**Influence maximization in multiple online social networks**

**1.introduction**

贡献：

**2.model and problem definition**

**A graph notations**

作者使用G1,G2,…,Gk来表示k个网络，每个网络，其中W为权重集合，每个用户的阈值集合为cita。在k个网络中，存在着积极参与多个网络的重叠用户。我们可以确定两个网络中的两个顶点是否是同一个人的帐户（用的参考文献17）。如果两个顶点ui∈V i和uj∈V j表示相同的用户u，则称为彼此重叠的顶点。网络Gi和Gj的重叠由重叠映射Cij来描述，其中如果ui和vj是重叠顶点，则Cij(ui, vj) = 1。最后，用G1…k表示系统由k个网络组成，其中包含一组重叠映射{Cij}，并用U表示全体用户集合

**B influence propagation model**

作者使用LT模型，将其扩展到多个网络。

在多个网络G1…k中，信息在每个网络中分别传播，可以通过重叠的用户从一个网络转移到另一个网络。信息开始从种子用户集合S (i)开始传播。S中的所有用户都处于活动状态，其余用户都处于非活动状态，在t时刻，如果某个网络中来自其活动邻居的总影响超过其阈值，则用户u变为活动用户。在每个时间步骤之后，新的非活动用户将被激活，并继续激活其他用户。此过程将继续，直到不再激活不活动的用户为止。如果我们将传播时间限制为d，那么进程将在t = d时间步长之后停止。时间d后由种子集S引起的活跃用户集合记为

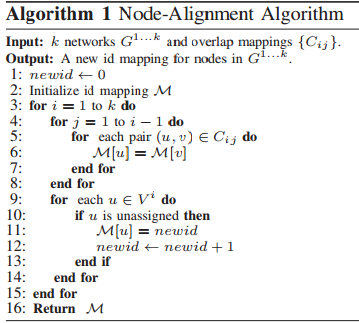
**C problem definition**

定义1：Influence Maximization Problem (IMP)：给定一个网络G1…k，用户集合U，正整数d，b，问题是要找到一个集合S属于U，并且S的大小为b，使得在d跳之后，影响的范围最广。

3.network alighment

我们首先将一个通用标识(id)重新分配给网络中的每个节点，以便同一用户的所有重叠节点都具有相同的id。每个网络拓扑通常使用自己的系统来命名节点，因此同一个人在不同的网络中可能有不同的id。因此，它需要一个复杂的机制和额外的努力，不断更新跨网络的用户状态。我们通过为每个用户分配一个惟一id并将其用作所有网络中的节点id来减轻这种负担。但是，如果我们将新id逐个分配给未分配的节点及其重叠节点，则需要每次扫描所有重叠映射，这在大型网络中几乎是不切实际的，接下来，我们设计了一个算法，它只在线性时间内分配新的id。

我们的目标是扫描每个重叠映射并检查每个节点一次以分配新的id。我们不同时为用户的所有重叠节点分配id，而是检查其中一个重叠节点是否已经分配了id。具体来说，我们分两个阶段处理每个网络:将id分配给其映射列表中的节点和处理过的网络，然后将新id分配给其余节点。算法一描述了这个过程：



节点对齐算法将相同的id分配给同一用户的所有重叠节点，将不同的id分配给不同用户的节点。

**4 coupling scheme**

本节提出一种将多个网络耦合成一个新的单一网络的方案，该方案保留了每个网络的影响扩散特性。因此，我们可以通过在耦合网络中求解来降低问题的复杂度。然而，由于以下原因，它具有挑战性：

1.表示用户在多个网络上的参与。在不同的网络中，每个用户的不同之处在于:(1)他从朋友那里获得的影响力的大小;(2)他对朋友的影响力。我们需要在不破坏单个网络结构的情况下为每个用户维护多个参数。此外，参数的数量因每个用户而异，因为用户可能不会加入相同数量的网络。

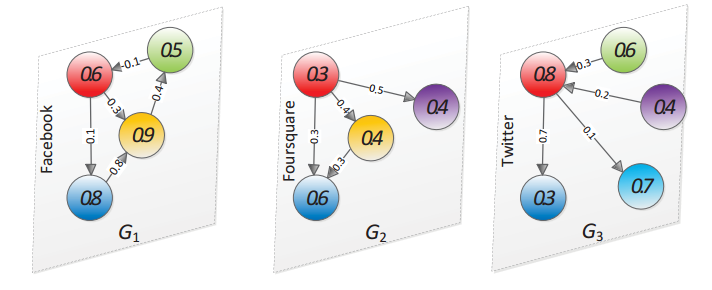
2. 编码用户的多重传播。一旦一个用户受到影响，他倾向于将信息传播到所有的网络。它要求我们在一个网络中编码多个传输，而这个网络不允许有多个边。

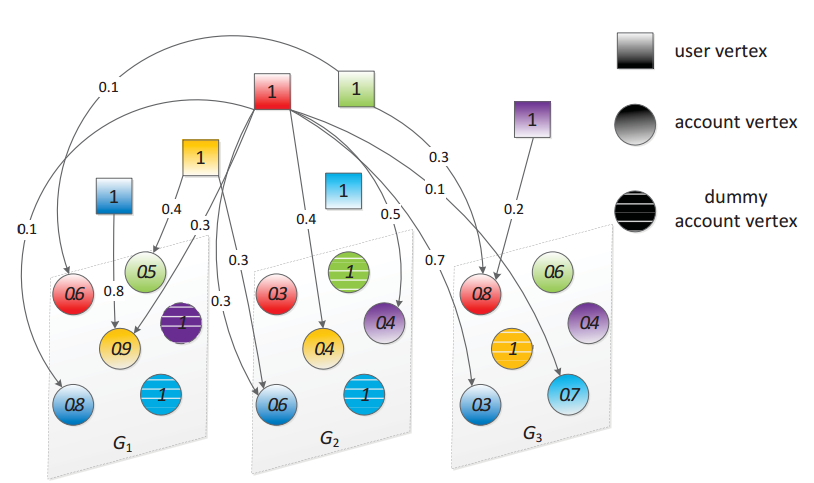
我们为网络中的每个用户使用代表节点，并将传播步骤分解为子步骤来处理多个传播，从而处理这些挑战。对于每个用户有一个全局id up，作者使用up0来表示up的状态。表示 up在不同网络的状态。如果up没有加入网络Gi，那么upi是一个孤立节点。用户up对uq在网络Gi中的影响被表示为up0对uqi的影响。接下来，我们将多个网络中的传播步骤分解为子步骤，以准备所有网络中的信息可用性，并同时进行传播。具体的说，给定k个网络，用户集合，作者会按照如下步骤构建一个新的图。

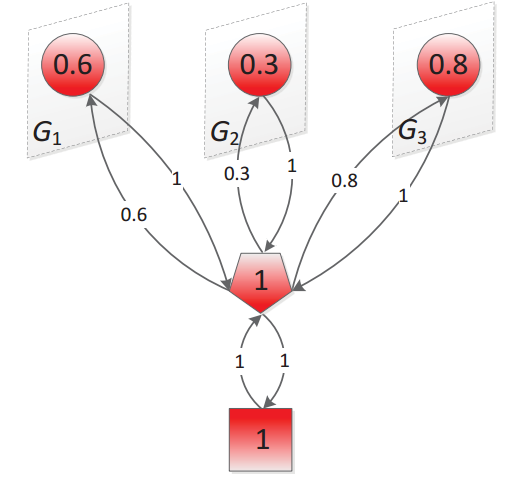
第一，构建节点集合V={upi | 0<=i<=k+1,1<=p<=n}，用户节点up0和account节点upi代表up和她在不同网络中的加入情况。转变节点upk+1以证明耦合网络中传播过程的等价性。关于节点的阈值，，所有其他节点的阈值都为1。

第二，作者用户顶点u0对网络Gi中用户v账户顶点的影响来表示用户u对网络Gi中用户v的影响。如果边，那么边和权重被加入到边集合中。

最后，我们在同一用户的用户顶点、转换顶点和帐户顶点之间添加边，以保证它们具有相同的激活状态，如果其中一个节点是活动的，它将激活所有其他节点。这些同步边的权值等于目标节点的阈值。







接下来，我们给出了耦合网络G中节点的激活状态与原始网络的等效关系，在G1.。。k上用d跳求解IMP等价于用G上的3d跳来求解IMP。

**引理2**：假设在G1.。。k传播过程和在G的传播过程是从一个相同种子集合S中开始的，用户u在d跳中被激活的条件是节点u0在3d跳被激活。

**引理3**：假设耦合网络G上的传播过程只从用户顶点的种子集S开始，那么活动顶点跳数就是k + 2乘以经过3d跳数后的活动用户顶点数（d>0）

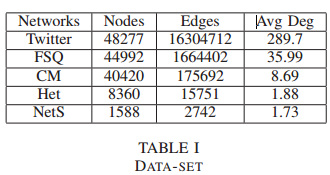
引理2和引理3证明了以下一点：IMP问题在G1.。。k上的d跳问题等于在耦合网络G中的3d跳问题。

定理1：种子集合在网络G1.。。k上经过d跳之后影响A个用户等价于种子集合为在耦合网络中3d跳之后影响了(K+2)A结点。

**5 实验**

在本节中，我们通过对真实网络和合成网络的大量实验，证实了考虑多个网络对识别最具影响力用户的重要性。

Real network：重叠节点数为4100,2860,517，每两个网络中都有90个重叠结点，将边随机在（0,1）上赋权，每个节点的阈值都是一个随机值，取值范围在0到其传入边的总权值之间。

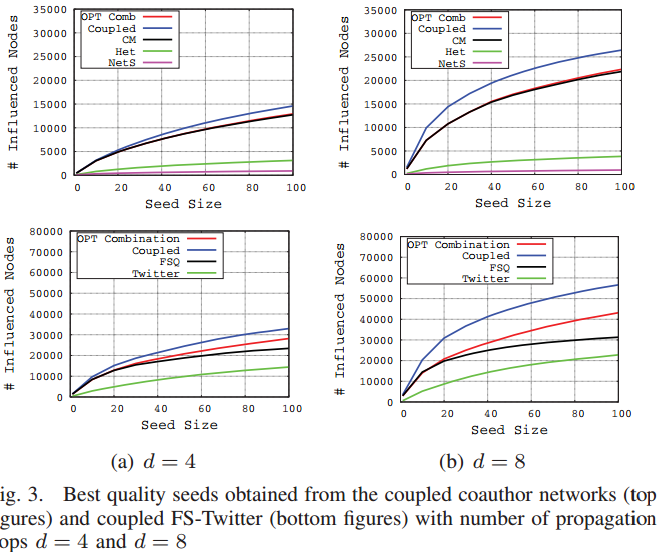


Synthesized networks.两个网络G1和G2分别有5000和10000个节点

**A 耦合网络的优点**

在这一部分中，我们评估了考虑多个网络时种子集的质量以及耦合网络的优点，在没有耦合的情况下，我们可能需要从每个网络中单独选择总种子集的一部分，以便最大限度地增加受影响用户的总数。这种策略提供了一个比只从一个网络中选择所有种子更好的种子集。

然而，从每个不同的网络中获得种子的最优组合(OPT)，不仅比从耦合网络中选择最具影响力的用户需要更多的努力，而且与此同时，它只能计算出一个质量低得多的种子集，如图3所示。



我们也可以使用耦合网络来识别更有影响力的用户，从而只在一个网络中传播信息。如图4所示，考虑到网络与其他网络的关系，我们得到了更好的种子集

**B 种子集和影响集的组成**

如图5所示，耦合网络的种子集包含来自特定网络的节点数量比其他网络多得多。在FSQTwitter中，首先选择FSQ用户，因为他们更容易受到影响。在合著者网络中，我们发现大多数种子节点都是从conmath网络中选择的。在前100名有影响力的用户中，约23%是重叠用户，而只有4.7%和8.5%的用户是FSQ-Twitter和合著者网络中的重叠用户。这意味着如果信息更容易在一个网络中流动，那么这个网络就会吸引和传播更多的信息。重叠节点是高潜在目标，一些网络比其他网络更有效地进行广告。

**C 重叠用户的用户的角色**

接下来，我们研究重叠用户对信息扩散的影响，由于缺乏具有不同重叠用户数量的真实数据集，我们对合成网络进行了实验。图6展示了重叠用户数量对top用户影响的稳定影响。这种影响几乎随着重叠节点的数量线性增加。当重叠节点数量从1000增加到5000时，前100名用户的影响力增加1.8倍。这意味着用户加入多个网络的趋势正在为病毒式营销的前景施肥。这也显示了耦合网络的优势，在未来，当重叠用户的数量可以更高。