Least Cost Influence Maximization Across Multiple Social Networks

**1.introduction**

随着现有OSNs的快速发展，已经有相当数量的用户同时拥有多个帐号，这使得他们可以在不同的网络上传播信息。

在本文中，我们研究了最小成本影响(LCI)问题，它的目的是寻找一组具有最小基数的用户来影响多路网络中一定比例的用户。由于复杂网络的扩散过程比较复杂，很难在单一网络中直接扩展已有的解来进行求解。此外，研究多路网络中的问题也带来了一些新的**挑战**:

1)如何准确评估重叠用户的影响;2)在哪个网络中，用户更容易受到影响;3)哪个网络传播的影响更好。

为了回答这些问题，我们首先引入一个模型表示来说明信息如何通过耦合方案在多路网络中传播，通过将多个网络映射到一个网络中，不同的耦合方案可以保留原网络的部分或全部特性。然后，我们可以利用单一网络上现有的解决方案来解决多路网络中的问题。此外，通过综合实验，我们验证了耦合方案的有效性，并对复杂网络中的影响传播过程提供了一些有趣的见解。作者的**贡献**如下：

1. 提出了一种基于多种耦合方案的模型表示方法，将复杂网络中的问题简化为单个网络上的等价问题。所提出的耦合方案可适用于最常用的扩散模型，包括线性阈值模型、随机阈值模型和独立级联模型

2. 提出了一种可扩展的贪婪算法来解决LCI问题。特别是，改进因子随着网络的大小而增加，这使得该算法可以在具有数百万个节点的非常大的网络上运行

3. 我们对真实数据集和合成数据集进行了广泛的实验。结果表明，考虑多网络而不是单一网络可以有效地选择最具影响力的用户

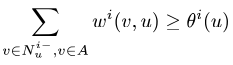
**2. MODEL AND PROBLEM DEFINITION**

**A graph notation**

作者假设有k个网络，G1，G2…,Gk。每个都是一个带权有向图，结点集合表示在网络中有ni个用户，边集合代表图中有mi条边。表示边上的权重。此外，每个用户都与一个阈值相关联，该阈值指示其观点的持久性。阈值越高，u越不会被影响。此外，积极参与多个网络的用户称为重叠用户，可以使用文献[10]中的方法进行识别。这些用户被认为是跨网络传播信息的桥梁用户

**B 影响传播模型**

作者使用LT模型并且扩展到多个网络中。在经典的LT模型中，每个节点可以是活动的，也可以是不活动的:如果它被选择到种子集中，则处于活动状态，或者来自In -degree邻居的总影响超过其阈值。在多路网络系统中，给定一定数量的网络，信息在每个网络中单独传播，只能通过重叠的用户流向其他网络。信息开始从一组种子用户开始传播。此时，如果某个网络中来自其活动邻居的总影响超过其阈值，则用户变为活动用户，即，存在这样的情况：

，A是活跃用户集。在每个时间步骤中，一些不活跃的用户被激活，并试图在下一个时间步骤中影响其他用户，直到无法激活不活动的用户时，进程将终止。如果我们将传播时间限制为d，那么在执行时间步骤t=d之后，进程将停止。在时间d时活跃的用户集合可以被表示为：

**C problem definiion**

在这篇论文中，作者解决了病毒式营销在多重网络中的根本问题:LCI问题。该问题要求找到影响大部分用户的最小基数的种子集。

定义1：LCI问题：给定k个网络G1…k，还有用户集合U，一个正整数d，一个数beta，LCI问题要求找到一个最小基数的种子集，以便根据LT模型进行激活后的活动用户数量至少是用户的beta部分，即：



**3. LOSSLESS COUPLING SCHEMES**

在此，作者针对每个参与网络的影响扩散过程，提出了将多个网络耦合成一个新的单一网络的无损方案。这个新耦合图的一个显著优点是，我们可以在一个网络上使用任何现有的算法来生成具有相同质量的多路网络的解。

**A Clique Lossless Coupling Scheme**

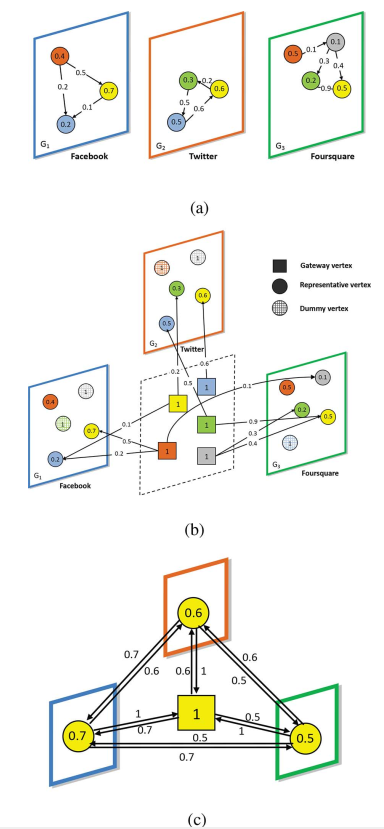
在LT模型中，第一个问题是通过为网络中的每个用户引入它不属于的虚拟节点来解决的。这些虚拟节点是孤立的。现在，在第i个网络中的节点集合可以表示为，

是所有用户的集合。是up在网络Gi中的代表结点。在这种表示下，节点之间有一条边相连如果up和uq在Gi中是相连的。现在我们可以联合所有的网络来形成一个新的网络。克服第二个挑战的方法是允许节点u的代表结点可以相互影响，通过加边和权重来实现。当ui被影响了，uj在下一个时间步骤中也会被影响，因为他们是重叠结点。因此，信息从网络Gi传输到Gj。然而，一个出现的问题是，当信息在两个网络之间传输时被延迟了。ui一旦被激活，就会影响到它的邻居，而uj在它开始影响邻居之前还需要一段时间。如果两者同时开始影响他们的邻居，那就更好了。为了实现这个，加入网关顶点u0到G中，ui和uj只能通过u0来影响其他结点。特别的，所有边会被替换成。此外，更多的边的加入使得u0，ui，uj互相影响，由于网关与同一用户的代表顶点之间的连接形成了一个小圈子，所以作者称这个为

Clique lossless coupling scheme。在形成耦合网络之后，作者对边权值和节点的阈值作如下分配：

**结点阈值**：所有虚顶点和网关顶点的阈值都为1（个人理解为没有连边的节点）。剩下的代表节点都与up有相同的阈值

边权值：如果在图Gi中，用户u和v之间有一条边，那么边的权值为，将网关结点与同一用户的代表顶点之间的边赋值为，图1给出了一个简单的无损耗耦合方案示例。



接下来，我们将证明原多路网络和耦合网络的传播过程实际上是相同的。影响在网关和代表性顶点之间交替传播，因此在多路网络中有d跳的问题等价于在耦合网络中有2d跳的问题。

**引理1**：假设耦合网络中的传播过程从种子集开始，种子用户集中只包含网关结点然后只在偶数跳时激活代表顶点。

引理2：假设在G1…k和图G传播过程中的初始种子集合是同样的，下面的条件是等价的

（1）用户u在G1…k中d跳后被激活

（2）存在一个i使得在图G中ui在2d-1跳之后被激活

（3）节点u0在图G中经过2d跳之后被激活

**定理1**：给定一个k个网络的系统G1…k和一个用户集合U，由无损耦合方案产生的耦合网络G，和一个种子集合，如果 是被种子集合S在多路网络中经过d跳后所激活的节点集合，那么是被S经过2d跳传播之后在耦合网络中所激活的节点集合。

定理1为从单网络上的解导出多路网络中LCI的解提供了基础。

**定理2**：当使用无损方案时，结合S