

Multiplex Influence Maximization in Online Social Networks With Heterogeneous Diffusion Models

背景

用户可能不只活动在一个社交网络上，他们可以通过账号连接穿越多个网络，而各个网络往往传播模式互不相同。因此多重网络通过重叠用户连接在一起，而每个网络拥有自己的局部传播方式。

那么问题来了，每层社交网络要满足什么条件才能使得多重传播具有子模性？是否可用已存在的针对单一社交网络的方法来解决？在一个多重社交网络中重叠用户到底扮演一个什么样的角色？本文研究了多重网络传播影响最大化（MIM）问题，并针对上述问题做了阐述。

之前相关工作

之前大量工作研究单层网络的传播影响最大化，比如经典的IM问题，在两个传统传播模型——独立级联和线性阈值上用贪心算法求解。为了改进贪心算法的效率，偷懒估值法的思想被提出。之后提出触发模型和采样技术来改进算法的效率和有效性。

多重网络的影响最大化问题的研究开始于离线和在线网络之间的连接。第一项工作研究在随机网络上的SIR模型；第二项工作研究在线互动和线下事件结合，聚焦于信息流动和网络聚类，但是没有解决异质影响问题。

Shen等人通过将重叠节点表示为一个超节点，从而将所有网络合并成一个网络，这也不能保证网络层的异质性；Nguyen等人考虑同质性的多层网络传播；Pan研究在多重网络上的阈值激活问题，基于一个连续时间的传播模型。

创新之处

- 先前问题上的创新：提出了一个新问题——异质多重影响力最大化问题（HMIM）。
- 先前方法上的创新：提出了结合偷懒估值法的贪心算法和基于背包思想的种子选择方法来解决HMIM问题。

解决问题

- 考虑了多重网络传播的异质性，解决了异质多重影响力传播最大化问题。
- 为异质多重影响力传播最大和问题设计了两个近似算法求解。

解决方法

HMIM:

给定一个多重网络 $\mathcal{G} = (G_1, \sigma_1), \dots, (G_k, \sigma_k)$ ，有 k 层网络，每层都有独自的影响传播模型 σ_i ，寻找一个集合 $S \subset V$ ， $|S| = l$ ，最大化激活用户数量期望 $\sigma(S)$ 。一个问题的实例被表示为 $(\mathcal{G}, k, l, \sigma)$ 。

近似算法:

ISF贪心方法

首先证明了如果影响传播在每层满足确定子模性那么在多重上也是子模的。

基于这个理论，使用ISF（贪心算法结合偷懒估值）得到 $1 - 1/e$ 的近似率。时间复杂度为 $O(nl(m + n)\log n)$ 。

ISF思想：为所有顶点标号，初始化种子集 S 为空，将所有层节点并入集合 V 。遍历所有节点 v ，初始化它的边际增益为 $\sigma(v)$ 和回合数为0，将其加入到优先队列中。每次从优先队列弹出一个节点，如果是当前回合的节点，那么并入 S ，否则计算边际增益重新加入到优先队列中。直到选取种子集合 S 大小为 l 。

KSN 可并行多重算法

KSN通过近似解决多重选择背包问题，获得了在单独各层上影响力最大化问题的解的组合来解决MIM问题。KSN算法得到了 $((1 - \epsilon)\alpha)/((o + 1)k)$ 近似率的解。

KSN思想：KSN算法将算法 A 作为输入解决了 k 层网络的各层影响力最大化问题，选择的种子节点个数为 l 。对于 $j \in \{1, \dots, l\}, i \in \{1, \dots, k\}$ ，算法 A 并行的运行在 G_i 得到种子集合 T_{ij} ， $|T_{ij}| = j$ 。然后用多重选择背包问题近似算法决定每层有多少种子节点被选择。

实验

算法:

- ISF
- KSN
- ES: 平均每层种子节点为 l/k
- BSN: 将所有 l 个种子节点放在最大化 $\sigma_i(S_i)$ 层中。

数据集:

1. 根据Barabasi-Albert模型生成三个规模自由的网络 H_1, H_2, H_3 ，其拥有1000节点，4000边，平均度为4。三个网络传播模型分别设置为LT, IC和MLT。从每个层中随机选择索引，然后添加两个层间边以将三个单独的用户连接成重叠用户，该过程直到选择 o 个重叠用户为止。 \mathcal{H}_o 代表重叠用户数量。

TRACES OF REAL NETWORKS

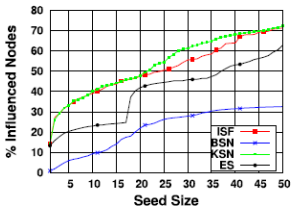
2.

Networks	Nodes	Edges	Avg Deg
Twitter	48277	16304712	289.7
FSQ	44992	1664402	35.99
CM	40420	175692	8.69
Het	8360	15751	1.88
NetS	1588	2742	1.73

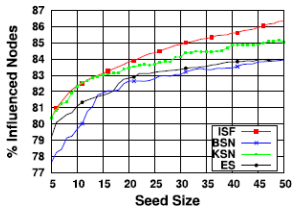
测试:

1. 综合多重数据上

a. 综合多重网络上的算法性能



(a) Overlap = 0 (Scale-free, \mathcal{H}_0)

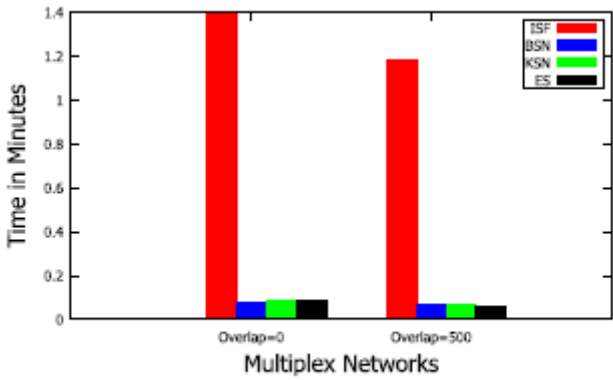


(b) Overlap = 500 (Scale-free, \mathcal{H}_{500})

无重叠用户时，KSN略优于ISF，两者都比其它方法要好。随着种子集合增大，影响节点数量增多。

有重叠用户时，ISF略优于KSN，两者比其它方法要好。随着种子集合增大，影响节点数量增多。

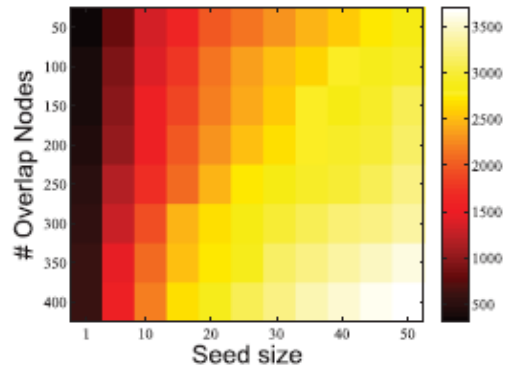
b. 运行时间



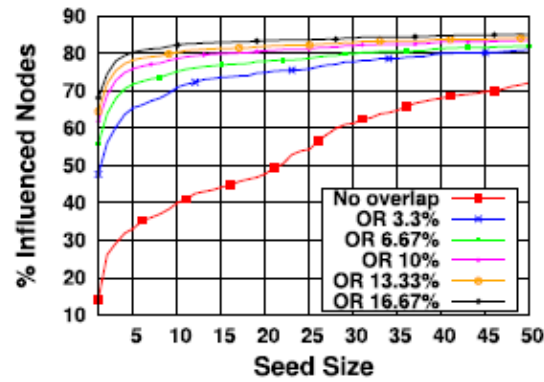
(c) Scale-free, \mathcal{H}

ISF运行时间最慢。有重叠用户时算法运行时间较快。

c. 重叠用户的角色



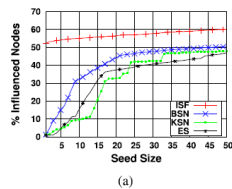
(a)



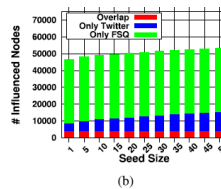
(b)

通过这两个实验图像可以看出，重叠用户数量越多，最终得到的影响用户数量越多。

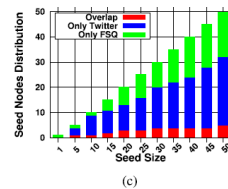
2. 真实社交网络数据上：



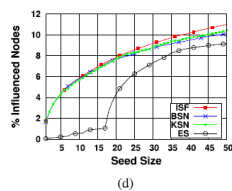
(a)



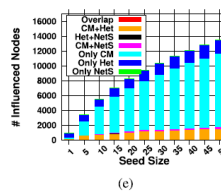
(b)



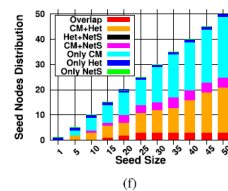
(c)



(d)



(e)



(f)

上图展示了在真实社交网络组合成为多重网络上，算法的性能以及种子节点的分布情况。

性能方面ISF最优，KSN和BSN相差不大。种子集合分布和影响节点分布方面，在Twitter和FSQ数据集上，种子集合分布在Twitter上更多，影响节点分布在FSQ上更多。在第二组数据上，种子集合分布在CM上更多，影响节点分布在CM上更多。

3. KSN的扩展性

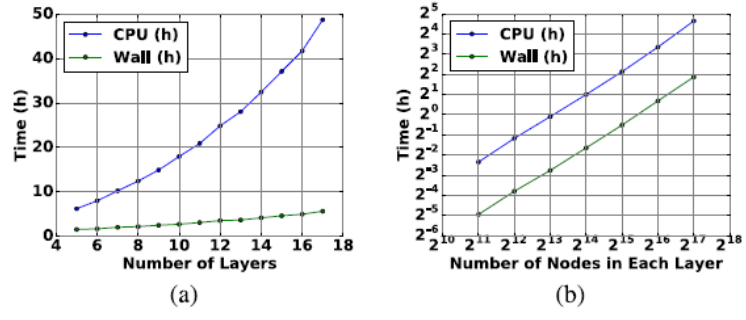


Fig. 9. Scalability of KSN. (a) $n_{ER} = 10^5$. (b) $k = 10$.

选择重叠用户数量为 $0.1n_{ER}$ ，随着网络层数增加，KSN运行时间指数性增长。随着网络每层节点数量增加，KSN运行时间线性增长。