《影响力最大化》文献综述

|  |  |  |  |
| --- | --- | --- | --- |
| 文档名称： | 《影响力最大化》文献综述 | | |
| 单位名称： | 北京邮电大学网络与交换技术国家重点实验室宽带网研究中心 | | |
| 文档作者： | 王子珩 | 文档版本： | Draft1.0 |
| 创建日期： | 2019/04/22 | 修改日期： |  |

修订历史记录

|  |  |  |  |
| --- | --- | --- | --- |
| 日期 | 版本 | 说明 | 作者 |
| 2018/02/10 | Draft1.0 | 修订详细信息。 | 王子珩 |
|  | Draft2.0 |  |  |
|  | Draft3.0 | 。 |  |
|  | Draft4.0 |  |  |
|  | Draft5.0 |  |  |

目 录

[1. 简介 4](#_Toc6995471)

[1.1 编写目的 4](#_Toc6995472)

[1.2 定义 4](#_Toc6995473)

[1.3 首字母缩写词和缩略语 4](#_Toc6995474)

[1.4 概述 4](#_Toc6995475)

[2. 引言 6](#_Toc6995476)

[3. 影响力最大化的传播模型 9](#_Toc6995477)

[3.1 独立级联模型（IC）及其变种 9](#_Toc6995478)

[3.2 线性阈值模型（LT） 11](#_Toc6995479)

[3.3 触发模型（TR） 15](#_Toc6995480)

[3.4 其他模型 15](#_Toc6995481)

[3.5 小结 16](#_Toc6995482)

[4. 影响力最大化的算法 17](#_Toc6995483)

[4.1 基于模拟的方法 17](#_Toc6995484)

[4.2 基于代理的方法 17](#_Toc6995485)

[4.3 基于草图的方法 17](#_Toc6995486)

[4.4 小结 17](#_Toc6995487)

[5. 上下文感知的影响力最大化 18](#_Toc6995488)

[6. 总结 19](#_Toc6995489)

[7. 参考文献 20](#_Toc6995490)

[8. 附录 23](#_Toc6995491)

# 1. 简介

## 1.1 编写目的

本文针对国内外影响力最大化的主要研究方法和相关研究现状进行了总结，对现有研究所采用的理论方法和主要思路进行了较详细的分析。

## 1.2 定义

|  |  |
| --- | --- |
| **英文全称或缩写** | **定义** |
| Diffusion Model | 给定一个社交网络G=(V,E)和一个用户集合S，传播模型M捕获了S在G上传播信息的随机过程 |
| Influence Spread | 一个用户集合S的影响传播表示集合S所能影响到的用户数。 |
| Influence Maximization | 给出一个图G=(V,E)，一个扩散模型M和一个正整数k，IM算法的目标是从V中挑出k个节点组成一个集合S使得它的影响范围最广。 |
| RR Set | 给定一个图G’为图G上的草图，RRG’（v）表示在G’中可以到达v的所有结点。 |

## 1.3 首字母缩写词和缩略语

|  |  |  |
| --- | --- | --- |
| **缩略语** | **英文全称** | **中文解释** |
| IM | Influence Maximization | 影响力最大化 |
|  |  |  |

## 1.4 概述

本研究报告分为概述、前言、国内外研究现状和发展趋势以及总结四个部分：

第一章，概述部分简要说明本文的编写目的，对一些定义和缩略语进行解释，概括整个报告的总体内容；

第二章，前言部分简要阐述IM问题的重要性。

第三章，给出近年来IM问题在国内外研究现状和发展趋势，主要介绍IM的模型。

第四章，给出近年来IM问题在国内外研究现状和发展趋势，主要介绍IM的算法。

第五章，

第六章，

第七章，

第八章。

# 2. 引言

过去的几十年见证了在线互动的蓬勃发展，产生了数量空前的内容。在线社交网络的普及引起了人们对信息传播的关注，因为一条信息可以通过在网络上朋友之间的“口碑”传播迅速变得无处不在。这种扩散现象已经在许多的应用中被证明是强大的，例如采用政治立场和技术创新。最近的一个例子是唐纳德·特朗普的总统竞选。2016年，Twitter几乎每天都被用作竞选工具。因此，在线社交网络中的信息传播吸引了计算机科学、物理学、流行病学等多个领域的广泛研究[1]

随着科学技术的发展，越来越多的个人和公司使用社交网络来促进他们的业务，传播新闻、思想和观点，社交网络有许多研究应用，如链接预测、谣言控制、影响力最大化。然而，大多数研究工作只考虑单一影响的传播。事实上，不同的观点甚至相反的观点在社交网络中同时传播的情况经常发生。一个显然的问题是，当社会网络中存在竞争时，影响力如何传播，我们称之为影响力竞争问题。影响力阻塞最大化( IBM )问题是一种影响力竞争问题。IBM问题的目标是尽可能阻止竞争对手的影响力的扩散[2]。

# 3. 影响力最大化的传播模型

传播模型在影响力最大化问题中非常重要，不同的模型应用不同的机制来捕获用户如何从非活跃的状态到激活的状态。

## 3.1 独立级联模型（IC）及其变种

独立级联模型是一个非常经典并且被研究很多的模型[1]。

传统的独立级联模型的主要思想是当一个节点u被激活时，它会以概率p(u,v)对它未激活的出边邻居节点v尝试激活，这种尝试仅仅进行一次，而且这些尝试之间是互相独立的，即u对v的激活不会受到其他节点的影响。

它的传播过程如下：

·给定初始的活跃节点集合S，当在时刻t节点u被激活后，它就获得了一次对它的邻居节点v产生影响的机会，成功的概率为p(u,v)，是随机赋予的系统参数。

·若v有多个邻居节点都是新近被激活的节点，那么这些节点将以任意顺序尝试激活节点v。如果节点u成功激活节点v，那么在t+1时刻，节点v转为活跃状态。

·在t+1时刻，节点v将对其他节点产生影响，重复上述过程。

·结束条件：网络中不存在有影响力的网络节点。

在文献[2]中，作者提出了竞争的独立级联模型（CIC）：一个社交网络G，每个节点有三种状态，未激活，正向状态，反向状态。每个边有正向传播概率PP（u,v）和负向传播概率PN（u,v）一开始有两个种子集合，一个是正向种子集合一个是负向种子集合。两种传播方式各自独立。在t步时，一个未激活的节点u，变为负向的，他有一次机会对他的未激活邻居节点进行负向传播，反之亦然。如果一个节点被影响到了了，它在t+1时刻被激活。U只有一次机会激活别人。如果在某一步时，没有新的节点被激活，传播结束。

在文献[3]中作者提出了延迟感知的独立级联模型（LAIC）：通过考虑时间延迟来扩展传统IC模式，这与现实社会网络更加一致。在这个模型下，每个结点的状态分为三种，激活，延迟激活，非激活。对于一个在种子集合中的结点u，可以认为他在t=0时刻处于活跃态，如果它成功影响到了一个非活跃邻居节点v，那么节点v变为延迟活跃节点。根据之前定义的影响力延迟的分布，每条边e（u，v）可以扩展到，其中代表了一个非活跃节点v被他的活跃邻居u在一个激活延迟以概率激活。是一个延迟分布。

LAIC模型的过程如下:在任意一步t>0，在种子集S中的结点u可以有一次激活邻居节点v（v是处于非激活状态或者延迟激活状态）与传统IC模型不同的是，被激活的节点v会变味延迟激活状态而不是激活状态，在经过延迟之后，延迟激活节点v会被种子集合S激活。

## 3.2 线性阈值模型（LT）

线性阈值模型的主要思想它对图中每个节点v都有一个激活阈值，当v的入邻居影响力大于它的阈值时，就会被激活。

传统线性阈值模型的传播过程如下：

·集合中的任意节点v随机分配阈值θ[v]∈[0,1]。只有当节点v的新处于激活状态的邻居节点对它的影响力大于该阈值时，节点v才能被激活。

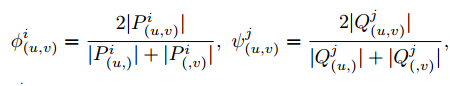
·用权值b[u.v]表示节点v被它的邻居节点u的影响，∑u∈in(v)  b[u,v]≤1表示节点v的处于活跃状态的邻居节点对它的影响力之和。这里in(v)是v的入边邻居节点集合。

·给定初始的活跃节点集合S，在t时刻，所有在t-1时刻处于活跃状态的节点仍保持活跃，并且当这一时刻节点v的邻居节点的影响力之和大于节点b的阈值时，节点v被激活.

·节点v被激活后，下一时刻将对它的邻居节点产生影响，重复上述过程。

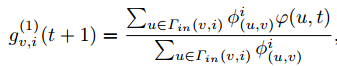
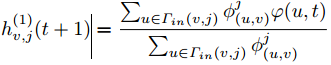
·结束条件：当网络中已存在的所有活跃节点中任意活跃节点的影响力之和都不能激活他们的处于非活跃状态的邻居节点时。

在文献[4]中作者将扩展传统的线性阈值(LT)模型来处理多对齐多关系网络(MMNs)的信息扩散，提出了M&M模型。基于MMNs M = (U, E, R)，通过pathsim估计每对具有不同扩散关系的用户的权重。式中，用户u与v之间的网内(网间)扩散权与关系i(j)定义为:



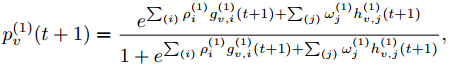
其中是为网内(网间)扩散元路径实例集，根据关系i(j)从u开始到v结束。表示以用户u、v分别作为起始用户和结束用户的元路径实例数。

在传统LT模型的基础上，影响在网络中以离散的步骤传播。在步骤t中，所有激活用户都保持活动状态，如果接收到的影响超过其阈值，则可以激活非激活用户。只有在步骤t时被激活的用户才会影响到其在步骤t+1时的邻居，且在一个网络中(如G(1))用户v在网络内关系i和网络间关系j下的激活概率分别为

其中是用户v在关系i和关系j下的邻居集合。表示用户u在时间t被激活。

通过对网络内部和网络间的各种关系进行聚合，可以得到v(1)的整体激活概率



代表了每个关系的权重，满足

## 3.3 触发模型（TR）

触发模型的主要思想是结合了IC模型和LT模型。在扩散过程中，TR模型会节点v随机选择一个触发集合Tv。

它的传播过程如下;

·定义一个初始集合S，为每个节点随机选择一个触发集合。

·非激活的v在t时刻会被激活的条件为：v的一个属于Tv中的邻居在t-1时刻被激活。

·被激活结点继续上述传播。

## 3.4 其他模型

由于贪心算法代价太大，文献[5]提出了影响力最大化树状结构（MIA）来计算影响的传播。MIA的主要思想如下：给定一个社交网络G，两个结点u，v属于结点集合V，结点u可以影响结点v的条件是两个结点之间有一条路径，这条路径可以表示为，结点u能激活v的概率可以表示为。在MIA模型下，当u和v之间存在多条路径时，选择概率最大的路径，因为通过这条路径u有最大的几率去激活v，作者将这条路径定义为。在MIA模型下，结点u只能通过去影响结点v。

## 3.5 小结

0

# 4. 影响力最大化的算法

虽然影响力最大化问题的计算是非常复杂的，但是当影响函数满足单调性和子模性的时候，该问题可以得出近似的最优解。影响函数的单调性和子模性的定义如下：

单调性：一个影响函数是单调的当且仅当它满足，直观来说就是在集合S中加入一个结点后，它的影响范围不会减少。

子模性：一个影响函数是子模的当且仅当它满足如下所示的条件：

。直观来说就是集合S中加入一个结点，它所带来的边际收益要小于上一个结点的加入带来的边际收益。

影响力最大化算法大都是基于贪心算法的改进。贪心算法的思想是首先初始化一个空的种子集S，在每次迭代中都选择一个有最大边际收益的结点加入到S中，当种子集S中已经有k个种子的时候算法停止。

## 4.1 基于模拟的方法

基于模拟的主要思想是利用蒙特卡洛模拟去估计每个节点的影响范围函数进而计算每个结点的边际收益，之后使用贪心算法每次迭代都选择边际收益最高的结点加入种子集合。为了提高估算边际收益的精确度，此算法一般要跑10000轮再取平均。显然，由于需要进行的蒙特卡洛模拟次数过多，传统的基于模拟的方法无法在一个规模较大的图中应用。所以有学者对其进行改进。

在文献[7]中作者提出了UBLF算法。作者使用代表一个节点v在步骤t被种子集S激活的概率。影响函数可以表示为。对于t=1,2,…N-|S|，可以得到这样一个不等式。随后作者利用矩阵形式对其进行简化，PP是传播矩阵，pp(u,v)是对应的两个节点之间的传播概率。我们使用一个行向量来表示未激活节点在t时刻被激活的概率。上述不等式可以表示为。影响函数的上界可以表示为。UBLF的算法步骤如下：1. 将带有传播概率的扩散网络信息转换成传播概率矩阵PP。

2. 以PP为处理对象，利用上界计算理论公式，计算每个节点作为初始扩散点的传播范围上界。

3. 依次选取上界较大的节点进行蒙特卡洛模拟，用真实值进行序列对比；若某节点v的真实值比其它点的上界值还要大，那么节点就入选整个网络的关键节点集，如此直到选满k个节点为止。

## 4.2 基于代理的方法

用代理模型的方法去计算影响传播函数，而不是用很繁琐的蒙特卡洛模拟，这样可以使得IM算法在大规模网络中有很好的延展性。基于代理的方法一般都是为特定的传播模型量身定做的，它的通用性不好。

在文献[9]中作者提出了SimPath算法，他是基于LT模型。作者首先使用经典的回溯算法找到所有的简单路径，进而计算出种子集合的影响函数。有了种子集合的影响函数，作者将它代入贪婪算法(用CELF优化)中，得到一个影响力最大化的算法。这构成了SIMPATH算法的核心。为了进一步提高SIMPATH的效率，作者提出了两个优化方法，第一个是顶点覆盖优化也就是找到最小顶点覆盖，可以减少影响函数的计算量，第二个是一个前向的优化。

## 4.3 基于草图的方法

这种方法的主要焦点是提高基于模拟的方法的理论效率，同时使得近似值可以得到保证。为了避免运行大量模拟，基于草图的方法基于特定的扩散模型预先计算多个草图，然后利用草图来评估影响扩散。它主要分为两大类，前向影响草图和反向可达草图。

### 4.3.1 前向影响草图

前向影响草图的主要思想与蒙特卡洛方法相似[1]。在一个图G=(V,E)中，假设每条边上的传播概率为P1,P2,,,Pn。利用蒙特卡洛模拟计算一个节点的影响范围时，首先将图中所有边以1-Pk的概率移除，在这个子图中计算当前节点的影响范围。S的影响传播函数为，其中Gi为去某条边之后的子图，为在子图Gi中种子集合S可以到达的结点的集合。

### 4.3.2 反向可达草图

文献[8]中，作者提出了两阶段的影响力最大化算法（TIM）。

## 4.4 小结

异构

# 5. 上下文感知的影响力最大化

现如今，

# 6. 总结

0

# 7. 参考文献

1. Li, Y., Fan, J., Wang, Y., & Tan, K.-L. (2018). Influence Maximization on Social Graphs: A Survey. IEEE Transactions on Knowledge and Data Engineering, 1–1. doi:10.1109/tkde.2018.2807843
2. Zhu, W., Yang, W., Xuan, S., Man, D., Wang, W., & Du, X. (2018). Location-Aware Influence Blocking Maximization in Social Networks. IEEE Access, 1–1. doi:10.1109/access.2018.2876141
3. Liqing, Q., Jinfeng, Y., Xin, F., Wei, J., & Wenwen, G. (2019). Analysis of influence maximization in temporal social networks. IEEE Access, 1–1. doi:10.1109/access.2019.2894155
4. Zhan, Q., Zhang, J., Wang, S., Yu, P. S., & Xie, J. (2015). Influence Maximization Across Partially Aligned Heterogenous Social Networks. Lecture Notes in Computer Science, 58–69. doi:10.1007/978-3-319-18038-0\_5
5. Chen, W., Wang, C., & Wang, Y. (2010). Scalable influence maximization for prevalent viral marketing in large-scale social networks. Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’10. doi:10.1145/1835804.1835934
6. 123
7. Zhou, C., Zhang, P., Zang, W., & Guo, L. (2015). On the Upper Bounds of Spread for Greedy Algorithms in Social Network Influence Maximization. IEEE Transactions on Knowledge and Data Engineering, 27(10), 2770–2783. doi:10.1109/tkde.2015.2419659
8. Tang, Y., Xiao, X., & Shi, Y. (2014). Influence maximization. Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data - SIGMOD ’14. doi:10.1145/2588555.2593670
9. Goyal, Amit, Wei Lu, and Laks VS Lakshmanan. "Simpath: An efficient algorithm for influence maximization under the linear threshold model." 2011 IEEE 11th international conference on data mining. IEEE, 2011.

# 8. 附录

该附录介绍