# A new algorithm for positive influence maximization in signed networks

## 文章信息

文章：*符号网络中正影响最大化的新算法*

期刊：《information sciences》

时间：2019

作者：Weijia Ju, Ling Chen2, Bin Li, Wei Liu, Jun Sheng, Yuwei Wang

作者信息：扬州大学信息工程学院；南京大学新型软件技术国家重点实验室

分区：



### 摘要

随着在线社交网络的快速发展，影响力最大化问题受到了研究者的广泛关注，并被应用到营销、金融等多个领域。由于社会网络中个体之间可能存在正负关系，因此符号网络中的影响力最大化问题有着广泛的应用。文章提出了在独立级联模型下符号网络中积极影响力最大化的一种有效算法。首先，我们提出了一个独立的基于路径的算法来计算节点对之间的激活概率。在激活概率的基础上，我们定义了一个扩展函数，以避免模拟种子节点选择的影响扩散。我们提出了一种选择种子节点的算法，以使得选定的中子节点在有符号网络中能将积极影响扩散最大化。在社交网络中的实证结果表明，我们的算法比其他方法具有更广泛的正向传播影响。

## 文章内容

符号网络（signed network）是指包含正、负两种关系的二维复杂网络，是对一般复杂网络描述能力的一种推广。符号网络广泛存在于社会，生物等多种复杂系统中。例如，在社会系统中，“喜欢”、“尊重”和“表扬”属于正关系。而 “厌恶”、“轻视”和“责备”则属于负关系。再如，在神经系统中，神经元之问的“相互促进”属于正关系，而“相互抑制”属于负关系。符号网络社团结构具有社团内正关系稠密，同时社团间负关系也稠密的特点。

例如，C1和C2这两家公司是竞争关系。如果A客户更喜欢C1的产品，且与B客户的关系是正向的，那么B可能会受到A的正向影响，更喜欢C1的产品。如果A客户和B客户之间存在负向关系，那么B可能会受到A的负面影响不喜欢C1的产品，而更喜欢C2的产品。因此，社交网络中的顾客可以隐式地分为两类：喜欢C1产品的顾客和喜欢C2产品的顾客。当C1想要在社交网络上推销自己的产品时，他会尽量避开C2中可能会对C1产品产生负面影响的成员。此外，C1可能会通过提供折扣的方式，初步激活几个有积极影响的顾客，使其将积极影响传播到最大范围并消除对手C2的负面影响。

然而，现有的研究中存在两个问题。一是很少有研究者关注负面优势对竞争影响力传播的影响。在签名网络中，当影响通过负向传播时，有益的影响将变成有害的影响，反之亦然。一些情况下，负向传播对影响最大化结果的影响会大于正向。二是现有的积极影响最大化方法大多采用蒙特卡罗模拟来估计每个候选种子集S的影响扩散。然而，在IC模型下，评估给定种子集的影响扩散是一个棘手的问题。这种耗时的模拟过程阻碍了这些方法在解决现实问题中的应用。

**文章工作：**

1. 我们正式定义了**IC模型下签名网络中的正影响最大化问题**。还提出了一套传播规则来模拟竞争影响力的传播。
2. 我们提出一种算法，使用**独立路径**计算每个节点对之间的正激活概率。
3. 为了避免模拟影响传播的耗时过程，我们定义了一个**传播函数**来估计种子集的影响范围。
4. 提出了一种称为**PIM-SN**的算法，用于使用传播函数检测最佳种子集。

### PIM-SN问题和独立路径

在符号网络中，当影响力通过负边传播时，有益的影响力就会变成有害的，反之亦然。然而，现有的方法不能有效地利用符号网络的这一特性将有益的影响传播到最大范围。在本工作中，我们将该问题定义为有符号网络中的正影响最大化问题，并研究如何有效利用有符号关系的信息来检测种子节点，使正影响的范围最大化，并消除负影响。

在符号网络中，正连接表示节点之间存在正关系，如亲戚、朋友或同学，负链接表示消极关系，如敌人或竞争对手。在一个竞争激烈的市场中，顾客可能会因为朋友的建议而购买一件产品，而当对手的广告宣传时，顾客可能会拒绝。传播者希望自己的影响力能影响到大部分人，并限制竞争对手的影响力。因此，我们为每个活动节点定义了两个激活状态: 积极和消极激活状态表明节点受到了积极或消极的影响。我们的目标是利用节点之间的正、负关系，积极激活尽可能多的节点。我们将这样的问题定义为有符号网络中的积极影响最大化:

**定义1. 签名网络中的积极影响最大化，PIM-SN**

符号网络用表示，是有向连边集合，是网络的节点集。定义概率函数为中各链路传播的概率。每个链接上都存在正、负号，表示为。表示积极影响，表示消极影响。表示种子集遵循**符号网络影响传播规则**（SIP）的**正激活节点数**，PIM-SN在于**要找到一个种子节点集，使得正影响传播规模最大：**

**定义2.符号网络影响传播规则，SIP**

在符号网络上的影响传播中，当节点处于正（或负）激活状态时，如果边为正，则节点被激活到正（或负）状态的概率为，保持不活跃状态的概率为；如果边为负，节点将被激活到负（或正）状态，概率为，保持不活跃状态，概率为。

从定义可以看出，在这样的签名网络中，当有益的影响通过负链接传播时，它就会变成有害；反之，有害的影响将成为有益的影响。重要的是，要利用这种消极关系来有效地选择能够积极激活最多节点的种子。

### 路径传播概率

在大多数有符号网络的影响最大化算法中，都采用了传统的贪心算法。在贪心方法中，通过蒙特卡罗模拟，在每次迭代中估计所有候选种子集的正影响传播范围。这样一个耗时的模拟过程阻碍了贪心方法在现实问题中的应用。为了避免这种耗时的模拟过程，我们提出了一个传播函数来估计种子集的积极影响传播。

常用的影响传播模型有两种:线性阈值模型(LT)和独立级联模型(IC)。我们研究了IC模型下符号网络的正影响最大化问题。设u是一个被种子v间接激活的节点。在IC扩散模型下，激活过程可以用从v到u的路径L来表示。种子v通过L激活u的概率可以由路径L的概率估计，定义如下:

**定义3. 路径概率**

假设从v到u的路径L由一系列边，是在边上的概率。的概率从v到u可以估计为：

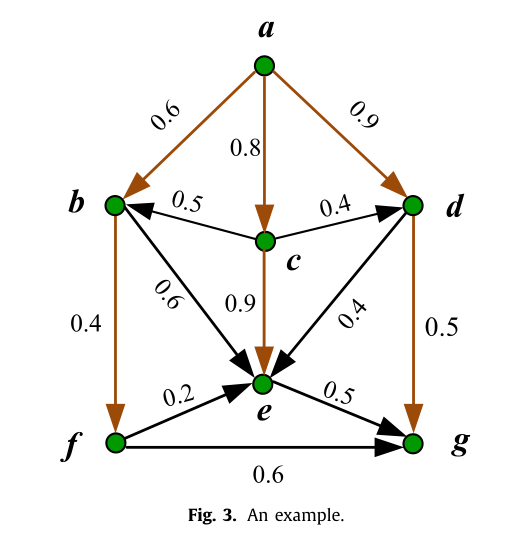
由于从v到u可能存在多条路径，所以种子v在独立路径上激活u的概率可以计算出来。

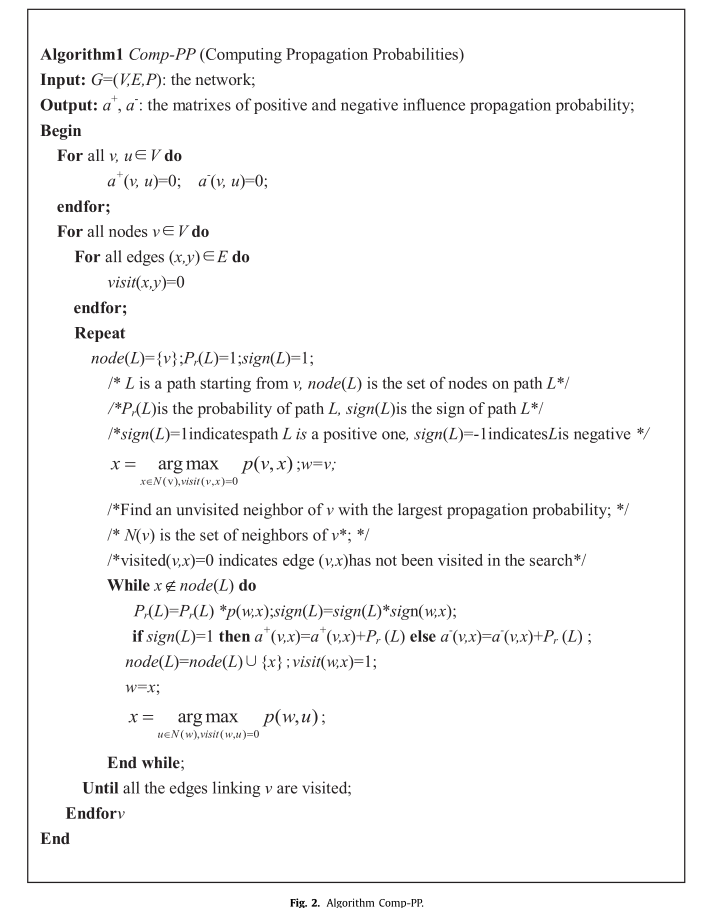
**定义4. 独立路径**

如果两条路径l1和l2不重叠，即它们没有共同边，我们称它们为相互独立的路径。

种子v正激活u的概率记为，种子v负激活u的概率记为。为了估计和，我们需要找到**一组从v到u的独立路径，并把它们的概率相加**。由于v到u的独立路径集合有很多，我们提出了一种算法来获得**概率最大的独立路径集合**。由于带有循环的路径对影响传播没有意义，我们只考虑没有重复节点的路径。

基于以上讨论，计算节点间传播概率的算法总结如图2所示。算法1采用深度优先搜索(deep -first search, DFS)方法检测路径。该算法中的DFS方法与经典的深度优先搜索方法有两个不同之处。首先，由于每条边在集合中只出现在一条路径上，所以搜索结果是一组独立路径的集合，而不是深度优先树。其次，当搜索到达一个节点w时，所有连接的边都已经访问过了，搜索将返回到起点v，这与传统的深度优先搜索不同，传统深度优先搜索回溯到访问边。算法1通过深度优先搜索，从网络中的每个节点v开始寻找独立路径。在深度优先搜索的每个节点w上，**算法选择传播概率最大的节点w的一个邻居作为路径中的下一个节点**。为了保证所构造路径的独立性，每个节点只能连接一条路径。因此，只能选择未访问的节点并将其添加到新路径中。当选择节点u时，路径L从v到u的概率加为或，这取决于路径L的符号。如果路径L上有偶数(奇数)条负边，则从v到u的路径L为正(负)边。设路径L的概率为如果路径L是正的，那么被加到；否则，它被加到中。



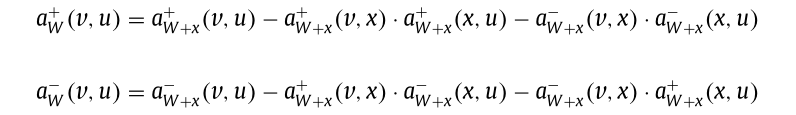


### 子图中的传播概率

，

设W，V是两个节点集，节点v是V中节点，用V-W, V-x, W+x 和V-S+v表示集合

**引理1.** 令是节点的集合，v和u是W中的两个节点，x是中的一个节点，则有：

****

**证明：**

假设和分别是和中v到u的独立正路径的集合。可以证明：

****

这里，是中从v到u经过节点x的所有独立正路径的集合。v通过的路径正向激活u的概率是。因为x 不属于W, 和中的所有路径是相互独立的，所以有：

****

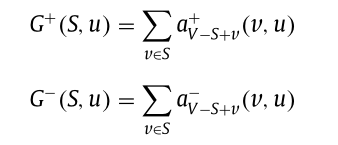
负路径同理

设W为贪婪方法选取的部分种子集，x为待加入W的候选种子节点。由引理1可知，将节点x加到W后，v正激活u的概率增量为。在贪心算法中，这个增量可以用来为种子集选择候选节点。

### 传播函数和性质

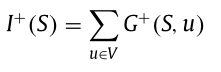
在定义的激活概率的基础上，我们定义了一个传播函数，该函数可以估计通过种子集传播的积极影响。之后将使用传播函数来解决PIM-SN问题。

为了计算种子节点S积极或消极激活节点u的概率，我们可以使用集合S中所有种子节点通过独立路径激活的概率进行计算。对于一个种子节点集，我们使用和分别表示S积极和消极激活u的概率，得到:

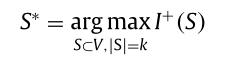
****

在上式中，是S中的种子v通过子图中的独立路径激活u的概率。对于S中的两个种子v和w，子图和中v和w激活u的独立路径是相互独立的。因此，我们可以将种子集S中所有节点的概率相加，得到。同理，将种子集S中所有节点的概率相加，得到。

种子集S的正传播函数估计如下:

****

我们的目标是找到大小为k的种子集S：



### 计算传播增量

在我们的算法中，我们采用贪心方法来选择最优的种子集。该算法将种子集S初始化为空集。然后依次选择传播增量最大的节点加入到S中。这个种子选择过程不断迭代，直到种子集的大小达到k。为了避免模拟影响扩散来选择最佳候选种子节点，我们在传播函数的基础上定义了正影响传播增量。

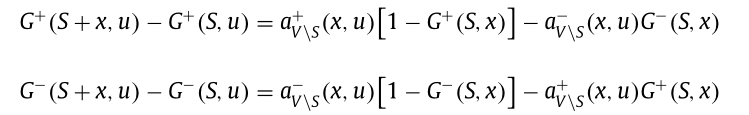
定义5. 正影响传播增量

设S为部分种子集，x∈V \ S为候选种子节点。将x加入S后的正传播增量定义为:

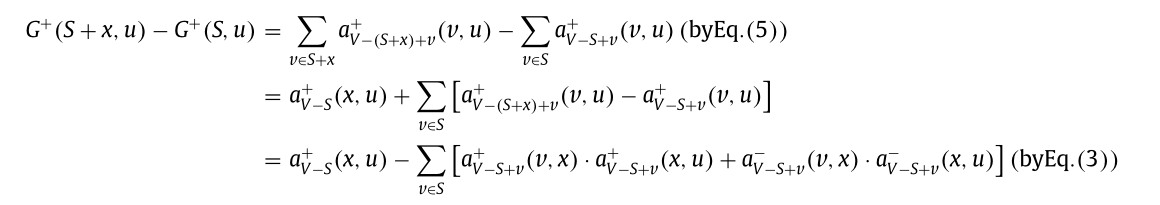


首先，我们提出以下定理来分析节点u在S中加入x后的激活概率。

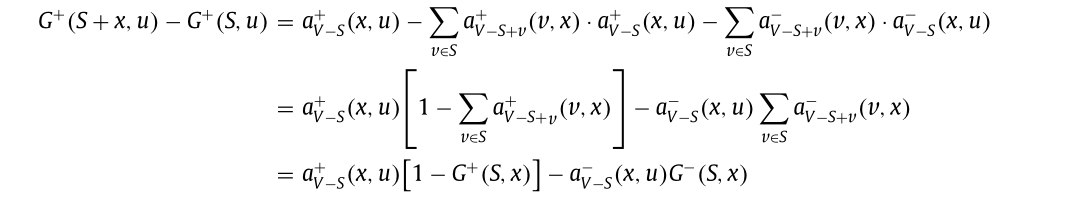
定理1. 设S为部分种子集，x∈V \ S为候选节点。对于节点u∈V，有:

****

**证明：**

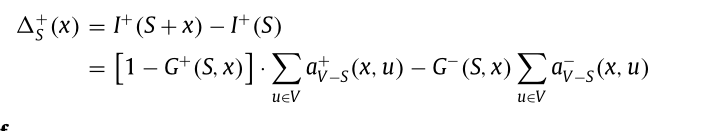
****

在上式中，考虑了子图中v到u经过x的所有独立正路径，即形式的路径。因为这些路径中没有循环，所以除了起始点，节点v不会出现在每条路径上。即v不能出现在的子路径上，所以这里的和可以分别替换为和。因此，我们可以得到：

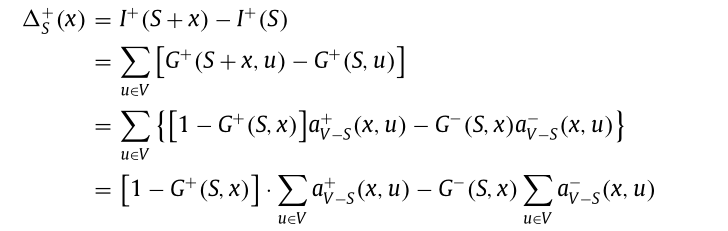


根据定理1，我们提出以下定理来得到的估计。

**定理2.** 令是当前种子集，x是V \ S中的一个候选节点。将x加到S的正传播增量是

****

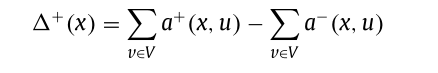
**证明：**

****

### PIMSN(符号网络中的积极影响最大化)的框架

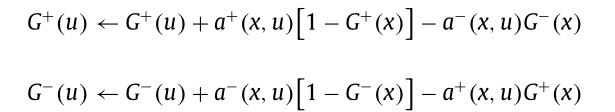
该算法首先通过调用算法Comp-pp计算节点对的激活概率和。算法PIMSN基于定义的传播函数，采用贪心策略选择传播增量最大的节点作为候选种子。

该算法将种子集S初始化为空集。我们分别用变量和表示当前种子集S下和的激活概率。首先，利用Compp-pp算法计算所有节点对(v, u)之间的正激活概率和负激活概率;我们用变量和分别表示当前种子集S正、负激活u的总概率和。由于种子集S最初是空的，因此和的初始值为0。我们用变量表示，它是在当前种子集S中加入x时正传播的增量。的初始值为

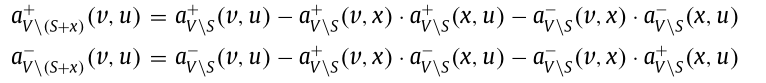
****

在每次迭代中，从集合V \ S中选择最大的节点x加入种子集。将x加入种子集S后，变量、、和和的值相应更新：

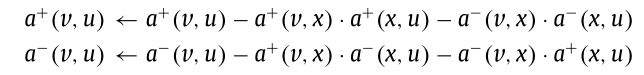
将x加入种子集S后，和的值更新如下:

****

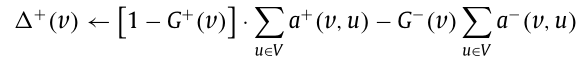
将集合W + x替换为V \ S，将W替换为V \ (S+x)，得到:

****

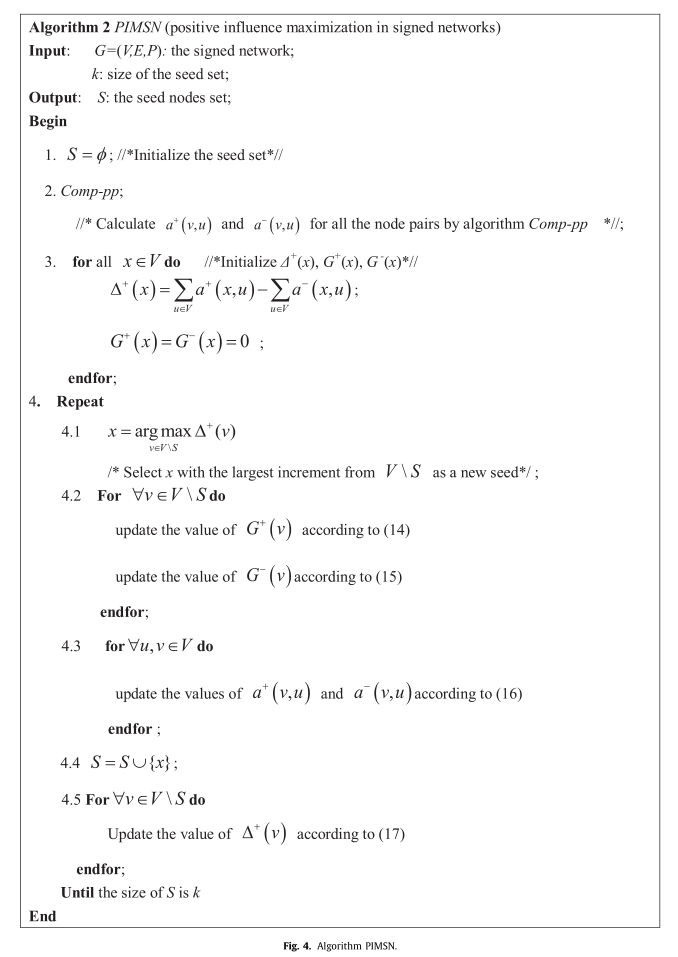
将x加入种子集S后，和的值需要更新如下:

****

利用更新后的、、和的值，可以更新正扩散的增量：

****

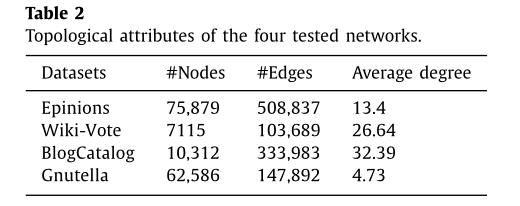
在下一次迭代中，可以选择正传播增量最大的新种子。这个迭代过程不断重复，直到种子集中的节点数达到k。

****

## 实验解析

在四个真实的网络上对算法PIM-SN的有效性和性能进行了测试，四个网络分别为：Epinions, Wiki-Vote, BlogCatalog和Gnutella。

|  |  |
| --- | --- |
| Epinions | 一个消费者可以讨论商品的网站。他们也可以对其他用户的评论给出自己的意见。根据他们的观点，可以定义他们之间的积极和消极关系。 |
| Wiki-Vote | 它在维基百科上记录了投票历史。网络中的节点代表维基百科的用户，从用户u到v的签名链接表示用户投票赞成或反对向管理部门推荐v。 |
| BlogCatalog | 社会博客目录网站，其中每个博客都与信息相关联，如在社会网络中爬行的积极和消极关系以及各种群体的成员。节点代表用户，无向链路表示用户之间的正、负关系。 |
| Gnutella | 第一个去中心化的点对点大规模网络。在网络中，主机由节点代表，而主机之间的互连则由链接代表。Gnutella由一个月内观察到的9个快照组成。 |

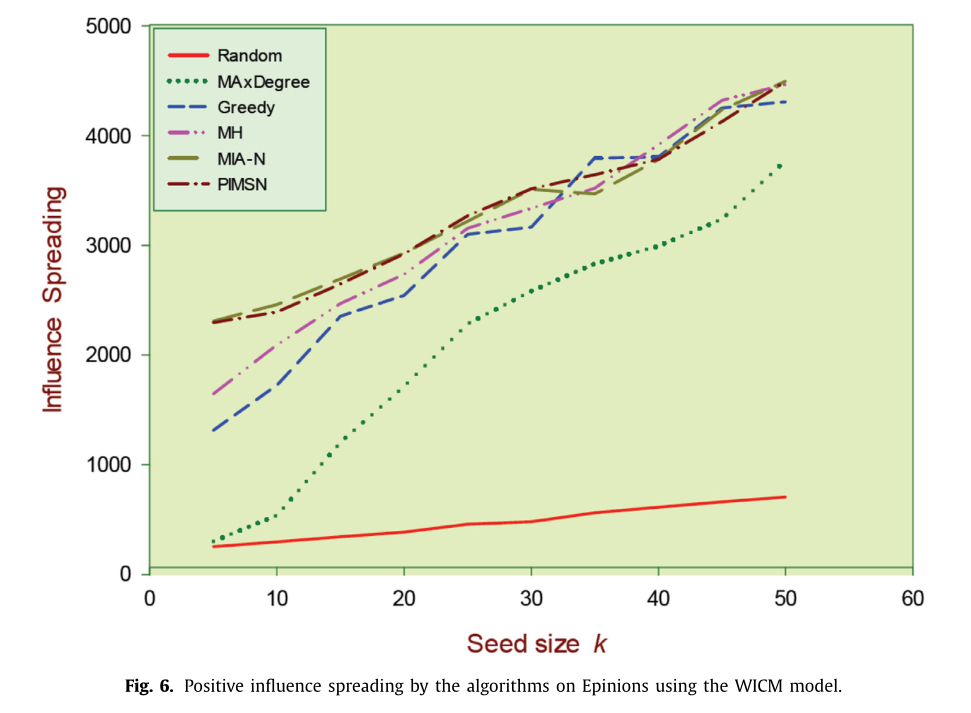


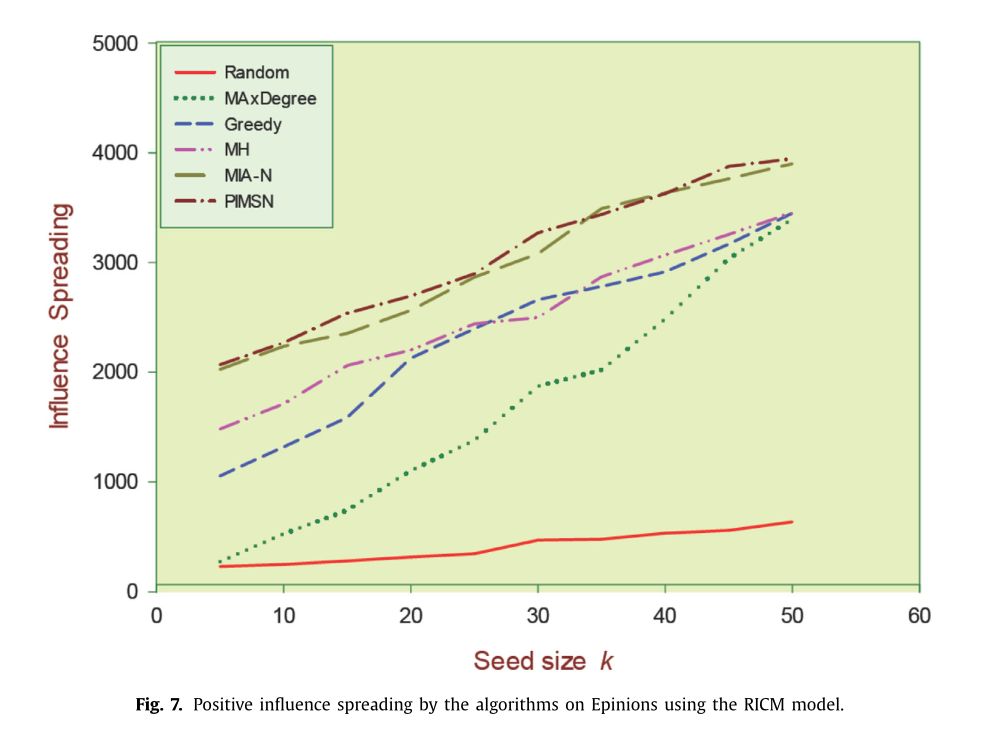
在上述数据集中，**边缘的传播概率由两种模型确定:随机IC (RICM)和加权IC (WICM)。**在RICM模型中，对每个链接(u,v)随机设置一个概率，表示节点v的活跃度。在WICM模式中，每个链路(u,v)被设定为激活v的概率为

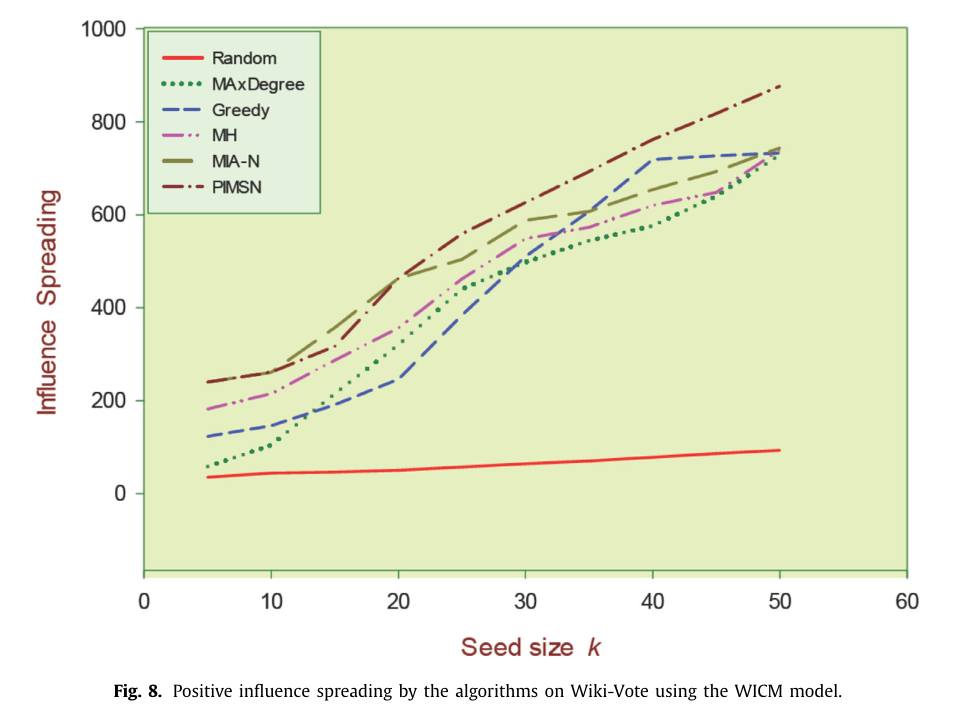
实验测试并比较了PIMSN算法与其他五种算法的影响扩散。

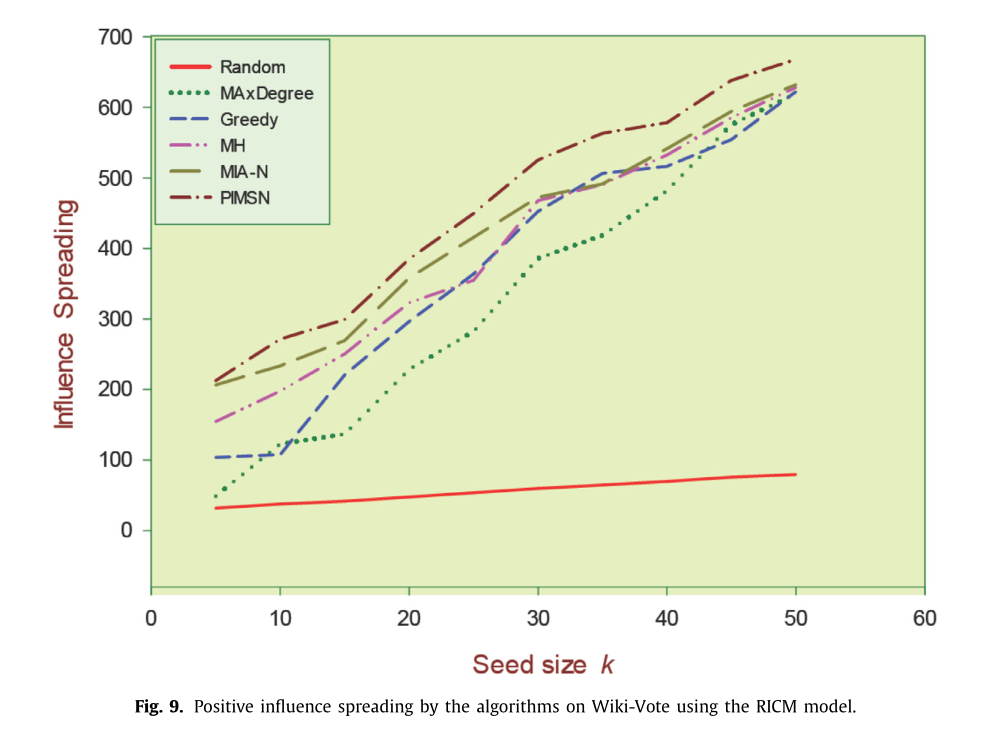
|  |  |
| --- | --- |
| Random | 这是影响最大化最常用的方法之一。该方法随机选取k个节点构建种子集。在我们的实验中，这样的随机种子选择过程重复了100 0 0次，这100 0 0个种子集的平均传播就是算法的输出结果。 |
| MaxDegree | 这是一种使用程度中心性的启发式算法。MaxDegree认为度中心性最高的节点影响最大。算法会反复选择中心性最高的节点加入到种子集中，直到种子节点数量达到k。 |
| Greedy | 该算法将种子集初始化为空集，并反复选择一个潜在的候选节点加入种子节点。在每一轮种子选择中，该方法选择影响扩散边际收益最大的节点作为候选种子。通过蒙特卡洛模拟估计节点的边际增益。 |
| Meta heuristic (MH) | Canh V. Pham提出了在d跳内传播过程中解决正面影响最大化而负面影响受限的算法(d- pimcn)。该方法将d -PIMCN问题视为0 - 1整数线性规划，并采用启发式优化方法求解 |
| MIA-N | Chen wei提出了两种对立关系网络中影响最大化的MIA-N算法。基于树形结构中的影响传播草图，MIA-N执行启发式优化，使积极影响最大化。 |

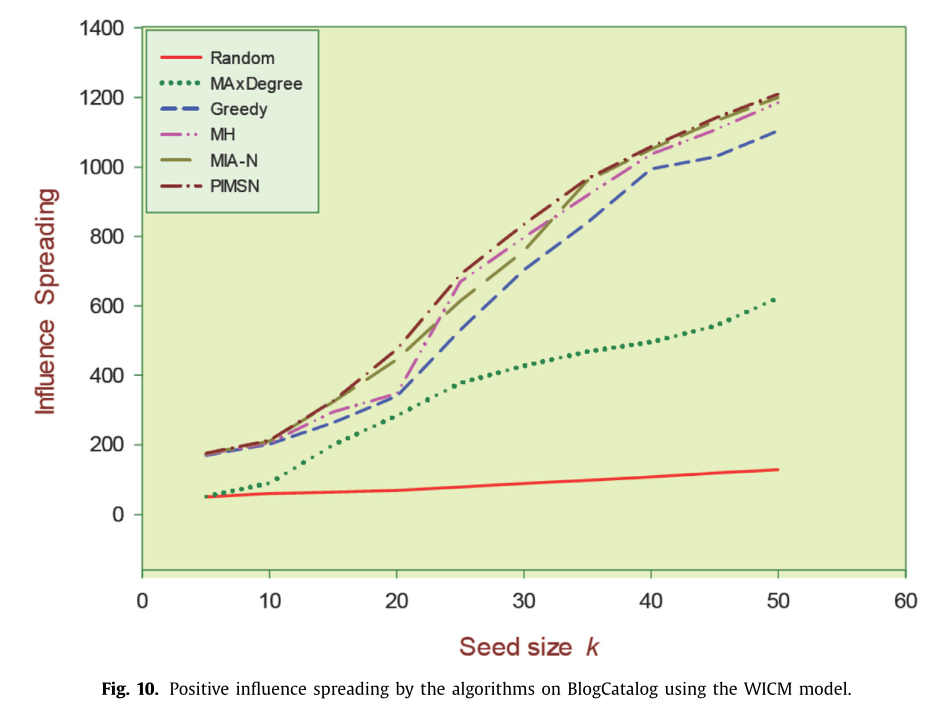
### 积极影响扩散

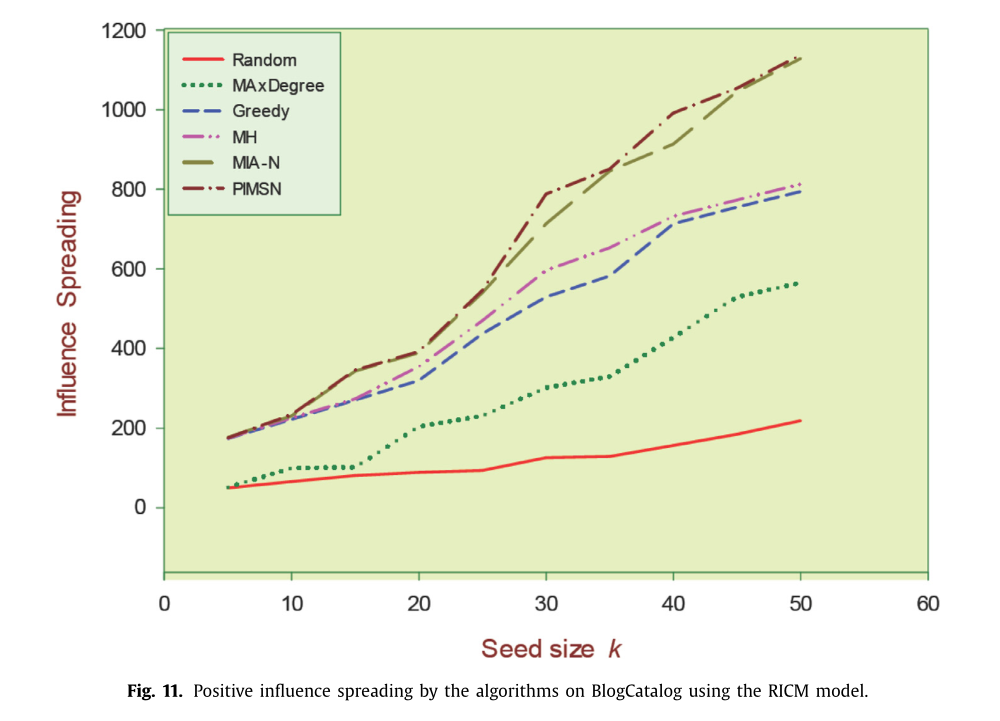


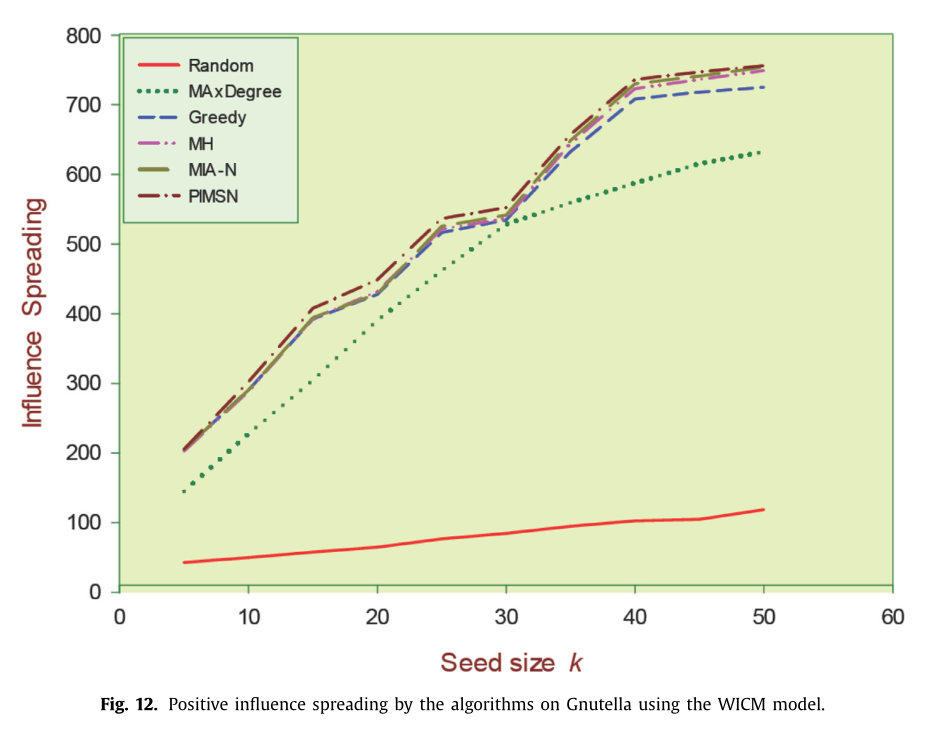


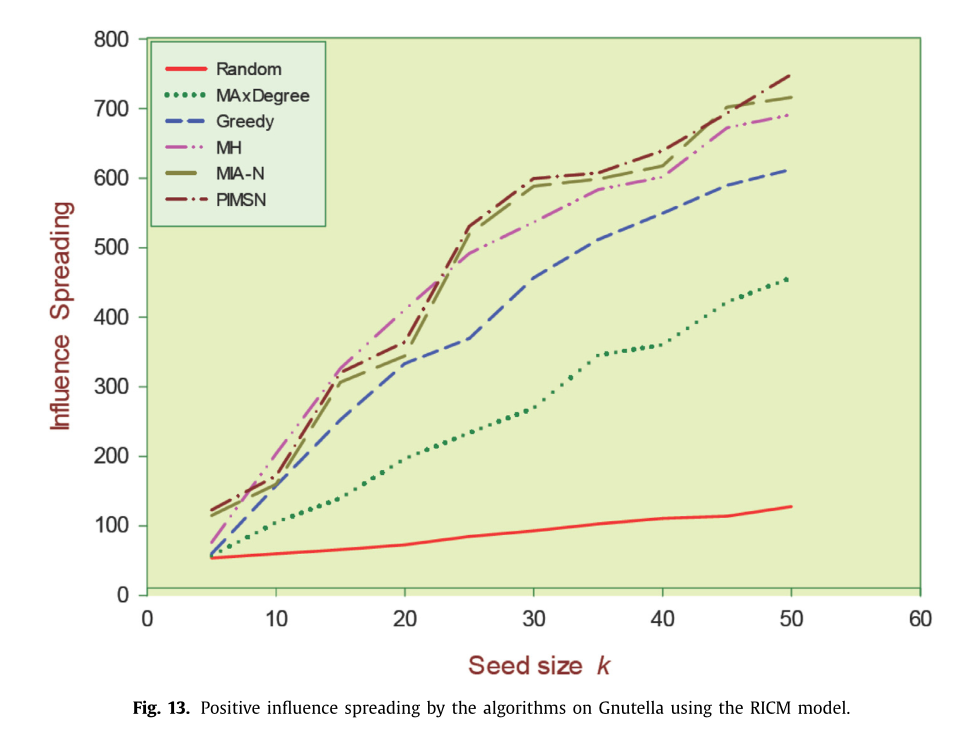




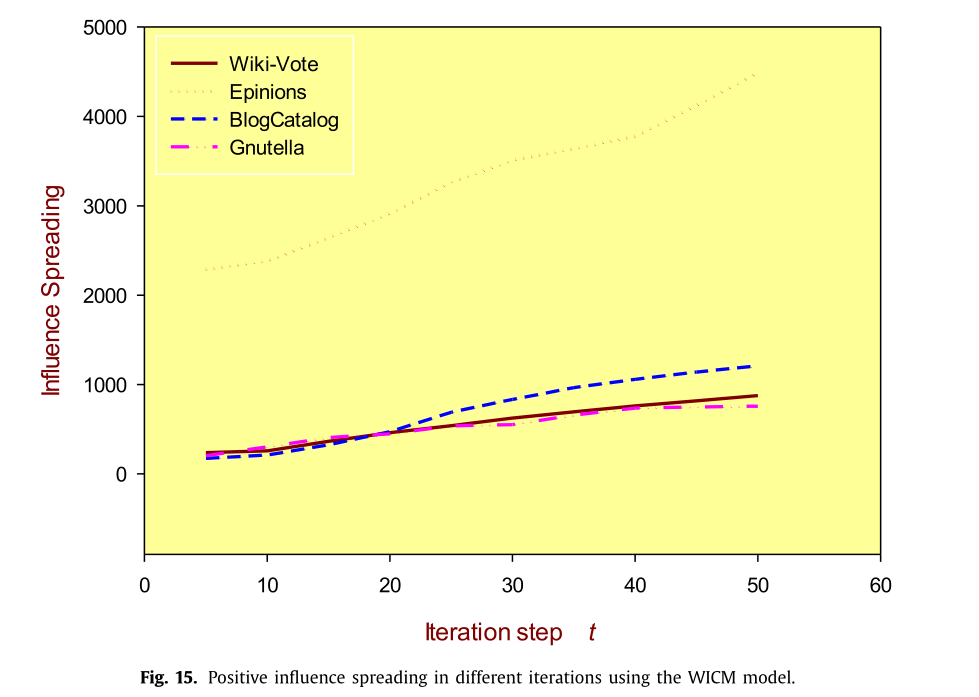
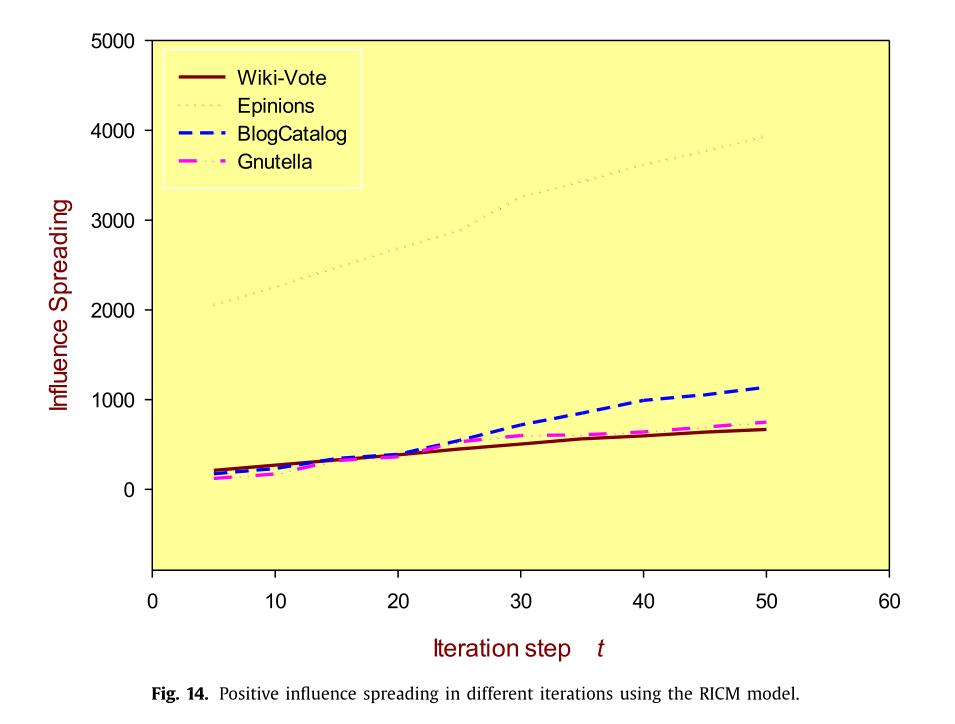




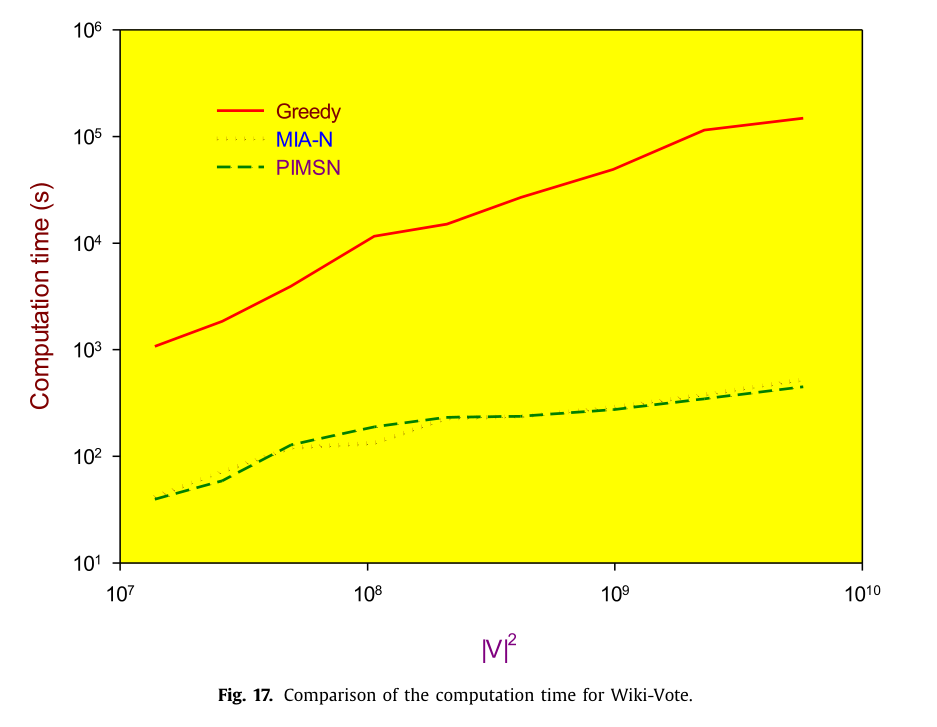
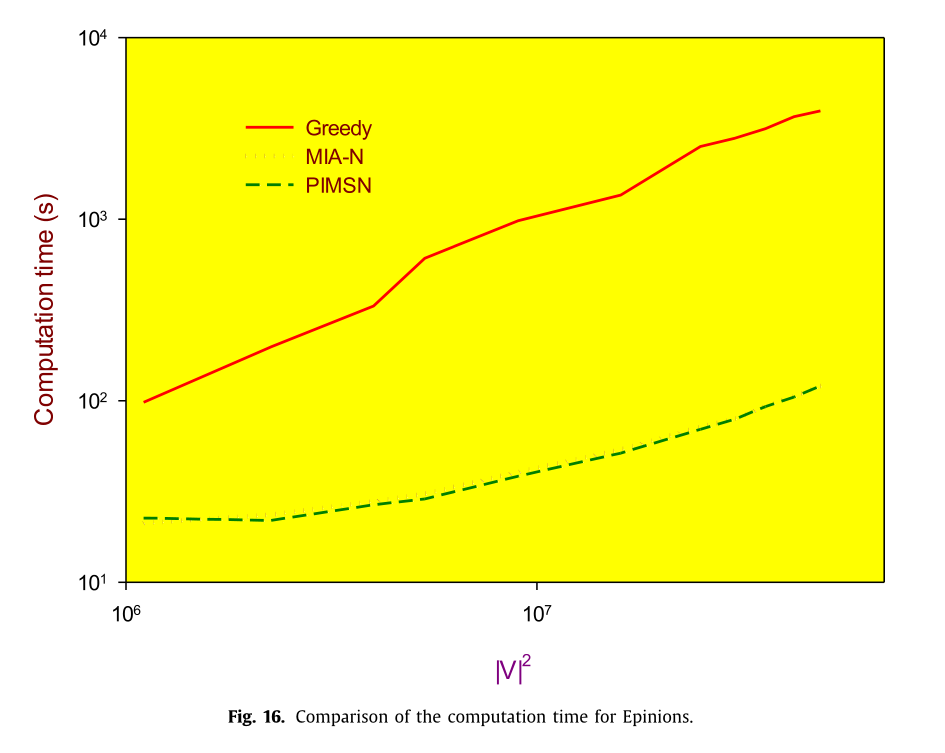


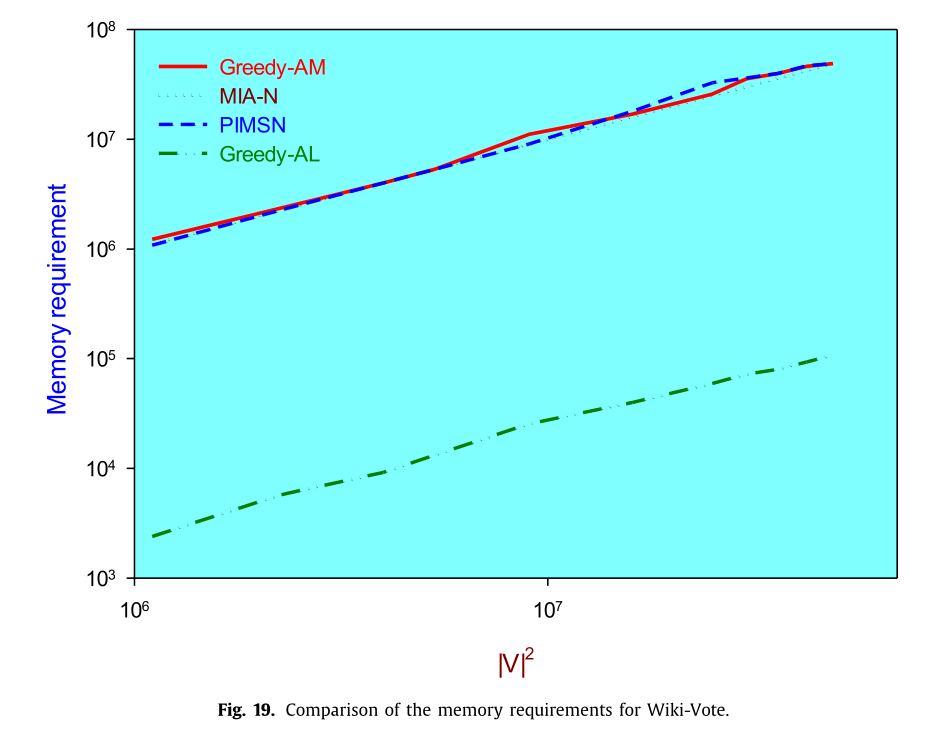
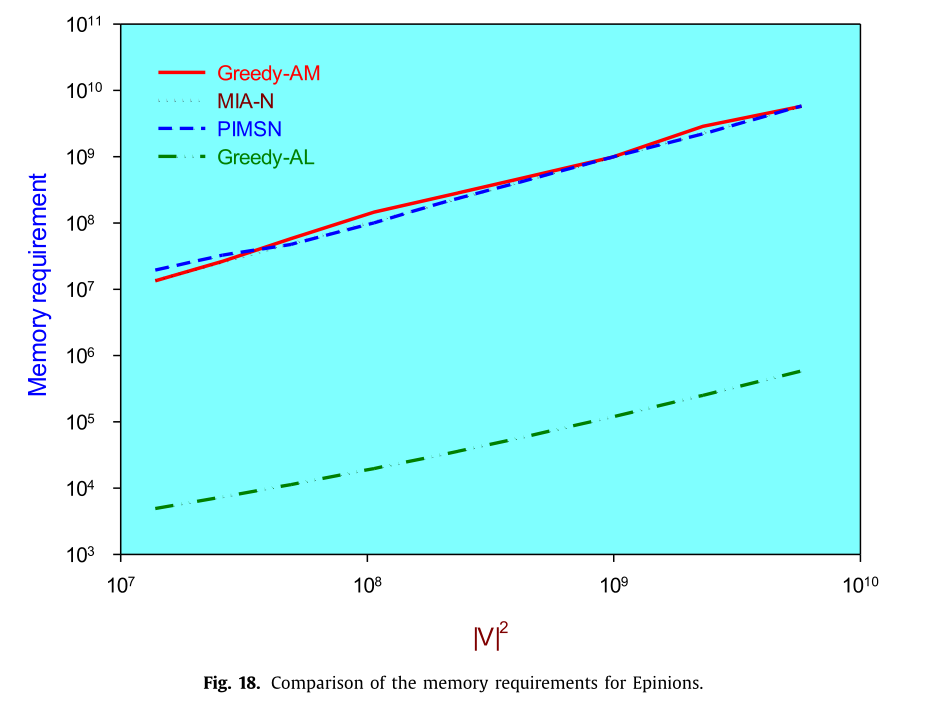


### 时间步下不同网络的影响规模变化



### 测试各算法的时间和内存需求





## 总结

在签名社交网络中，个体之间存在着积极或消极的关系。特别是，消极链接比积极链接发挥更重要的作用，因为它们可以扭转影响。如何利用不同的关系来有效地选择影响节点是有符号网络影响最大化的关键问题。