**Influence maximization Across Partially Aligned Heterogenous Social Networks**

**Abstract：**

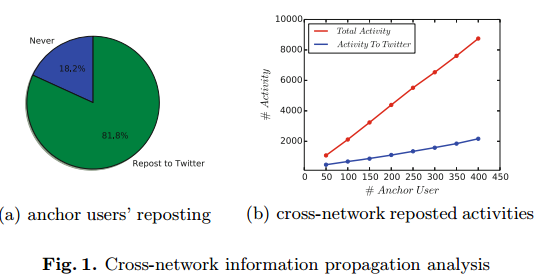
影响最大化问题的目的是寻找一个种子用户子集谁可以最大限度地扩大影响在在线社交网络(OSNs)。现有的工作大多集中在一个单一的同质网络。但是在现实世界中，OSNs(1)通常是异构的，用户可以通过多个渠道相互影响;(2)共享公共用户，信息可以通过这些用户在网络中传播。

本文首次研究了多部分对齐异构OSNs的影响最大化问题。针对这一问题，提出了一种新的多对齐多关系网络影响最大化模型。

**Introduction**

传统的病毒式营销问题的目的是选择一组种子用户，仅仅基于一个单一的社交网络中用户之间的社会关系，最大限度地对产品进行宣传。然而，在现实世界中，社交网络通常包含异构信息，例如各种类型的节点和复杂的链接，用户通过这些节点广泛连接，并有多个渠道相互影响。

目前用户通常同时参与多个社交网络，以享受更多的社交网络服务。跨多个社交网络的共享用户被称为锚用户，影响力社交网络之间传播。为了支持这一观点，我们研究了本文研究的部分对齐网络数据集（twitter和foursquare）结果如下



在图1(a)中，我们随机抽取Foursquare的一个锚用户子集，观察到500个锚用户中有409个(即， 81.8%)抽样用户已将他们的活动(如提示、地点签到等)重新发布到Twitter。同时，他们在twitter上发布的活动只占他们实际活动的一小部分。

本文研究的AHI问题非常重要，在现实社会网络中有着广泛的具体应用，跨社区甚至跨平台的产品推广和意见传播。

为了帮助说明AHI问题，我们在图2中给出了一个例子

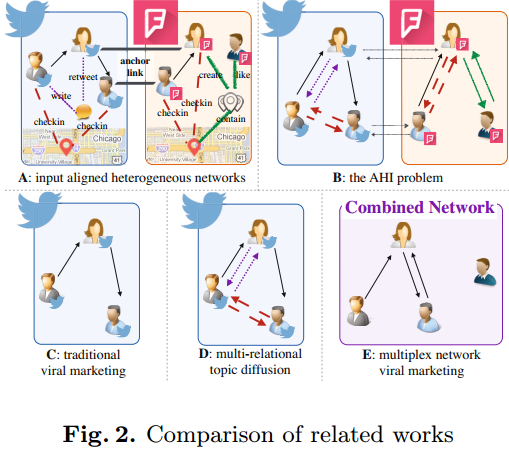


图2-A显示了两个部分对齐的异构输入网络。为了在输入网络中进行病毒式营销，解决AHI问题，我们首先提取多个影响渠道即在具有异构信息的用户之间(如传统的follow links、retweet、location checkins、anchor links等)，根据构建的多关系网络选择最优种子用户集，如图2-B所示。

AHI问题是一个全新的问题，完全不同于传统的关于信息扩散和影响最大化的研究， AHI问题中:(1)社交网络是异构的[18];(2)同时研究多个社交网络[20]，不同的异构网络可能具有不同的结构或网络模式，如图3所示;(3)本文研究的社交网络通过锚点链接[20]进行部分对齐，而不是简单地合并在一起。

AHI的难点：异构网络中的信息扩散:异构网络中的用户通过不同类型的链接广泛地联系在一起，信息可以通过不同的渠道在用户之间扩散。异构社交网络中的信息扩散建模是一项非常具有挑战性的工作；跨网络信息传播:通过锚点链接，信息可以跨网络传播。网络间信息扩散的建模仍然是一个有待解决的问题；NP-hard: AHI问题被证明是NP-hard问题，不能在多项式时间内求解。

针对上述问题，本文提出了一种新的多对齐多关系网络影响最大化模型（M&M）M&M首先基于一组网络内部和网络之间的社交元路径，提取具有跨输入OSN异构信息的多对齐多关系网络[18,20]。M&M扩展了传统的线性阈值(LT)模型来描述这些多对齐的多关系网络内部和之间的信息传播。同时影响函数也被证明是具有单调性和子模性的。

**2.problem formulation**

我们将遵循在[20]中提出的“锚用户”、“异构网络”、“对齐网络”等概念的定义。基于这些定义，AHI问题可以定义为：给定两个部分对齐的网络G(1)，G(2)以及一个在二者之间的无向锚链接。G(1)，G(2)的用户集合可以表示为U(1)，U(2)。其中Z属于两个网络用户集合的并集，这代表影响函数，也就是用户种子集合Z可以影响到的人数。AHI问题的目标就是寻找一个最优的包含d个用户的种子集Z\*去在网络中影响到最多的人。

**3.proposed Model**

**3.1 多对齐多关系网络抽取**

我们利用基于网络模式定义的元路径[18,20]，利用对齐网络中的异构信息提取多对齐的多关系网络。

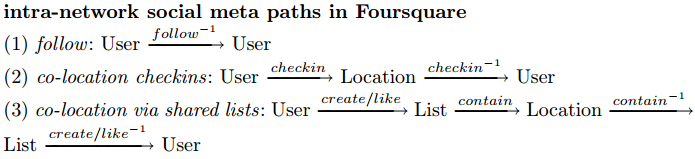
**定义1：网络模式**：对于一个网络G，它的网络模式可以定义为SG=（O,R），O代表结点种类集合，R代表链接集合。

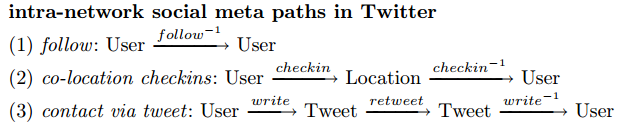
对于图2-A所示的部分对齐的输入网络。我们注意到，这两个网络的网络模式不同，因此不能像在同构情况下那样简单地将异构网络合并在一起

**定义2：Intra-network Social Meta Path：** ，Oi是属于集合O的。Ri属于R。

**定义3：Inter-network Social Meta Path：**给定两个部分对齐的异构网络G(1)，G(2)它们的网络模式分别为和。是在两个图中的元路径。Ri属于二者与{Anchor}（他是anchor链接种类）的并集，O也属于二者并集。

在Foursquare和Twitter上，用户可以关注其他用户，并在地点签到，形成用户之间的两个网络内影响渠道。同时，(1)在Foursquare中，用户可以创建包含一组位置的列表;(2)在Twitter中，用户可以转发其他用户的tweets，这两种方式将分别在Foursquare和Twitter中形成用户之间的网络内影响通道。本文考虑的网络内社会元路径集合及其物理意义如下:





用户可以通过锚用户形成的锚链接在网络中传播信息。这可以抽象为网络间的社会元路径user--anchor—user。考虑到网络间的元路径，由于两个网络中的非锚用户也可以通过网络内元路径和网络间元路径进行连接，因此所研究的问题变得更加复杂，因此，社会元路径实例的数量急剧增长。

每个元路径定义了链接用户之间的影响传播通道。如果链接用户u、v只通过网络内元路径连接，则表示u与v之间存在网络内关系，反之则表示它们之间存在网络间关系。基于这些关系，我们可以为对齐的异构网络构建多对齐的多关系网络。多对齐多关系网络的形式化定义如下

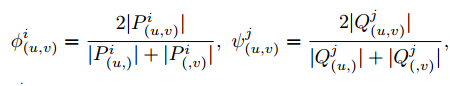
**定义4：多对齐多关系网络**：给定两个部分对齐的网络G(1)，G(2)，我们给出它的定义为M=(U,E,R)，其中U是U(1)，U(2)的并集，代表在网络中所有节点的集合。E是网络中所有节点之间链接的集合，它的元素e可以用（u，v，r）表示，代表了（u，v）之间有一个链接关系r。R是结点种类集合。

**3.2 在MMN中的影响传播**

在本节中，我们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。

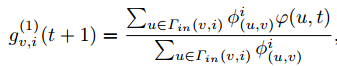
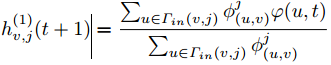
在传统LT模型中，给一个同构网络G=（V，E），一个节点uk可以被激活，是根据uk与它的入邻居之间的权重之和是否大于激活它所需要的阈值来定义的。

基于MMNs M = (U, E, R)，通过pathsim[18]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:



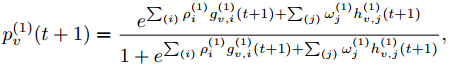
其中是为网内(网间)扩散元路径实例集，根据关系i(j)从u开始到v结束。表示以用户u、v分别作为起始用户和结束用户的元路径实例数。

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

**4.IM problem in M&M Model**

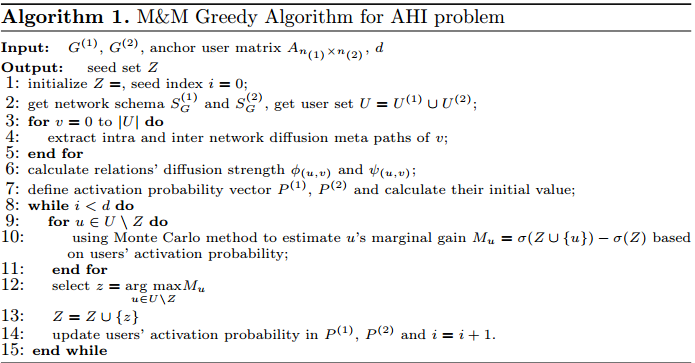
**4.1 IM问题的分析**

我们将表明,M&M模型下的影响力最大化问题也是np-hard并证明影响扩散函数σ(Z)是单调的和子模的。

**定理1**：AHI中的影响力最大化问题是NP-hard。证明如下：AHI问题可以很容易地映射到np-complete的“顶点覆盖”问题。因此AHI问题是NP-hard。

**定理2**：对于MM model，影响函数是单调的

**定理3**：对于MM model，影响函数是子模的

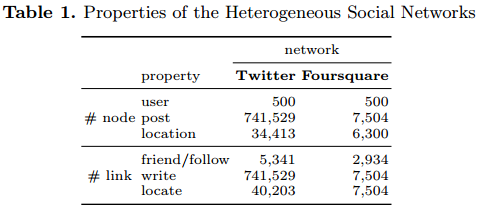


**4.2对于AHI的贪心算法**

**5 实验**

**5.1实验准备**

数据集：实验中使用的部分对齐异构网络数据集为Foursquare和Twitter，这两个数据集的统计数据如表1所示。有关数据集及其爬行方法的更详细信息，请参阅[13]



对比方案：

1. 本文提出了一种从提取的MMNs中贪婪地选择种子用户的方法。根据选择种子用户的网络，比较了M&M的不同变体:(1)M&M(从Foursquare和Foursquare中选择种子用户)(2) M&M-Foursquare(只从Foursquare中选择)和(3)M&M-Twitter(仅从Twitter中选择)。

2.LCI：方法LCI是针对[16]中复杂网络提出的影响最大化方法，从合并网络中选择种子用户。

3. Greedy method for single heterogenous network：贪心方法基于一个多关系网络(如图2-D所示)，选择能够在单个网络中获得最大影响增益的种子用户。与M&M类似，它也有两种变体:greedy-foursquare和greedy-twitter。

4. Seed Selection method based on traditional LT model：基于一个单一的同构网络(如图2-C)， LT选择能够带来最大影响增益的种子用户。实验比较了LT方法的两种变体，LT- foursquare和LT- twitter。

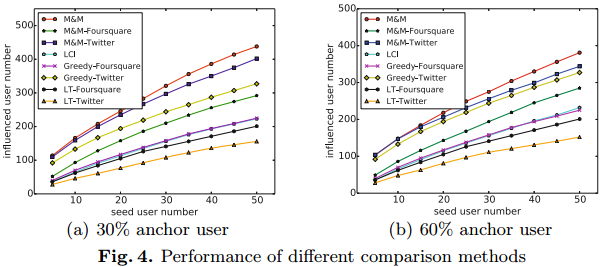
**5.2 实验建立**

输入对齐的异构网络，基于一组网络内部和网络之间的社会元路径提取MMN。利用各关系中用户之间的影响评分，用logistic函数计算聚合的激活概率。为简单起见，将所有关系(包括网络内部和网络之间)的权重设置为相等（0.25）。在 [0,1]内随机选取用户阈值。从{5,10，···，50}中选择种子用户数量。为了模拟不同的部分对齐网络，我们从不同锚点比率的网络中随机抽取锚点链接样本:{0.3,0.6}，其中0.3表示保留了30%的锚点链接，其余70%被删除。

为了评价各种比较方法的性能，实验中以种子用户最终激活用户数作为评价指标，其中锚用户最多一次

**5.3实验结果**

实验结果如图4所示

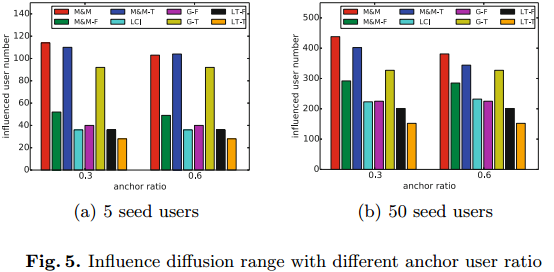


可以看出MM是最好的。

通过与M&M- foursquare和M&M- twitter的比较，我们发现在图4(a)- 4(b)中，M&M都比M&M- foursquare和M&M- twitter表现得更好。它表明了全局选择种子用户(即在两个网络中选择种子)与局部选择种子用户的方法相比，该方法可以取得更好的效果。

**5.4 参数分析**

为了研究锚定比参数的影响，我们将这些比较方法在锚定比0.3和0.6时的性能进行比较，结果如图5所示，其中图5(a)-5(b)对应种子用户集大小分别是5和50



可以看出来MM模型在锚用户比例低的情况下影响的更多。但是，随着锚点比率的降低，更多的用户将是非锚点用户。在计算的时候锚点用户只能算作一次，所以非锚点用户的增多也会导致结果的上升。