节点集合  $B$  为增益节点后存在一条激活路径到达根节点时,  $y_R(B) = 1$ ; 否则  $y_R(B) = 0$ 。假设已经采样了足够数量的潜在反向可达图,用  $\mathcal{R}$  表示这些图集合。在给定的备选种子集合  $\hat{S}$  下,增益节点集合  $B$  的传播影响期望可以估计为

$$\mathbb{E}[I(B)] = n/|\mathcal{R}| \cdot \sum_{R \in \mathcal{R}} y_R(B) \quad (10)$$

### 2.2.2 期望阈值边际增益函数估计

计算  $\mathbb{E}[I_\phi(x_i | U_{i-1})]$ , 一种简单的思路是直接利用  $\mathbb{E}[I_\phi(x_i | U_{i-1})] \cdot \mu_i/n_i$  进行估计,然而当节点数量较多时,这种方法可能会严重偏离实际值。Tang 等<sup>[15]</sup> 在种子数量最小化任务中提出,估计阈值边际增益函数时,可以采用将反向可达图拓展成多根的方法,从而得到有效的近似值。假设潜在反向可达图的根节点数量为  $k$ , 根节点对于阈值边际增益函数的期望贡献可以表示为  $\min(n_i/k, \eta_i)$ ,  $k$  越接近  $n_i/\eta_i$ , 准确率越高。

基于这一思路,通过采样  $k$  根潜在反向可达图,可以估计给定增益节点在阈值下的期望边际增益。采样生成一个随机的  $k$ PRR-Graph, 需要从  $V \setminus [\rho(U_{i-1}) \cup \hat{S}]$  中随机选择  $k$  个节点作为根节点,从这些根节点出发构造潜在反向可达图  $R$ 。定义二值函数  $\hat{y}_R(B)$ , 当且仅当种子节点存在激活路径到达任一根节点,或者提升节点集合  $B$  为增益节点后,存在激活路径到达至少一个根节点时,  $\hat{y}_R(B) = 1$ , 否则  $\hat{y}_R(B) = 0$ 。

定义  $\mathcal{R}$  表示采样的随机  $k$ PRR-Graph 集合, 则期望阈值边际增益函数可以近似计算为

$$\tilde{\Gamma}_{\mathcal{R}_i}[(B_i, t) | U_{i-1}] = \frac{\eta_i}{|\mathcal{R}_i|} \cdot \sum_{R \in \mathcal{R}_i} \hat{y}_R(B_i) \quad (11)$$

**定理 1**<sup>[15]</sup> 定义概率  $r = n_i/\eta_i - \lfloor n_i/\eta_i \rfloor$ , 采样潜在反向可达图时,按照概率  $r$  设置根节点数量  $k = \lceil n_i/\eta_i \rceil$ , 按照概率  $1 - r$  设置  $k = \lfloor n_i/\eta_i \rfloor$ , 则有

$$(1 - 1/e) \mathbb{E}[I(x_i | U_{i-1})] \leq \mathbb{E}[\tilde{\Gamma}(x_i | U_{i-1})] \leq \mathbb{E}[I(x_i | U_{i-1})] \quad (12)$$

### 2.2.3 采样和增益节点选择

在目标函数满足单调性和子模性时,贪心策略可提供严格的近似保证。贪心策略从空集合开始,逐一选择使目标优化函数最大化的节点加入集合。但  $\tilde{\Gamma}[(B_i, t) | U_{i-1}]$  只满足单调性而非子模性。Lin 等<sup>[12]</sup> 提出三明治近似策略为非子模的目标优化函数最大化问题,并提出了一个解决范式。由于直接在非子模函数上使用贪心策略得到的结果是没有近似保证的,可能出现结果偏小的情况。构造满足单调子模性的紧凑下界函数,计算使下界函数最大化的节点集合可以弥补直接贪心导致的结果偏小的情况,从而提供最大化  $\tilde{\Gamma}[(B_i, t) | U_{i-1}]$  的近似保证。

在多根潜在反向可达图中定义关键点<sup>[12]</sup> 集合  $\hat{C}_R = \{v | \hat{y}_R(\{v\}) = 1\}$ 。定义  $\hat{y}_R(B)$  的单调子模下界函数  $\hat{y}_R^-(B) = \mathbb{I}(B \cap \hat{C}_R \neq \emptyset)$ 。期望阈值边际增益函数的下界函数可以估计为

$$M_{\mathcal{R}_i}[(B_i, t) | U_{i-1}] = \eta_i/|\mathcal{R}_i| \cdot \sum_{R \in \mathcal{R}_i} \hat{y}_R^-(B_i) \quad (13)$$

由于  $\hat{y}_R^-(B_i)$  具有单调子模性,  $M_{\mathcal{R}_i}[(B_i, t) | U_{i-1}]$  和  $\mathbb{E}[M[(B_i, t) | U_{i-1}]]$  也具有单调子模性。

TRIM<sup>[15]</sup> 算法框架通常用于解决单调子模的下界函数的期望最大化问题。算法 2 描述了这一过程。TRIM 算法在当前所维护的采样集合  $\mathcal{R}$  上应用贪心策略(算法 3)动态计算采样造成的误差,直至误差满足近似性要求。如果不满足则继续扩大采样的规模。

**定理 2**<sup>[15]</sup> 在  $1 - \delta$  的概率下,通过下界贪心策略可以得到期望阈值边际增益下界函数最大化问题的  $1 - \hat{\varepsilon}$  近似解。

可以证明,通过下界贪心策略求得的近似解,与阈值边际增益函数的最优解比例近似小于  $(1 - 1/e)(1 - \hat{\varepsilon}) \frac{\mathbb{E}[M(\cdot)]}{\mathbb{E}[\tilde{\Gamma}(\cdot)]}$ , 其中  $\frac{\mathbb{E}[M(\cdot)]}{\mathbb{E}[\tilde{\Gamma}(\cdot)]}$  表示下界函数和估计函数的近似比。

第 3 节中实验结果展示,直接采用下界贪心策略也可以获得一个逼近最优解的结果。

下界贪心策略会返回一个  $k$ PRR-Graph 的采样集合  $\mathcal{R}$ , 三明治近似策略在采样集合  $\mathcal{R}$  上贪心计算使  $\tilde{\Gamma}[(B_i, t) | U_{i-1}]$  最大化的增益节点集合  $B_{\tilde{\Gamma}}$ , 具体来讲,从空节点开始,逐一选择节点  $v_0 = \arg \max_v \eta_i/|\mathcal{R}| \cdot \sum_{R \in \mathcal{R}} \hat{y}_R(B_{\tilde{\Gamma}} \cup \{v\})$  加入  $B_{\tilde{\Gamma}}$ ,

## 算法 2 下界贪心策略

Alg. 2 Lower bound greedy strategy

输入: 社交网络  $G$ , 备选种子集合  $\hat{S}$ , 已激活节点集合  $A$ , 传播影响阈值残差  $\eta_t$ , 每轮次增益节点数量  $b$ , 误差参数  $\beta$

输出: 使期望阈值边际增益下界近似最大化的增益节点集合  $B_L$

设置参数  $\delta \leftarrow \varepsilon / [100(1 - 1/e)(1 - \varepsilon)\eta_t]$

$\hat{\varepsilon} \leftarrow 99\varepsilon / (100 - \varepsilon)$

$\theta_{\max} \leftarrow 2n_t \left( \sqrt{\ln \frac{6}{\delta}} + \sqrt{\ln \binom{n}{b} + \ln \frac{6}{\delta}} \right)^2 / (b\hat{\varepsilon}^2)$

$\theta_* \leftarrow \theta_{\max} \cdot \frac{b\hat{\varepsilon}^2}{n_t}$

$z \leftarrow \lceil \log_2 \frac{\theta_{\max}}{\theta_*} \rceil + 1$

$a_1 \leftarrow \ln \frac{3z}{\delta} + \ln \binom{n}{b}, a_2 \leftarrow \frac{3z}{\delta}$

生成  $\theta$  个随机多根潜在反向可达图, 保存在  $\mathcal{R}$  中

对于  $t$  从 1 到  $z$  进行循环:

$B_L \leftarrow \text{Greedy}(\mathcal{R}, \eta_t, b)$

$\Lambda^l(B_L) \leftarrow (\sqrt{\Lambda_{\mathcal{R}}(B_L)} + 2a_1/9 - \sqrt{a_1/2})^2 - a_1/18$

$\Lambda^u(B_m) \leftarrow (\sqrt{\Lambda_{\mathcal{R}}(B_L)} + a_2/2 + \sqrt{a_2/2})^2$

如果  $\frac{\Lambda^l(B_L)}{\Lambda^u(B_m)} \leq 1 - \hat{\varepsilon}$ :

提前结束循环

否则:

采样  $|\mathcal{R}|$  个随机多根潜在反向可达图, 加入  $\mathcal{R}$

## 算法 3 迭代过程中的贪心策略

Alg. 3 Greedy strategy on iterative step

输入: 潜在反向可达图集合  $\mathcal{R}$ , 传播影响阈值残差  $\eta_t$ , 每轮次增益节点数量  $b$

输出: 贪心策略下使期望阈值边际增益下界函数近似最大的增益节点集合  $B_L$

初始化  $B_L \leftarrow \emptyset$

当  $|B_L| < b$  时:

计算节点  $v_0 = \arg \max_v \eta_t / |\mathcal{R}| \cdot \sum_{R \in \mathcal{R}} \hat{y}_R(B_L \cup \{v\})$

$B_L \leftarrow B_L \cup \{v_0\}$

直到节点数量达到要求。

对比  $\tilde{\Gamma}_{\mathcal{R}}[(B_T, t) | U_{t-1}]$  和  $\tilde{\Gamma}_{\mathcal{R}}[(B_L, t) | U_{t-1}]$ , 选择使影响力更大的节点集合作为本轮次最终选择的增益节点集合  $B_{sa}$ , 即

$$B_{sa} = \arg \max_{B \in \{B_L, B_T\}} \mathbb{E}[\tilde{\Gamma}[(B, t) | B_{t-1}]] \quad (14)$$

算法 4 描述了这一过程。可以证明, 通过三

明治近似策略求得的增益节点集合是阈值边际增益函数最大化问题的  $(1 - 1/e)(1 - \hat{\varepsilon}) \cdot \frac{\mathbb{E}[M(\cdot)]}{\mathbb{E}[\tilde{\Gamma}(\cdot)]}$  近似最优解。

## 算法 4 三明治近似策略

Alg. 4 Sandwich approximation strategy

输入: 社交网络  $G$ , 备选种子集合  $\hat{S}$ , 已激活节点集合  $A$ , 传播影响阈值残差  $\eta_t$ , 每轮次增益节点数量  $b$

输出: 使近似期望阈值边际增益最大化的增益节点集合  $B_{sa}$

$B_L \leftarrow \text{Greedy}(G, \hat{S}, A, \eta_t, b)$

获取最终生成的随机潜在反向可达图集合  $\mathcal{R}$

初始化  $B_T \leftarrow \emptyset$

当  $|B_T| < b$  时:

计算节点  $v_0 = \arg \max_v \eta_t / |\mathcal{R}| \cdot \sum_{R \in \mathcal{R}} \hat{y}_R(B_T \cup \{v\})$

$B_T \leftarrow B_T \cup \{v_0\}, B_{sa} = \arg \max_{B \in \{B_L, B_T\}} \mathbb{E}[\tilde{\Gamma}[(B, t) | U_{t-1}]]$

## 2.3 复杂度分析

## 2.3.1 准确性近似分析

对下界贪心策略和三明治近似策略的理论近似性进行分析。

**定理 3** 在  $1 - \delta$  的概率下, 通过下界贪心策略和三明治近似策略求得的近似解  $B_L$ , 与使期望阈值边际增益函数  $\mathbb{E}\{\Gamma[(B, t) | U_{t-1}]\}$  最大化的最优解的比值近似小于  $(1 - 1/e)(1 - \hat{\varepsilon}) \cdot \frac{\mathbb{E}[M(\cdot)]}{\mathbb{E}[\tilde{\Gamma}(\cdot)]}$ 。

采用自适应贪心策略  $\pi$  可以得到近似比为  $(\ln \eta + 1)^2 / \alpha$  的近似解, 其中  $\alpha$  为每轮次求解最大阈值边际增益时的近似比例。下界贪心策略和三明治近似策略两种方法中, 均有

$$\alpha = (1 - 1/e)(1 - \hat{\varepsilon}) \frac{\mathbb{E}[M(\cdot)]}{\mathbb{E}[\tilde{\Gamma}(\cdot)]} \quad (15)$$

综上, 采用自适应贪心策略得到的传播轮次与最优解的近似比小于  $\frac{(\ln \eta + 1)^2}{(1 - 1/e)(1 - \hat{\varepsilon})}$ 。

$$\frac{\mathbb{E}[\tilde{\Gamma}(\cdot)]}{\mathbb{E}[M(\cdot)]}$$

## 2.3.2 时间复杂度分析

下界贪心策略和三明治近似策略的复杂度主要来自三个方面: 采样随机多根潜在反向可达图的总数量、生成一个随机多根潜在反向可达图的时间复杂度和节点贪心选择的复杂度。

**定理 4** 第  $t$  轮次应用下界贪心策略产生的

随机多根潜在反向可达图的数量为  $O\left[\frac{b \ln n + \ln(\eta_i/\varepsilon)}{\varepsilon^2 O_{b,i}} \eta_i\right]$ ,  $O_{b,i}$  表示选择数量为  $b$  的增益节点集合可以产生的最大期望阈值边际增益下界。

**定理 5** 定义  $Q$  表示一个随机多根潜在反向可达图的最大期望,在每轮次采用下界贪心策略和三明治近似策略的计算复杂度为

$$O\left\{\frac{Q}{O_{b,i}} \cdot b \cdot [b \cdot \ln n + \ln(\eta_i/\varepsilon)] \cdot (n+m)/\varepsilon^2\right\} \quad (16)$$

综上,采用自适应贪心策略的算法总复杂度为

$$O\left\{T \cdot \frac{Q}{O_{b,i}} \cdot b \cdot [b \cdot \ln n + \ln(\eta_i/\varepsilon)] \cdot (n+m)/\varepsilon^2\right\} \quad (17)$$

式中,  $T$  为采用自适应贪心策略达到阈值所需要的传播总轮次,采用自适应贪心策略可以达到近似线性的复杂度。

### 3 实验结果

为了验证 AdaMBTS 算法的正确性与时间效率,在多个真实数据集上进行实验,采用启发式的增益节点集合选择算法作为基线方法进行对比,实验结果展示 AdaMBTS 在正确性上具有明显的优势。所有的实验在 CPU 型号为 Intel Xeon Gold 5220R@2.20 GHz, 1 T 内存的机器上进行。

#### 3.1 实验设置

##### 3.1.1 数据准备

**数据集:** 实验在真实社交网络数据集 Digg<sup>[17]</sup>、Flixster<sup>[18]</sup> 和 Twitter<sup>[19]</sup> 上进行。采用数据集中提供的图结构建立社交网络  $G = (V, F)$ 。图结构中每个节点代表一个用户,在 Digg 数据集上,若两个用户之间存在发帖和回帖的关系,则存在一条边。Flixster 数据集上的边表示用户之间的好友关系。Twitter 中的边表示用户主页之间的关注关系。实验仅在这些数据集的最大弱联通分量上进行。

**估计传播概率:** 实验前,需要学习图上每条边的影响概率  $p_w$  和增益概率  $p'_w$ 。利用 Lin 等<sup>[12]</sup> 提出的方案构建影响增益模型。实验采用 Goyal 等<sup>[22]</sup> 提出的影响概率模型学习社交网络上每条边的影响概率,采用计算结果的 2 倍作为边上的传播影响概率。各数据集点、边统计如表 1 所示。定义增益参数  $\beta$ , 对于社交网络上的边  $e_w$ , 若这条边的影响概率为  $p_w$ , 那么提升后的影响概率为

$$p'_w = 1 - (1 - p_w)^\beta (\beta > 1)。$$

表 1 数据集点、边统计

Tab. 1 Statistics of points and edges in the dataset

数据集名称	点数量	边数量	平均度数	平均传播概率
Digg	26 707	200 189	15.0	0.36
Flixster	95 969	484 865	10.1	0.34
Twitter	323 185	2 148 717	13.3	0.73

**备选种子集合:** 采用基于鞅的影响力最大化 (influence maximization via martingales, IMM) 算法<sup>[6]</sup> 选择 500 个传播影响最大的节点作为备选种子集合  $\hat{S}$ , 从这些种子出发可以得到有效的传播。设定种子节点成为各轮次传播的种子节点的概率为 0.1。

##### 3.1.2 基线算法

采用启发式算法和非自适应的增益节点集合选择策略作为基线算法与 AdaMBTS 进行对比。实验时,预先生成了 20 组传播 120 轮次的传播实例,统计在 20 组传播实例上到达阈值时所进行的传播轮次的平均值,在设计传播阈值时,约束了传播轮次不超过 120。

**基线算法采用最大化节点度数 (Degree)<sup>[12]</sup>、PageRank 和非自适应的基于单轮次增益节点选择策略 (single round based multiply boosting nodes selection, S-MBTS)<sup>[12]</sup>。**

**Degree<sup>[12]</sup>:** 在每轮次中,最大节点度数启发式算法迭代选择有权度数最大的节点加入增益节点集合,直到增益节点集合数量达到要求。参考文献<sup>[12]</sup>的节点度数权重定义,实验中采用 5 种权重度数,节点  $u$  的权重度数定义为:①所有出边传播概率的总和;②不包括终点在增益节点集合中的出边的传播概率总和;③所有入边提升概率与原传播概率差的总和;④不包括起点在增益节点集合中所有入边提升概率与原传播概率差的;⑤不包括终点在已经激活节点集合中的出边的传播概率总和。实验结果采用 5 种不同有权度数定义中平均传播轮次最少的结果。

**PageRank<sup>[20]</sup>:** 在每轮次中,选择 PageRank 分数最大的  $b$  个节点加入增益节点集合。

**S-MBTS<sup>[12]</sup>:** 在传播开始前通过三明治近似策略<sup>[12]</sup> 贪心选择  $b$  个使传播影响增益最大的增益节点,在每轮次传播的时候,均提升这  $b$  个节点为增益节点,观察最终传播影响达到阈值时所需

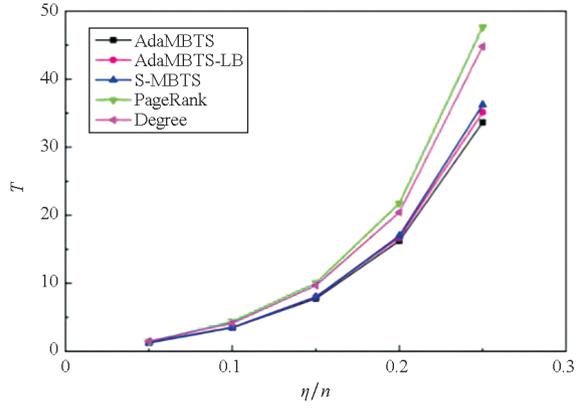
的传播轮次。

### 3.2 结果分析

在 Digg<sup>[17]</sup>、Flixster<sup>[18]</sup> 和 Twitter<sup>[19]</sup> 数据集上进行测试,实验结果表明,在不同阈值与不同参数下,基于 AdaMBTS 和 AdaMBTS-LB 的增益节点选择策略相比于最大化节点度<sup>[12]</sup>、PageRank<sup>[20]</sup>、非自适应增益节点选择策略<sup>[12]</sup> 3 个基线算法表现出了明显的优势。达到阈值所需的期望传播轮次远低于其他基线方法。从时间效率上,AdaMBTS 的复杂度近似线性,AdaMBTS-LB 采用下界贪心策略估计期望边际增益最大解。在保持相当的正确率的情况下,相比于采用三明治近似策略的 AdaMBTS,AdaMBTS-LB 在时间效率有显著提升。

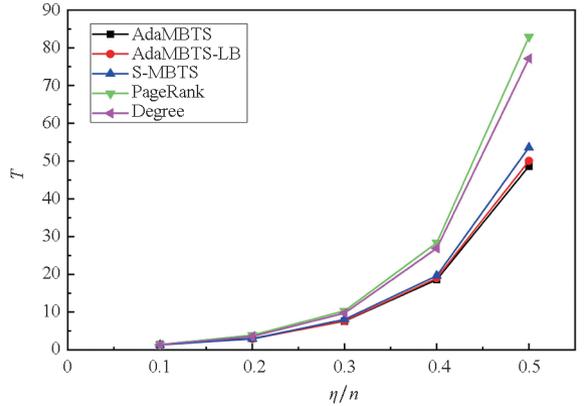
#### 3.2.1 性能对比

在不同传播影响阈值下,采用自适应选择策略所需的期望最少传播轮次始终优于其他基线算法。设置增益系数  $\beta = 3$ , 每轮次增益节点数量  $b = 50$ , 改变不同数据集上的影响传播阈值与节点数量的比例  $\eta/n$ , 计算采用不同增益节点选择策略所需的期望传播轮次  $T$ 。实验结果如图 1 所示。可以观察到,在 3 个数据集上采用 AdaMBTS 和 AdaMBTS-LB 要优于其他启发式节点选择策略<sup>[12,20]</sup>, 相比于 S-MBTS<sup>[12]</sup> 也存在相当大的优势。图 1 中 PageRank 和 Degree 的期望传播轮次远大于通过 AdaMBTS 选择增益节点所需的传播轮次。在 Digg 数据集上, Degree 和 PageRank 平均需要 44.8 和 47.7 轮次才能到达 25% 的影响阈值;而 AdaMBTS 和 AdaMBTS-LB 仅需要 33.65 和 35.15 轮次即可达到阈值。在 Flixster 数据集上, Degree 和 PageRank 平均需要 77.15 和 82.9 轮次才能达到 50% 的影响阈值;而 AdaMBTS 和 AdaMBTS-LB 仅需要 48.65 和 50.05 轮次即可达到指定传播阈值。在 Twitter 数据集上, Degree 和 PageRank 都平均需要 87 轮次才能达到 35% 的影响阈值;而 AdaMBTS 和 AdaMBTS-LB 仅需要 60.8 和 62.2 轮次即可达到指定传播阈值。而 S-MBTS 在 3 个数据集上分别需要 36.3、53.6、74.4 传播轮次达到阈值。采用 AdaMBTS 比采用 S-MBTS 在 3 个数据集上分别减少了 7.3%、9.2%、18.3% 的时间轮次, 而采用 AdaMBTS-LB 比采用 S-MBTS 在 3 个数据集上分别减少了 3.2%、6.6%、16.4% 的时间轮次, 两种方法都表现出了明显的优势, 而这一优势将随着传播轮次的增加而增大。



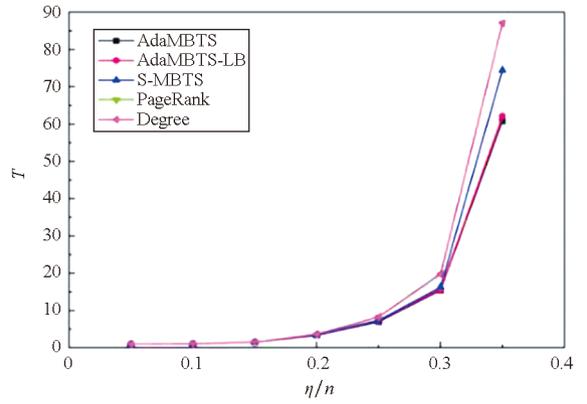
(a) Digg 实验结果

(a) Result on Digg



(b) Flixster 实验结果

(b) Result on Flixster



(c) Twitter 实验结果

(c) Result on Twitter

图 1  $\beta = 3, b = 50$  时不同  $\eta/n$  下所需的传播轮次

Fig. 1 The required spread rounds at different  $\eta/n$  when  $\beta = 3, b = 50$

图 2 分别展示了不同数据集上 AdaMBTS 和 AdaMBTS-LB 的平均每轮次运行时间, 线条上的数值为 AdaMBTS-LB 相对于 AdaMBTS 的加速比。虽然 AdaMBTS-LB 算法准确率略差于 AdaMBTS, 但在运行时间上有着显著优势, 在精确度需求不高的时候可以采用 AdaMBTS-LB 算法提升效率。随着传播影响阈值的提高, AdaMBTS 算法在

Digg 数据集和 Flixster 数据集上平均每轮次运行时间呈现上升的趋势,这是由于采样数量提升。而在 Twitter 数据集却先上升而后下降。这是由于随着  $\eta$  的增长,传播轮次增长,在早期传播轮次中每轮次采样的  $k$ PRR-Graph 数量增多,但  $k$ PRR-Graph 的根的数量减少,根的数量直接与  $k$ PRR-Graph 的最大期望大小有关。而 Twitter 数据集的规模最大, $k$ PRR-Graph 的期望大小对运行时间的影响更大,随着  $\eta$  的增长,总传播轮次数量增长,增长的尾部传播轮次带来的传播影响增益相对减少,导致对应的  $\eta_i$  较小,运行时间较短,从而使得整体的平均每轮次运行时间变小。

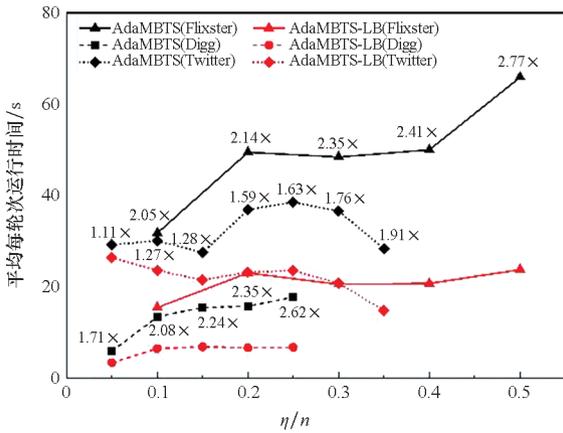


图2  $\beta = 3, b = 50$  时不同  $\eta/n$  下的平均每轮次运行时间

Fig. 2 Average time per round at different  $\eta/n$  when  $\beta = 3, b = 50$

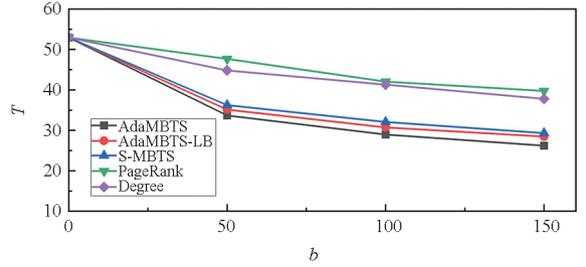
### 3.2.2 参数调整

为了测试不同参数对于算法准确性的影响,分别在  $\beta = 3$  以及  $b = 50, 100, 150$  的情况下进行对比实验。实验证明 AdaMBTS 在不同参数的设置下,达到传播影响阈值所需要的传播轮次远小于基线算法,始终保持一个较高的准确率。

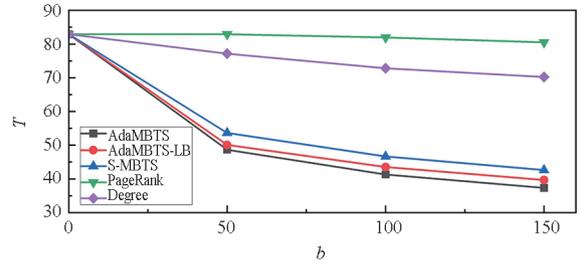
由于不同数据集的规模和结构不同,在不同的数据集设置了不同的阈值限制,从而能够在实验预生成的传播轮次中达到阈值。Digg 上设置  $\eta/n = 0.25$ , Flixster 上设置  $\eta/n = 0.5$ , Twitter 上设置  $\eta/n = 0.35$ 。

图3展示了在不同数据集上  $\beta = 3$  时,设置不同的增益节点集合数量  $b$ ,采用不同的增益节点选择策略时,达到阈值所需的传播轮次  $T$  的变化。可以观察到,随着增益节点数量的增加,期望传播轮次逐渐减少,但减少幅度逐渐变小,增加增益节点集合的数量对于传播轮次的减少起到的效果有限。基于 AdaMBTS 和 AdaMBTS-LB 选择的

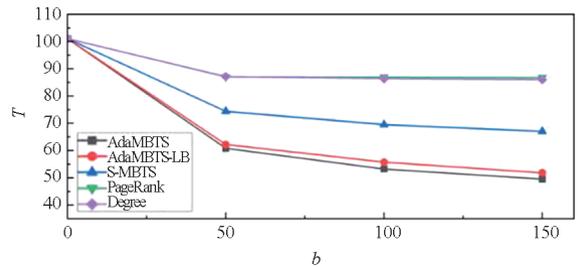
增益节点集合相比于基线算法,能够有效减少阈值要求下必需的传播轮次。



(a) Digg 实验结果  
(a) Result on Digg



(b) Flixster 实验结果  
(b) Result on Flixster



(c) Twitter 实验结果  
(c) Result on Twitter

图3  $\beta = 3$  时不同增益节点集合数量下达到阈值所需的传播轮次

Fig. 3 The required spread rounds at different boosting node set quantity when  $\beta = 3$

图4同时展示了 AdaMBTS 和 AdaMBTS-LB 的平均每轮次运行时间,AdaMBTS-LB 在准确性上略低于 AdaMBTS,但在时间效率上相对于 AdaMBTS 有较大的提升,AdaMBTS-LB 的时间效率受到增益节点数量的影响相对较小,能始终保持高效的状态。

基于 2.3 节中的分析,AdaMBTS 和 AdaMBTS-LB 的算法复杂度与  $\epsilon^2$  成反比,图5展示了 AdaMBTS 和 AdaMBTS-LB 在不同误差参数  $\epsilon$  下的运行耗时与基线算法<sup>[12,20]</sup>的对比,其中 Degree 与 PageRank 的平均每轮次运行时间曲线几乎重合。实验在 Digg 数据集下进行,设置增益节点集合数量  $b = 50$ ,增益系数  $\beta = 3$ 。虽然 AdaMBTS 和 AdaMBTS-LB 的运行速度相较于基

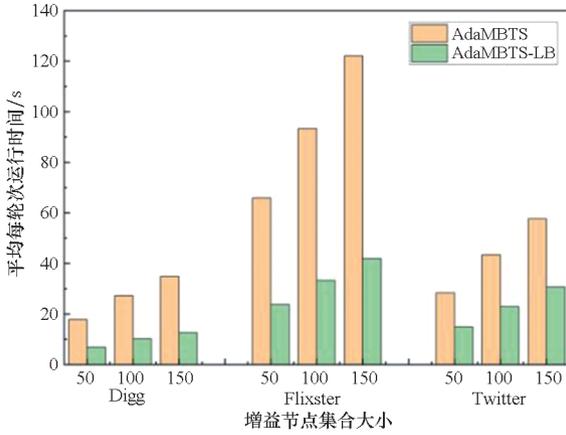


图 4  $\beta = 3$  时不同增益节点集合数量下  
平均每轮次运行时间

Fig. 4 Average time per round at different boosting  
node set quantity when  $\beta = 3$

线算法较慢,但通过调整误差参数,可以在结果准确率和运行效率之间取得一个均衡。即使设置较大的误差参数,在大部分情况下,AdaMBTS 和 AdaMBTS-LB 相比于基线算法仍然有效减少了达到阈值所需的轮次。

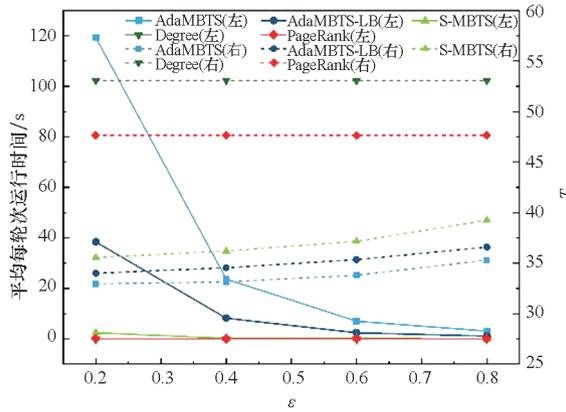


图 5 不同  $\epsilon$  时平均每轮次运行时间和  
达到阈值所需的轮次

Fig. 5 Average time per round and the required  
spread rounds at different  $\epsilon$

### 4 结论

提出了多轮次传播影响增益模型下的增益节点传播成本最小化问题。采用自适应的增益节点动态贪心选择策略和相应的下界策略,使得产品可以用最少的传播轮次达到传播影响阈值,并从理论上论述了 AdaMBTS 和 AdaMBTS-LB 与最优策略的近似性。在真实数据集上的实验结果表明,AdaMBTS 和 AdaMBTS-LB 能够高效地计算得到使期望传播轮次最少的有效近似解。

### 参考文献 (References)

- [1] AZAOUI M, MNASRI W, BEN ROMDHANE L. New trends in influence maximization models [ J ]. Computer Science Review, 2021, 40: 100393.
- [2] DOMINGOS P, RICHARDSON M. Mining the network value of customers [ C ] // Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2001: 57 - 66.
- [3] KEMPE D, KLEINBERG J, TARDOS É. Maximizing the spread of influence through a social network [ C ] // Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003: 137 - 146.
- [4] BORGS C, BRAUTBAR M, CHAYES J, et al. Maximizing social influence in nearly optimal time [ C ] // Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, 2014: 946 - 957.
- [5] TANG Y Z, XIAO X K, SHI Y C. Influence maximization: near-optimal time complexity meets practical efficiency [ C ] // Proceedings of the ACM SIGMOD International Conference on Management of Data, 2014: 75 - 86.
- [6] TANG Y Z, SHI Y C, XIAO X K. Influence maximization in near-linear time: a martingale approach [ C ] // Proceedings of the ACM SIGMOD International Conference on Management of Data, 2015: 1539 - 1554.
- [7] TANG J, TANG X Y, XIAO X K, et al. Online processing algorithms for influence maximization [ C ] // Proceedings of the International Conference on Management of Data, 2018: 991 - 1005.
- [8] KAZEMZADEH F, SAFAEI A A, MIRZAREZAEI M. Influence maximization in social networks using effective community detection [ J ]. Physica A: Statistical Mechanics and Its Applications, 2022, 598: 127314.
- [9] KAZEMZADEH F, SAFAEI A A, MIRZAREZAEI M, et al. Determination of influential nodes based on the communities' structure to maximize influence in social networks [ J ]. Neurocomputing, 2023, 534: 18 - 28.
- [10] ZHANG Y P, GUO J X, YANG W G, et al. Supplementary influence maximization problem in social networks [ J ]. IEEE Transactions on Computational Social Systems, 2024, 11(1): 986 - 996.
- [11] ALI K, WANG C Y, YEH M Y, et al. NEDRL-CIM: network embedding meets deep reinforcement learning to tackle competitive influence maximization on evolving social networks [ C ] // Proceedings of the IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA), 2021.
- [12] LIN Y S, CHEN W, LUI J C S. Boosting information spread: an algorithmic approach [ C ] // Proceedings of the IEEE 33rd International Conference on Data Engineering ( ICDE ), 2017: 883 - 894.
- [13] CHEN S M J, ZHANG Y P, YANG W G, et al. Information spread maximization with multi-boosting stages [ J ]. IEEE

- Transactions on Network Science and Engineering, 2022, 9(5): 3467 – 3477.
- [14] SUN L C, HUANG W R, YU P S, et al. Multi-round influence maximization [C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018: 2249 – 2258.
- [15] TANG J, HUANG K K, XIAO X K, et al. Efficient approximation algorithms for adaptive seed minimization [C]//Proceedings of the 2019 International Conference on Management of Data, 2019: 1096 – 1113.
- [16] NIKNAMI N, WU J. A budgeted framework to model a multi-round competitive influence maximization problem [C]//Proceedings of the IEEE International Conference on Communications, 2022: 4120 – 4125.
- [17] HOGG T, LERMAN K. Social dynamics of Digg [J]. EPJ Data Science, 2012, 1(1): 26.
- [18] JAMALI M, ESTER M. A matrix factorization technique with trust propagation for recommendation in social networks [C]//Proceedings of the Fourth ACM Conference on Recommender Systems, 2010: 135 – 142.
- [19] HODAS N O, LERMAN K. The simple rules of social contagion [J]. Scientific Reports, 2014, 4: 4343.
- [20] BRIN S, PAGE L. The anatomy of a large-scale hypertextual web search engine [J]. Computer Networks and ISDN Systems, 1998, 30: 107 – 117.
- [21] VASWANI S, LAKSHMANAN L V S. Adaptive influence maximization in social networks: why commit when you can adapt? [EB/OL]. (2016 – 04 – 27) [2023 – 11 – 01]. <https://arxiv.org/abs/1604.08171v1>.
- [22] GOYAL A, BONCHI F, LAKSHMANAN L V S. Learning influence probabilities in social networks [C]//Proceedings of the Third ACM International Conference on Web Search and Data Mining, 2010: 241 – 250.