# Social Network Sites: Definition, History,and Scholarship

## 社交网站的定义

1. 大多数社交网站支持原本已经存在的原始社交网络，一些则帮助具有共同兴趣、政治观点的陌生人建立连接。一些SNS面向多元化的用户，而其他的则针对特定的种族、语言、宗教等用户。网站也在不同程度上加入了其他性质的通信和信息交流方式，如通过分享图片音频等。
2. 将社交网站定义为：允许个人 1) 在限制的系统内创建开放的或半开放的主页

2）列出与他们有连接的其他用户列表

3）查看和访问与他们有连接的其他用户的连接列表，这种连接的方法和命名会随着网站不同而不同。

1. 社交网站独特的地方不在于说是能够让用户认识陌生人，而是在于能够使他们的社交网络能够表现出来。
2. 许多大型的社交网站的使用者并不是真正的在“社交”或者想认识陌生人，相反的，他们大多数时候知识为了跟已经处于他们原本社交网络的人联系。
3. 社交网站的个人主页由创建账户时用户所填写的信息生成，有的社交网络允许用户添加不同的媒体介质来定制化个人主页。关于个人主页是否公开也随着网站不同而不同。
4. 在加入社交网站后，用户会被鼓励与其他用户建立联系，不同的网站设置不同的标签来表明这种联系，如朋友、粉丝、关注者等。
5. 公开展示自己的社交网络是SNS关键的组成环节。
6. 大多数SNS都提供评论或私信功能。
7. 除了个人主页、朋友圈、评论、私信等功能，不同的网站具有不同的特色，如图片分析（ins）、视频分析（YouTube）。

## 社交网站的历史

1. 最先具有个人主页、朋友列表的社交网站是SixDegrees.com，发布于1997年，1998年加入了访问朋友列表的功能。
2. 从1997-2001许多社区都开始支持个人主页与公开朋友列表的功能。
3. Ryze.com发布于2001年帮助用户利用他们的商业关系网。

# 社交网络影响力传播研究

陈卫

## 一、引言

1. 影响力传播的三大支柱:
2. 影响力传播模型：主要描述影响力在社交网络中如何传播、有何特点和性质
3. 影响力传播的学习：如何利用网络大数据挖掘学习影响力传播模式和具体传播模型的参数
4. 影响力传播优化：着重考虑在不同传播模型下，如何通过施加外部作用力来扩大希望传播的影响力或者控制和减弱不希望传播的影响力，也包括有效地监控影响力的传播。



1. 将社交网络用有向图G =（V,E）来表示，对每一个节点即位图中的人。针对某一具体传播实体，将图中的每个点描述为两种可能状态：不活跃(inactive)和活跃(active).不活跃表示该节点还未接受传播实体，活跃表示已经接受传播实体，节点从不活跃转化为活跃，称之为被激活。

## 二、 影响力传播模型

1. 影响力传播模型大多数以随机模型(stochastic models)来描述，也有用博弈论模型(game-theoretic models)来描述的。随机模型更直接地反映了影响力传播中的不确定性，也是当前研究的主流。
2. 随机模型分为离散时间模型和连续时间模型、递进型(progressive)和非递进型(non-progressive)模型等。
3. 离散时间模型：便于分析和计算
4. 递进型：假设任意节点一旦从不活跃变为活跃就会一直保持在活跃状态。多用于信息、产品等的传播/
5. 非递进模型：允许节点在不同状态之间来回切换。这类模型多用于描述观点、看法的传播，因为人的观点和看法常会改变。
6. 经典离散时间递进性传播模型

包括独立级联(independent cascade)和线性阈值（linear threshold）等离散时间递进性传播模型。

1. 独立级联模型：

独立级联模型将图中有向边的权重值p(u, v)定义为当节点u为活跃状态时，节点u通过边（u，v）向节点v成功传播的概率值。对于离散模型，信息的传播均在离散点完成，独立级联模型的传播过程在离散时间点如以下形式完成：

在t=0时刻，一个预先存在的初始集首先被激活，其他节点均处于不活跃状态。

在t≥1的时刻，表示到时刻t为止所有激活的节点集合，所有上一时刻激活的节点()都会对它的每个尚未激活的出邻居以概率p(u,v)尝试激活一次，所有激活事件之间相互独立。当没有新的节点被激活时，传播过程结束。

注意：每个节点只做一次激活尝试。

使用来表示传播结束时的激活节点的集合，是一个随机集合，影响力传播中经常关心的时传播结束后被激活节点个数的期望值E[||],用表示，称为影响力延展度（influence spread）。

1. 线性阈值模型：

线性阈值模型将节点的激活条件定义为该节点v所有处于激活状态的入邻居的指向节点v的边上的权重之和,同时每一个节点都有一个阈值，当该节点的权重之和大于该阈值时，则该节点被激活。对于一个节点所有入边上的权重之和要求.

1. 其他模型
2. 传染病模型(epidemic model)
3. 选举模型(voter model)
4. 博弈论模型(game-theoretic model)
5. 多实体传播模型(multi-item diffusion model)

## 三、影响力最大化(influence maximization)问题

1. 影响力最大化时在给定社交网络结构G=(V, E)、影响力传播模型及其参数的情况下，选择k个节点作为种子节点集合,使得以为种子节点长生的影响力延展度最大，即.
2. 子模函数(submodular function)和影响力最大化的贪心算法技术

子模性是集合函数的一种性质。

对于一个集合函数： ,f(x)定义域为集合Ω的任意子集，值域为实数集，对f(x)满足：对于任何两个Ω的子集X,Y，且X是Y的真子集，则对于任何一个，下式成立：

即为具有子模性。

对于具有单调子模性的集合函数的贪心算法大致如下所示：

算法：单调子模函数的贪心算法。

输入：单调子模函数f，预算k。

输出：大小为k的子集S。

初始化：S=φ

for i =1 to k do

v=arg maxu∈V\S( f(S∪{u} )-f(S) )

S =S∪{v}

end for

返回S

贪心算法的到的值不会小于,OPT为最优解。

1. 可扩展的影响力最大化算法

对于贪心算法存在NP-hard问题，使用蒙特卡洛方法又存在时间效率很低的问题。为解决这个效率问题提出的可扩展的影响力最大化(scalable influence maximization)算法基本分为：启发式算法和改进蒙特卡洛算法的贪心近似算法。

1. 其他基于影响力的优化问题
2. 种子集合最小化
3. 利润最大化：要考虑选取种子成本与收益的均衡
4. 影响力传播监控
5. 多实体传播模型下的影响力最大化
6. 网络拓扑的优化
7. 非子模性的影响力优化问题

## 四、社会影响力传播学习

1. 影响力传播学习的基本思想

这类工作基于两类数据：

1. 社交网络结构的数据
2. 用户某一类行为的时间序列

需要解决的问题是：在一个节点执行一个动作之前，有多个该节点的邻居节点都执行了同样动作， 在这种情况下如何判定是哪一个或哪几个邻居节点真正影响了该节点？现有的方法基本分两种

1) 最大似然估计

2) 信用分配和频度分析

# 社交网络分析核心科学问题、研究现状及未来展望

## 一、社交网络分析研究的背景

* + - 1. 根据欧盟关于社会计算的研究报告，在线社交网络可分为4类

1. 即时消息类应用，即一种提供在线实时通信的平台。
2. 在线社交类应用，即提供一种在线社交关系的平台
3. 微博类应用，即一种提供双向发布消息的平台。
4. 共享空间等其他类应用，即其他可以相互沟通但结合不紧密的web2.0应用
   * + 1. 蓬勃发展的基于互联网的社交网络对国家安全和社会发展产生这深远的影响。

## 二、社交网络分析研究的核心问题