Location-Based Seeds Selection for influence blocking maximation in social network

**Abstract**

在竞争激烈的社交网络中，影响阻塞最大化(IBM)是病毒式营销的一个关键问题。虽然IBM问题已经得到了广泛的研究，但是现有的工作忽略了位置信息在影响传播中的重要作用。本文研究了IBM问题的基于位置的种子选择问题，其目的是在给定的查询区域内找到一个积极的种子集，以尽可能地阻止在给定的块区域内负面影响的传播。针对基于仿真的贪心算法效率低的问题，提出了一种启发式算法IS-LSS及其改进版IS-LSSC均基于最大影响树状结构和四叉树指数。

**Introduction**

一个自然的问题是，当两个竞争对手同时在传播自己的影响力时，如何尽可能地阻止竞争对手的影响力扩散。我们将此称为影响阻塞最大化(IBM)问题，He等人[5]首次提出了IBM问题，一些研究者也从不同的角度考虑了这个问题。

**Related work**

**Influence blocking maximization**

**Preliminaries**

**A propagation model**

一个竞争的社交网络用G=(V,E)表示。作者为每个边eu,v分配了两个权值ppu,v和pnu,v来模拟用户u对用户v的正面和负面影响。

传统IM问题中最常用的模型是IC模型和LT模型，但这些模型不适用于竞争影响的传播。应该用竞争独立级联模型和竞争线性阈值模型。在本文中，作者使用齐次CIC[10]模型作为CIC模型的扩展，来模拟竞争社交网络下的影响传播。

给定一个带竞争的社交网络G=(V,E)，初始负向种子集SN和正向种子集SP，CIC模型的工作原理如下：用SNt和SPt表示在步骤t被负激活和正激活的种子集合。在t+1步，在SNt中的节点u有一次负激活他未激活邻居的机会以概率pnu,v，在SPt中的节点u有一次正激活他未激活邻居的机会，以概率ppu,v。如果在步骤t中，v同时被正传播和负传播激活，负传播占主导地位，v被负传播激活。负面主导原则与谣言通常难以反驳的常识相符。我们认为一个CIC模型是同质的如果pnu,v=ppu,v。此后，当没有歧义时，作者用pu,v表示u到v的正传播概率和负传播概率。

**B MIA structure**

用最大影响路径评估一个节点到另一个节点的影响力。对于节点对u和v，从u到v有许多路径。u到v的最大影响路径为所有路径中概率最大的那一条。可以将MIP转化为迪杰斯特拉问题。

**C influence blocking maximization**

**定义1** influence blocking set：正影响集合SP的影响阻塞集合表示为IBS(Sp,SN)，定义为如果正种子集为空，则将被负激活的节点集，如果正影响集合不为空则不被负影响激活。

**定义2** blocked negative influence：对于一个正向影响集合SP的阻塞负影响被表示为是IBS(Sp,SN)的数量。

**定义3** Influence Blocking Maximization：给定一个竞争的社交网络G=(V,E)，一个负向种子集合SN，一个竞争的传播模型，影响阻塞最大化问题的目标是找出一个正向影响集合SP，使得阻塞负向影响最大。

**4 LSSIBM problem and the greedy algorithm**

作者首先定义了区域限制的影响阻塞集合(RIBS)和区域限制的阻塞负影响(RBNI)。

在基于位置的竞争社交网络中，每个节点v都有一个地理位置(x，y)。给定一个查询，RQ是一个用于正种子选择的查询区域，RB为阻断负种子影响传播的阻塞区，k为正种子集的大小。作者定义区域约束阻塞影响集如下

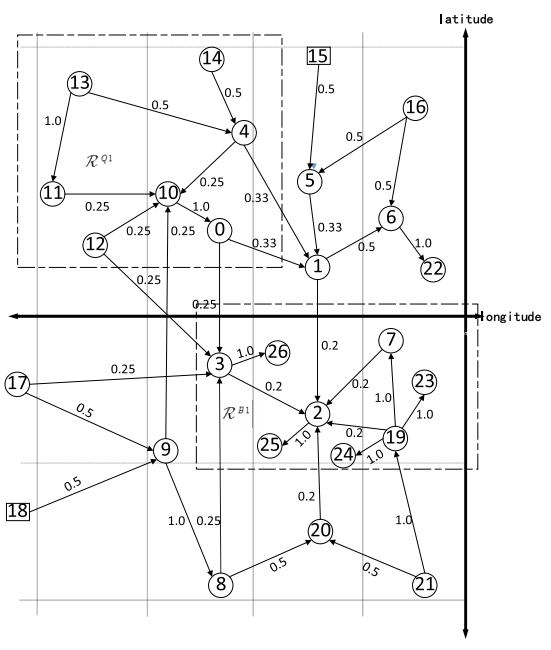
**定义4** **Region-Constrained Influence Blocking Set**：给定一个竞争的社交网络G=(V,E)，一个负影响集SN，一个竞争模型，一个查询区域RQ，一个阻塞区域RB，SP的Region-Constrained Influence Blocking Set表示为RIBS（SP,SN,RQ,RB）定义为位于RB中的节点集，当正种子集为空时，该节点集将被负激活，当正种子集为SP时，这些节点就不会被负激活。

定义5 Region-Constrained Blocked Negative Influence：给定一个竞争的社交网络G=(V,E)，一个负影响集SN，一个竞争模型，一个查询区域RQ，一个阻塞区域RB，一个SP的Region-Constrained Blocked Negative Influence表示为，定义为：

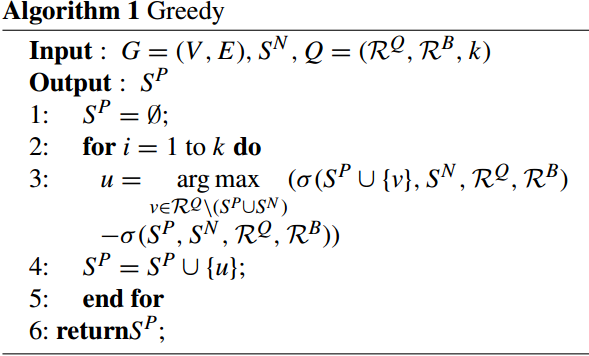


**定义6 Location-Based Seeds Selection For Influence Blocking Maximization**：给定一个竞争的社交网络G=(V,E)，一个负影响集SN，一个竞争模型，一个查询区域RQ，LSSIBM是要在查询区域Q内找到一个正向影响种子集合SP\*，使得Region-Constrained Blocked Negative Influence最大。

例子1：



在每次迭代中，贪心算法模拟均匀CIC模型下的影响传播，选取边际值最大的种子作为当前种子，直到正种子集大小为k。然而，贪婪算法使用蒙特卡罗模拟来选择最优的种子，即使使用最先进的优化方法，如CELF，在实验中，贪心算法需要花费超过20个小时来处理4000个节点的社交网络。在下一节中，作者将使用新的算法IS-LSS来解决这个效率问题。



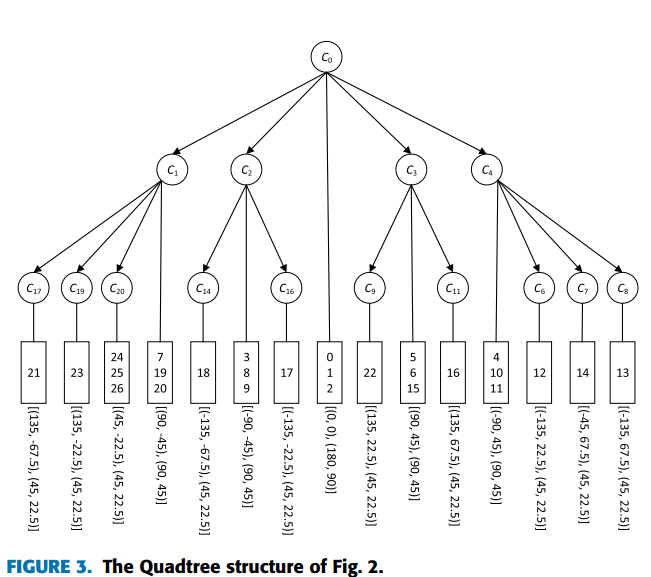
**5. IS-LSS algorithm**

**A The Quadtree structure**

四叉树是一种非常有用的二维点数据索引方法, 在四叉树中，每个四叉树单元形成如下形式

。Cid是一个四叉树单元格编号。M是单元格内位置的最小边界矩形，，x和y是M的边界中心的经度和纬度，xlim和ylim是M的长度和宽度的一半。NSET是位于M中的节点集合。SE,SW,NE,NW是指向子单元格的指针。此外，Quadtree的单元容量c限制了每个单元中节点的最大数量。当一个单元在插入后包含超过c个节点时，该单元根据其MBR的象限分成4个单元。

图3显示了基于地理位置的社交网路的四叉树结构



C0意味着四叉树单元格0，矩形中的值表示位于相应MBR中的节点，

**B the influence set**

社会影响是通过连接节点的路径来体现的。受MIA结构[11]的启发，我们在算法中使用了影响集，影响集包含两个部分，分别称为influencee set和influencer set。