Efficient Approximation Algorithms for Adaptive Seed Minimization

背景

社交网络成为人们讨论和分享他们各自的想法和评论的流行方式，基于此，个人的观点和想法可以通过口碑的方式快速传播。广告商可以利用这种特点提供免费的产品来给社交网络中的用户，让这些用户促销他们的产品。广告商可能想知道如何用最小的免费产品数量来达到所需要的利润。

种子最小化问题要求在影响传播过程中，最小化选择的种子节点的数量，并使其影响节点的数量至少为η，现存的工作大多数都聚焦于非渐进式的种子集合最小化，即所有的种子节点一次性被选择，然后观察被影响的节点。

解决问题

解决了以下研究存在的问题：

- 非渐进式种子集合最小化存在问题：求解返回的种子集合存在其影响传播达不到预设节点数量或选择了过量的种子节点使其影响传播远大于预设值的问题。
- 渐进式种子集合最小化存在问题：仅存的能够提供渐进式种子集合最小化的解的研究存在（1）对估算种子集影响期望精度要求过高从而导致计算开销大（2）不能提供任何非平凡的近似保证。

创新之处

- 提出了一个新颖的框架 $ASTI$ 用来解决种子集合最小化问题。
- 提出了一个新颖的评估截断影响传播的采样方法，基于提出的新概念-- mRR 集合。

解决方法

Def

给定一个概率式的社交网络$G=(V,E)$和一个阈值$\eta \in [1,n]$，种子集合最小化问题目的在于确定一个策略π，在概率实现$\phi \in \Omega$上，最小化完成影响传播至少为η的种子节点的数量，即：

$$
\begin{aligned}
&min_{\pi}\mathbb{E}[|S(\pi,\phi)|] \\
s.t. \quad &I_{\phi}(S(\pi,\phi)) \geq \eta \ for\ all\ \phi
\end{aligned}
$$

这里的Ω就是活边图的全集。

因为ASM问题将影响传播超过η的部分忽略，所以采用截断影响传播的概念：

给定一个种子集合S和一个阈值η，截断影响传播$\Gamma_{\phi} := min\{I_{\phi},\eta\}$。

ASTI

$ASTI$框架通过以下三个步骤处理ASM问题：

1. 迭代选择节点最大化期望边际截断传播，即选择 s_i ，使其 $\Delta(s_i|S_{i-1}) \geq \alpha\Delta(v|S_{i-1})\ all\ v \in V_i$ ，满足这个条件 $ASTI$ 可以提供一个非平凡的近似保证。
2. 观察新被影响的节点，将这些节点从 G_i 中移除
3. 更新相关信息，更新 i 和 S 。

直到$\Gamma(S) \geq \eta$，运行结束。

mRR sets

一个随机的$mRR-set$是在G的随机实现ϕ中，能够到达随机选取的k个节点集合K的节点集。它可以由以下两个步骤生成：

1. 随机选择 k 个节点组成的集合 $K \subseteq V$ 。
2. 从 K 开始，执行反向广度优先搜索，跟随到每个节点的边缘。将所有遍历到的节点插入到 R 中。

k的选取依据于：

$$k=\begin{cases}\lfloor\frac{n}{\eta}\rfloor+1 & p=\frac{n}{\eta}-\lfloor\frac{n}{\eta}\rfloor \\ \lfloor\frac{n}{\eta}\rfloor & otherwise\end{cases}$$

基于$mRR-set$，同过$TRIM$算法来选取每一批的种子节点：

1. 开始生成小数量的 mRR 集合。
2. 寻找一个节点 v，它覆盖了最多的 mRR 集合。
3. 判断终止条件，符合则返回 v，不符合翻倍 mRR 集合，再次寻找 v 。

这是每批选择一个节点，还可以每批选择b个节点，即$TRIM-b$算法，大致与上述相同，只是每次贪心的选择覆盖mRR集合最多的b个节点。

H3 时间复杂度和近似保证

- 基于 $TRIM$ 的 $ASTI$ 时间复杂度和基于 $TRIM-b$ 的 $ASTI$ 时间复杂度同为 $O(\frac{\eta(m+n)}{\varepsilon^2}ln(n))$
- $ASTI$ 得到的近似保证为 $\frac{(ln\eta+1)^2}{(1-(1-1/b)^b)(1-1/e)(1-\varepsilon)}$， b 即为每批选择的种子节点数。

H2 相关工作

1. 非渐进式
    - Goyal et.首次提出了种子集合最小化问题。
    - Chen在一个变种的线性阈值模型下研究种子结合最小化，每个节点都分配有固定的阈值。
    - Long and Wong在独立级联和线性阈值模型下广泛的研究种子集合最小化。
    - Goyal et al.提出了双标准近似算法，Zhang et al.做了改进。
    - Han et al.提出了 $ATEUC$ 算法，利用 RIS 评估节点的影响传播。
2. 渐进式
    - Vaswani and Lakshmanan考虑非渐进式种子集合最小化，应该是仅存的给该问题提供解的研究工作。

H2 实验

H3 数据集

Dataset	n	m	Type	Avg. deg.	LWCC size
NetHEPT	15.2K	31.4K	undirected	4.18	6.80K
Epinions	132K	841K	directed	13.4	119K
Youtube	1.13M	2.99M	undirected	5.29	1.13M
LiveJournal	4.85M	69.0M	directed	28.5	4.84M

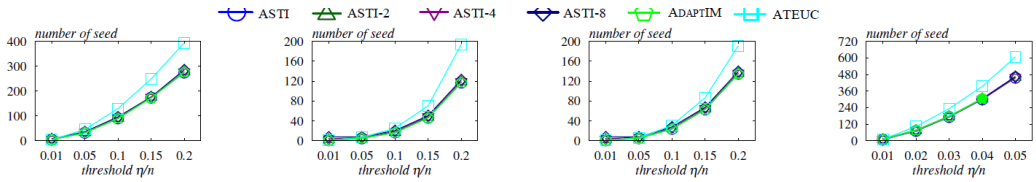
H3 算法

$ASTI$，$ASTI-2$，$ASTI-4$，$ASTI-8$，$ADAPTIM$ and $ATEUC$。其中$ASTI-b$是每批选择b个种子节点。$ADAPTIM$是渐进式影响最大化算法，$ATEUC$是最先进的非渐进式种子集合最小化问题的解决方案。

H3 参数设置

- 传播模型：IC和LT
- $p(<u,v>)=\frac{1}{indeg_v}$
- 对于NetHEPT, Epinions and Youtube，设置 $\frac{\eta}{n}=\{0.01,0.05,0.1,0.15,0.2\}$,对于LiveJournal设置 $\frac{\eta}{n}=\{0.01,0.02,0.03,0.04,0.05\}$

H3 实验结果



(a) NetHEPT (b) Epinions (c) Youtube (d) LiveJournal

Figure 4: Number of seed nodes vs. threshold under the IC model.

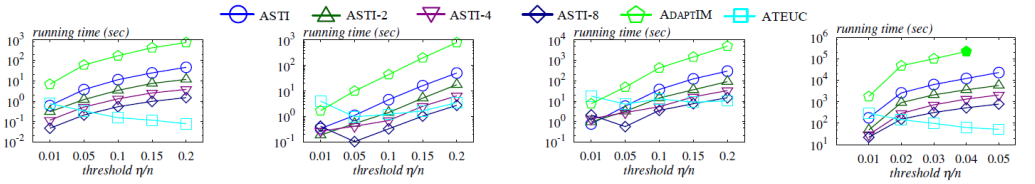
该图是在不同阈值$\frac{\eta}{n}$下，达到预期影响，最终选取的种子集合大小的变化情况。可以看出$ASTI$和$ADAPTIM$结果相近，且都比$ATEUC$选取的数量要低，最终大概差了%30-%40。

Table 3: Improvement ratio of ASTI over ATEUC

	η/n	0.01	0.05	0.1	0.15	0.2
IC Model	NetHEPT	N/A	40.8%	43.8%	43.0%	43.7%
	Epinions	N/A	N/A	50.7%	N/A	**65.7%**
	Youtube	0.0%	24.3%	N/A	37.5%	41.7%
	LiveJournal	N/A	43.0%	34.9%	N/A	33.0%
LT Model	NetHEPT	N/A	N/A	N/A	44.3%	47.5%
	Epinions	N/A	N/A	N/A	N/A	N/A
	Youtube	0.0%	39.5%	54.1%	N/A	47.9%
	LiveJournal	N/A	N/A	N/A	N/A	N/A

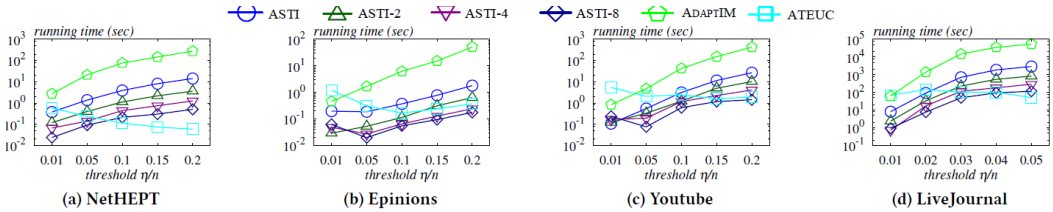
N/A: ATEUC does not meet the threshold for some realizations.

该表格可以看出在两个传播模型下，$ASTI$和$ATEUC$相差的百分比。



(a) NetHEPT (b) Epinions (c) Youtube (d) LiveJournal

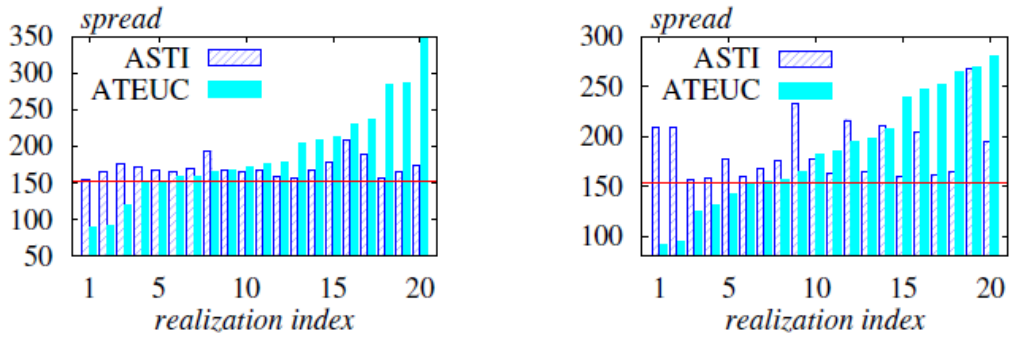
Figure 5: Running time vs. threshold under the IC model.



(a) NetHEPT (b) Epinions (c) Youtube (d) LiveJournal

Figure 7: Running time vs. threshold under the LT model.

该图反映了在两个传播模型下，各个算法运行的时间，因为$ATEUC$是非渐进式算法，而且阈值越大选取种子节点越多，遇到停止条件就越快，所有其运行时间呈下降趋势，$ADAPTIM$运行时间要是$ASTI$的10倍左右。



(a) IC model (b) LT model

Figure 8: Spread for 20 realizations on NetHEPT.

该图是在NetHEPT数据集下，实现20个活边图上，两个算法的影响传播期望分布。红线是要求达到的传播标准值。可以看出$ASTI$在20个实例上都达到了标准，而$ATEUC$在前5个都没有达到，而且在之后产生了过大的影响传播。因此$ASTI$相对来说更加稳定可靠。