**Discover Tipping Users For Cross Network Influence**

**1.introduction**

在线社交网络最大的优势之一是，当一个创意开始流行时，它可以以相当低的成本推动一个品牌迅速获得名声和财富。传统的营销主要是为产品在一个单一网络中选择种子用户。

然而，在现实生活中，在一些在线社交网络上应用病毒式营销是很困难的，这些社交网络在用户选择亲密朋友的过程中创造了亲密和私密的交流圈子。

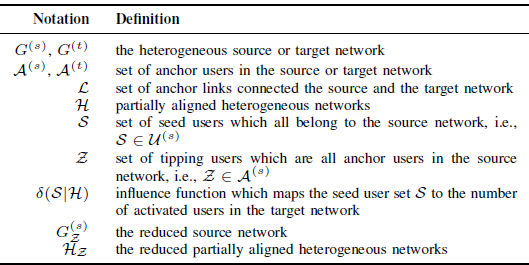
与此同时，如今的用户通常同时参与多个在线社交网络，那些加入Facebook的用户也在使用其他网络，比如Twitter，Foursquare和Instagram。这些社交网络的信息更加公开，用户更容易接触到，为广告公司与受众沟通提供了良好的渠道。并且这些网络提供了将图片分享到facebook上的操作，通过这种的方式，信息可以从这些公共社交网络传播到我们的目标Facebook网络。换句话说，病毒式营销实际上可以在这些公共社交网络中进行。，这些信息可以间接地扩散到目标网络并激活用户。为了区别于目标社交网络(如Facebook)，本文将这些易于访问的社交网络称为源网络（source network）

**2.problem formulation**

**A basic terms**

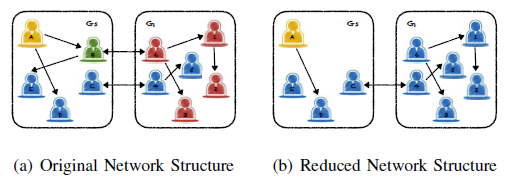
**定义1**：部分对齐异构网络：H=(G(s),G(t),L)，其中G(s)是源网络，G(t)是目标网络。L代表通过锚链接连接的锚用户。

在跨网络的信息扩散过程中，信息的传播是以一群源网络中的种子用户开始的离散步骤，一直到传到目标网络G(t) 网络中的用户有两种状态:活动状态和非活动状态，其中活动状态表示用户已经采用了某种产品。



**B Tipping Users and Influence Gain**

定义2：reduced network：给一个异构网络和一组锚用户Z，减少网络为：也就是把锚用户删掉再把与锚用户相连的边删掉



**定义3：**Reduced Partially Aligned Networks：给一对部分对其的异构网络H=(G(s),G(t),L)，一个用户集合Z，缩减的部分对其网络为，是缩减网络，LZ是与Z中用户相连的锚用户的集合。

**定义4**：influence function：给一对部分对其的异构网络H=(G(s),G(t),L)和一个种子用户集合，用表示影响函数，它表示从源网络中选的种子可以在目标网络中影响到的用户数

**定义5**：Cross Network Influence Gain：给一对部分对其的异构网络H=(G(s),G(t),L)，一个预先确定的种子用户集S，一组锚用户Z，是在G(t)中可以影响到的用户数目

定义用户集Z的跨网络影响增益为二者的差。

**C 问题定义**

**定义6**：Tipping users：给一对部分对其的异构网络H=(G(s),G(t),L)，一个种子用户集S，tipping users就是一组锚用户，它可以使得在源网络中的跨网络影响增益最大。

**定义7**：The TURN（Tipping Users for cRoss Network influencing） Problem：给一对部分对其的异构网络H=(G(s),G(t),L)，一个种子用户集S，一个预先定义好的数k，TURN问题的目标是在源网络中找到k个Tipping users，可以使得cross-network influence gain最大。



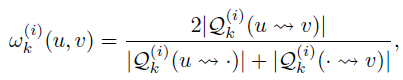
3. information diffusion model

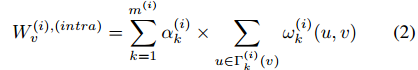
**定义8**：Network Schema：给定一个异构网络G=(V,E)，网络模式定义为S=(O,R),O代表实体的种类集合，R代表链接的种类集合。

**定义9：**meta path：一个元路径，基于网络模式，表示为：

在元路径中，开始和结束实体的类型都是User，其他实体的类型都不是User。元路径的实例称为扩散链接，它由两个特定的用户开始和结束。很明显，不同网络的网络模式、元路径和扩散链接是不同的。因此，需要分别考虑网络内部和跨网络的信息传播，将元路径分为网络内元路径和网络间元路径两类。

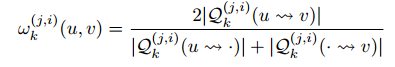
**定义10**：Intra-network and Inter-network Meta Path:给定部分对齐的异构网络H=(G(s),G(t),L)还有各自的网络模式。如果一条元路径的头尾都属于一个网络就叫做网络内元路径，如果一条元路径的头尾不属于一个网络就叫做网络内元路径。

除了上述链接之外，异构的在线社交网络通常还有各种各样的网络内元路径。用m(i)表示网络内元路径数目，表示在网络G(i)中的第k个路径，u，v之间在同一个网络内的链接可以表示为：，作者将定义为大量的信息从u传播到v通过元路径，它可以被计算为：

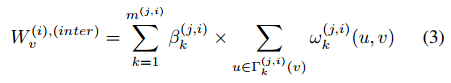
对于一个特殊的用户v，来自所有网络内扩散链路的信息之和表示为，定义如下：

其中是v的在网络内链接的邻居集合

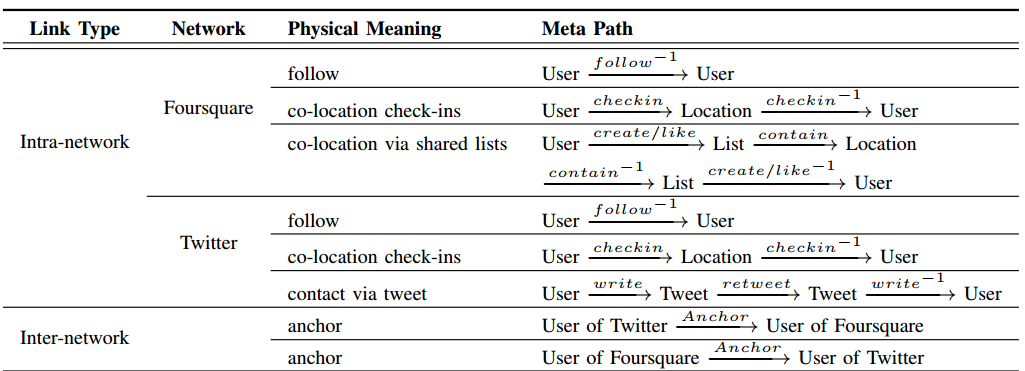
对于网络间元路径，与网络内元路径相似，作者使用表示网络G(j)到G(i)的第k条元路径，网络间元路径的数量表示为m(j,i)。因此节点v从u处接受到的信息可以表示为

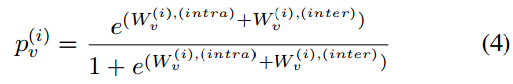


因此，v通过网络间扩散链路从其他网络传播的信息总量为:



作者将Foursquare和Twitter作为部分对齐的异构网络的例子。

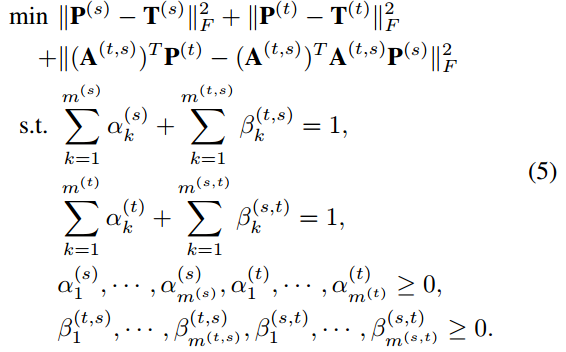


最后，我们可以定义激活函数和逻辑函数该函数将接收到的信息量映射到用户v的激活概率，计算为: 

**4.diffusion links weighting**

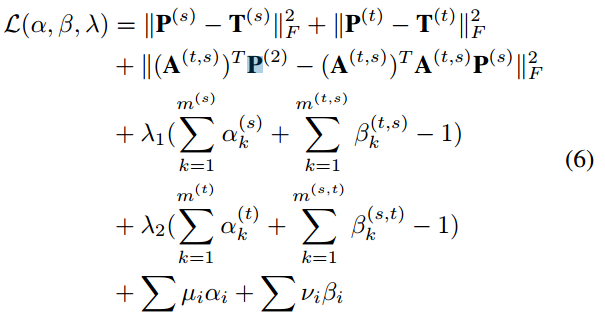
在ConForm模型中，每个权重衡量扩散过程中对应环节的重要性，根据重要性对不同环节进行排序，选择最上面的环节，增加个体的激活概率。对于网络G(t)，作者创建一个列向量其中 记录了用户v的激活概率。同时，从日志数据中提取的groundtruth用一个二进制向量表示，代表用户v最后被激活，否则的话学习权重的目的是缩小预测和基本事实之间的差距。。

另一个问题是锚用户的行为应该是一致的。虽然我们将一个锚用户视为各自网络中的两个独立用户，但从直觉上看，在现实生活中，一个人对同一个话题表现出一致的兴趣，因此，如果G(s)中的锚用户v(s)是激活的，那么在其他网络中v(s)的概率就很高，比如G(t)。因此锚用户的值应该很接近。限制条件是连接到特定网络G(t)或G(s)的权值之和应该等于1，所有的权值都应该是无负的，因此，通过求解以下目标函数，可以得到不同扩散环节传递的信息的最优权值。



代表在两个网路中的锚用户，可以根据锚链接集合L构建。代表v是个锚用户。

由三角形不等式和正齐性可知，每个范数都是凸函数。由于约束存在不等式，我们需要将拉格朗日乘子方法推广到Karush-Kuhn-Tucker (KKT)条件下，得到目标函数的局部和全局极小值:



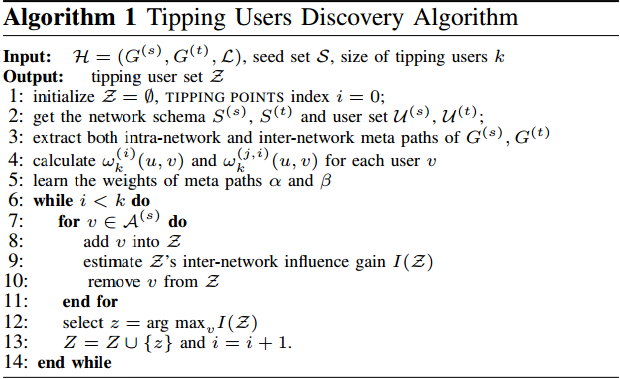
将能得出一组关于变量的方程，它们是隐函数，可以用源代码工具包求解，如Scipy非线性求解器。给定具有初值的参数，通过求解目标函数可以得到多个解。

**5.proposed algorithm**

**A problem analysis**

定理1：基于ConFROM的Turn问题是NP难的

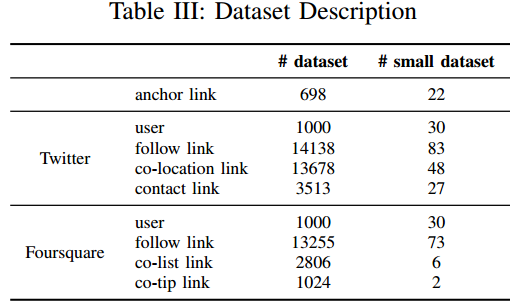
**B proposed algorithm**



**6.实验**

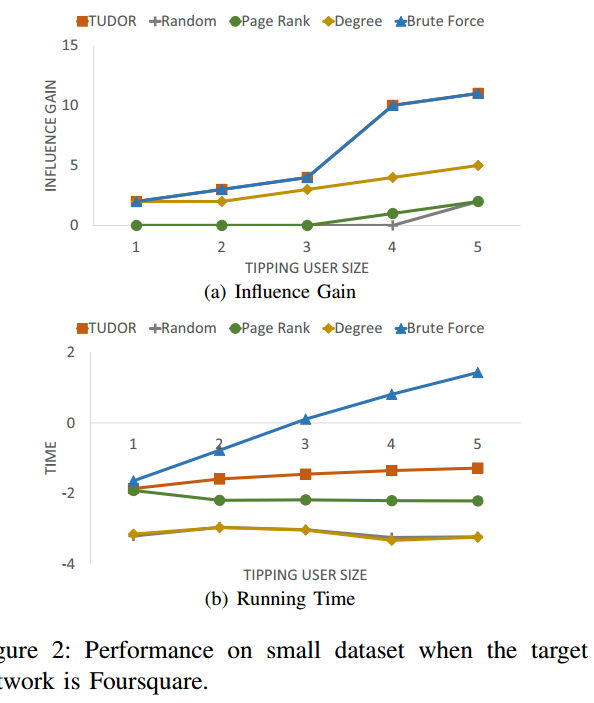
**A 小数据集实验**

我们从一个小数据集中取样，Foursquare和Twitter都是著名的异构在线社交网络，Foursquare用户可以将账户与他们的Twitter账户链接起来，Twitter账户显示在他们的个人资料中。基于这些锚点链接，我们对两个网络中的用户信息和链接进行了抓取，在这个小数据集中，我们总共抽样了60个用户，每个网络30个，其中22个是锚用户，在twitter中有83个社会链接，48个与地点相关的链接，27个接触链接。

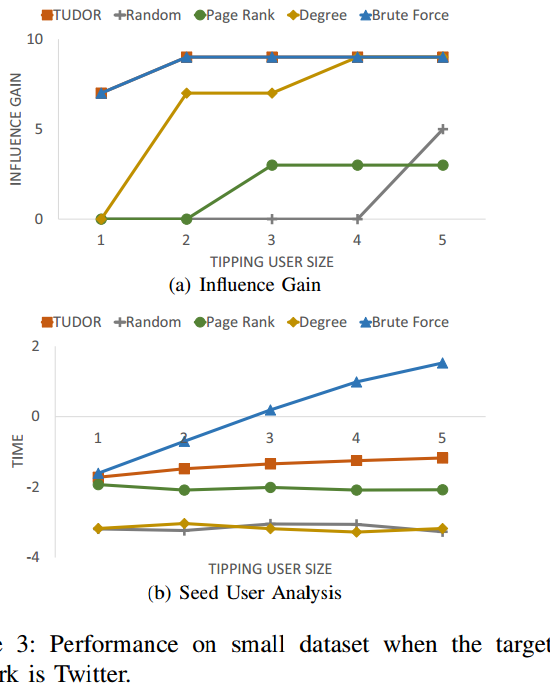


对比实验：TUDOR，Random，PageRank，Out-Degree，Brute Force

Result：当目标网络为Foursquare时，不同对比方法的实验结果如图2所示。



当Twitter作为目标网络时，结果如图3所示：



B原始数据集的实验

数据集：这个数据集也是从Foursquare和推特中获取的, Foursquare和Twitter都有1000个用户，其中689个用户在两个网络上都有账户。Twitter网络中有4138个社交链接、13678个共地点链接和3513个联系人链接，作者从Foursquare抓取了13255个社交链接、2806个共列表链接和1024个连接用户的共提示链接。

Foursquare的影响增益结果如图4(a)所示。twitter的影响增益结果如图4(b)所示。

