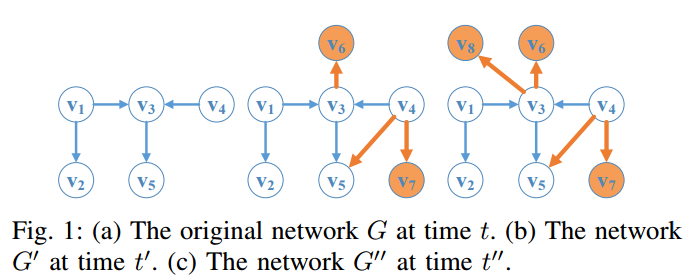
**Time is What Prevents Everything from Happening at Once: Propagation Time-conscious Influence Maximization**

**1. introduction**

**A 传统IM问题的限制**

经典IM问题及其解基于以下隐式假设。假设影响从种子集S传播到社交网络G中的最大节点需要t时间，然后，假设t很小，使G的拓扑在t时刻保持不变。因此，G的拓扑结构在传播过程中是完全已知的。综上所述，上述研究表明，影响传播可能需要相当长的时间才能达到社交网络发展的最大范围(几周)。换句话说，在t时刻，t’时刻的网络完整拓扑结构是未知的。座椅一开始会选v1作为种子节点，但是在t’时刻会选v4，因为网络中新加入了节点。



**B Can Recent IM Efforts on Dynamic Networks Address the Limitations?**

最近，人们在动态或带时间的社交网络的背景下对IM问题进行了研究，乍一看似乎可以解决前面提到的问题。不幸的是，事实并非如此，因为这些技术要么假设网络拓扑在特定的时间点是完全已知的，要么忽略了影响传播时间对网络状态的影响。一般来说，这些技术在不同的时间点重复运行经典的IM算法(或它们的增量版本)，以便找到最新的种子集。然而，无论时间点是如何分割的，也无论IM算法的选择如何，该策略都无法解决上述经典IM问题的局限性。

C 贡献

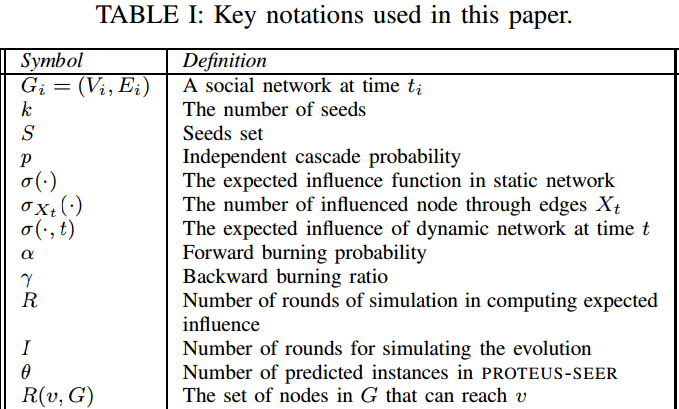
本文有四个贡献。

首先，我们从理论上证明，如果经典IM问题抛弃了上述与社会网络进化性质有关的假设，以及影响传播时间对种子选择的影响，那么贪婪算法会近似到。

第二，作者将经典IM问题重新定义为PROTEUS-IM（Propagation Time-conscious Influence Maximization)。给定一个网络在t0时是G0=(V0，E0)，在时间tr时，网络演进为Gr=(Vr,Er)，PROTEUS-IM的目标是在t0时刻，在G0中找到一个种子集合S属于V0，使得在Gr中影响最大。

第三，我们提出了一个贪心算法PROTEUSGENIE来解决PROTEUS-IM问题

第四，作者将PROTEUSGENIE算法应用于实际网络中



**2. classical influence maximization problem**

**3. propagation time-conscious influence maximization**

**A terminology**

作者将社交网络抽象为一个有向图G=(V,E)，V中的节点表示网络中的个体，E中的边表示它们之间的关系。G的阶是|V|，大小是|E|。传统的IM假设影响根据特定的级联模型在节点之间传播，并选择V中的k个节点作为种子结点来传播一条信息，使信息传播到其他节点的数量最大。然而，这种影响的传播在现实中可能需要tr时间(可能是几周)。在此期间，社交网络可能从t0时刻的G0 = (V0, E0)演化为tr时刻的Gr = (Vr, Er)。我们称G0为当前网络，称Gr为目标网络。相应地，t0和tr分别称为当前时间和目标时间。为了通用性，我们假设tr是由用户给出的，因为它依赖于应用程序和网络。我们假设|Vr| > |V0|和|Er| > |E0|，大多数现实社会网络随着时间的推移而增长。并且点的交集和边的交集不能是空。作者用表示在时刻t的影响期望。为了便于说明，作者假设传播符合独立级联(IC)模型，其中影响按照独立概率pij沿任意边vivj传播，因为它是文献中最流行的模型之一。

**B redefining IM problem**

经典的影响最大化问题(定义1)忽略了影响传播的时间，而影响传播的时间在现实中可能是非常重要的，在此期间，潜在的社交网络可能会进化。因此，我们正式地将这个经典的影响最大化问题重新定义如下。

定义2：Propagation Time-conscious Influence Maximization Problem

设G0 = (V0, E0)为t0时刻的当前网络，k为预算。Tr为影响传播时间，并且此时的G0演化到了Gr=（Vr,Er）。Propagation Time-conscious Influence Maximization的目标是在t0时刻选择一个种子集合S属于V0，使得在时刻tr最大。并且假设Gr的拓扑在t0时刻是不被知道的。

注意，根据上面的定义，种子是从网络G0的当前实例中选择的，而不是从网络的未来实例中选择的，这是因为在t0中很难知道将来哪些用户可能会加入或离开社交网络，他们将如何与其他用户连接，以及他们是否会成为种子的一部分。事实上，如前所述，准确预测未来社会网络具有的完整拓扑是不现实的，所以只能选择在t0时刻的节点作为种子。

影响函数是子模的。PROTEUS-IM是np难的。

**4. A greedy solution**

在本节中作者提出了一个贪心算法叫做PROTEUS-GENIE。请注意，设计这样的算法是有挑战性的。一方面，在t0时刻预测目标网络Gr的拓扑结构是不现实的。另一方面，在不了解Gr拓扑结构的情况下，利用现有的级联模型计算Gr的传播期望是非常困难的。

我们利用一个流行的网络演化模型——森林火灾（forest fire model）模型，从G0预测Gr的期望拓扑结构，从而解决了这一难题. 因此，我们利用Gr的这种预测拓扑结构，使用现有的级联模型来确定它的预期扩展。

**A forest fire model**

大多数社交网络本质上是进化的，表现出一系列的性质和现象，包括收缩直径、致密化幂律。文献中提出了几种网络演化模型[14]、[23]、[27]、[28]来模拟现实世界在线社交网络的演化。在这些模型中，我们选择了森林火灾模型(FFM)[23]，因为它优于其他模型[27]。该模型的正式定义如下。

定义3：假设Gt为时间t的网络，G1只包含第一个节点。在时间t给定输入节点v，网络Gt-1在时刻t-1可以被升级为Gt，只要遵循以下规则。

1：从Gt - 1中均匀选择一个ambassador节点w，建立v到w的有向边vw

2：从一对均值为 的二项分布中选取两个数字x和y作为样本。v均匀地选择x个入链和y个出链与w相连，让w1 w2。， wx+y为所选链接的另一端。特别是,α是一个预设燃烧的概率，γ是预设的反向燃烧比，γα是向后燃烧的概率。

3：建立从v到w1、w2…wx的有向边，并且建立从wx+1…wx+y到v的有向边。然后，我们递归地对每个w1、w2….wx应用步骤(2)，直到没有要添加的新链接为止。随着这个过程的继续，节点只能访问一次，这样就不存在循环子结构。

在[23]中已经表明，FFM生成的网络满足大多数实际网络的特性，不仅包括静态特性，还包括动态特性。在[27]中也证明了使用该模型可以很好地模拟和预测许多真实网络的演化。因此，我们利用FFM来预测网络在目标时间tr的演化。具体来说，我们的PROTEUS-GENIE算法在影响最大化的过程中，将FFM与节点选择相结合，便于发现优质种子。

**B the proteus-Genie algorithm**

PROTEUS-GENIE算法的目标是利用FFM预测底层网络从t0到tr的拓扑结构，贪婪地选择边际期望影响最大的节点.

从直观上看，在PROTEUS-GENIE中种子的选择是这样的。

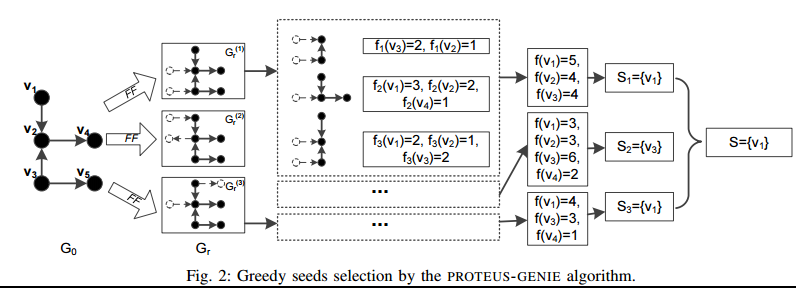
**首先**在t0给定一个网络G0，估算所有v属于V0的在时刻tr，在社交网络Gr中的边际影响，Gr是通过在G0上运行FFM模型来得出的。通过拟合选择正向燃烧概率和反向燃烧比。**其次**，选择期望影响最大的节点作为第一个节点，并将其从Gr中删除。

第三，对前两个步骤执行k轮迭代，选择k个种子作为S1。注意，在前面的步骤中，我们使用FFM生成一个目标网络Gr，这将导致tr时刻的确定性网络。然而，使用FFM进行t0到tr的网络演化是一个随机过程，单轮仿真无法准确描述。因此，前三个步骤分别执行I轮，导致I个不同的Gr实例，表示为因此，种子集S1，…， SI是在I轮之后生成的。最后，将这些种子的秩进行汇总，选择总体秩最高的前k个节点作为最终的种子集S。

算法描述如下：

首先，模拟了网络G0向Gr的演化过程，然后初始化一个种子集实例，Si为空(第3-4行)。然后，它迭代地将k个种子节点选择到Si中(第5-12行)。对于每个种子节点的选择，我们分别以1 - p的概率删除Gr中的每条边，生成图G ' r，得到生成图G ' r = (Vr '， Er ')，这样，Er’可以看作是tr时刻的活动边集Xr，从中我们可以计算每个v∈V0的边际影响。这个过程重复R次，并聚合每个节点的边际影响(第6-9行)。然后，算法选择到目前为止累积边际影响最大的节点(表示为v∗)，并将其插入Si并从V0′中删除。到目前为止，我们有种子节点集的I个实例，每个实例由k个节点及其排名Si(·)组成。因此，对于S1中出现的每个节点v，将它们的排名进行聚合，然后选择前k个节点。

例子：假设k=1，R=3，I=3。一开始时刻为t0



**5. A reverse reachable set-based solution**

观察到，PROTEUS-GENIE的时间复杂度受|Er|和|V0|的影响较大, 对于包含数百万个节点的真实网络，这些值非常大，因此在处理此类网络时，贪心算法的效率可能会受到负面影响。在本节中，我们通过提出一个称为PROTEUS-SEER的算法来解决这个问题.

A Reverse Reachable Set

设v为G中的一个节点，g为以概率1 - p删除G中的每条边e后得到的图。对于节点v在图g中的RR集可以表示为R(v,g)，是在图g中可以到达节点v的节点的集合。也就是说，对于RR集合中的每个节点u, g中都有一条从u到v的有向路径。

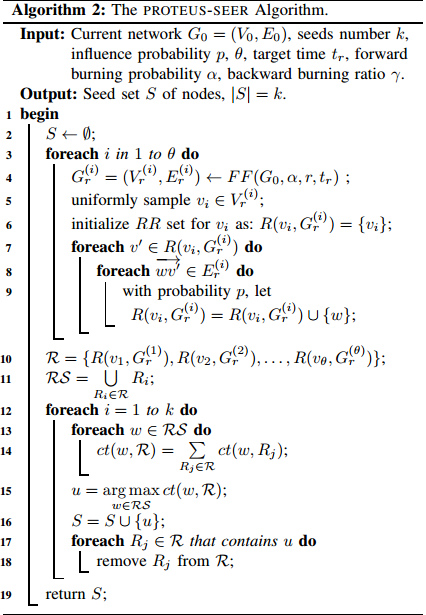
B the proteus-seer algorithm

首先，我们使用FFM来模拟网络G0的演进，得到Gr。这个过程重复θ次，这样可以得到θ个Gr的不同实例,表示为。Θ的次数在参考文献【24】。

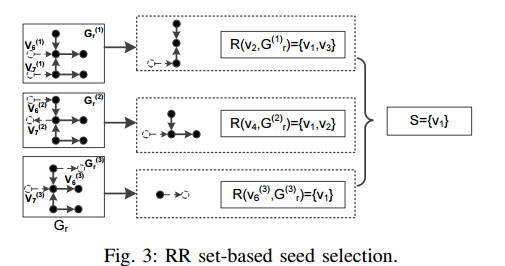
然后，对于每个实例，将Vr中的一个节点统一采样为vi，并为其生成RR集我们有θ个这样的集合,每个对应一个采样节点。最终把这些集合用R=来表示。

最后，我们贪婪地从R中的所有RR集中选择出现在RR集中最多的节点w，选完之后将w从R中删除，再迭代地选择k个节点，并且输出为种子集合。

算法如下：



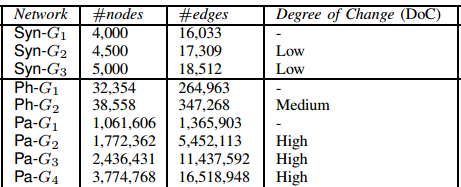
例子2：



**6. 实验**

**A 试验建立**

数据集：



FFM参数：





算法：

Greedy

IRIE

IMM

MaxG

PRO-GENIE

PRO-SEER

**B 实验结果：**