**Analysis of influence maximization in temporal social networks**

**Abstract**

影响最大化问题旨在指定最大限度地影响个体的少数初始个体，这已引起研究者的广泛关注。然而，大多数现有的工作局限于静态的社交网络，忽略了时间在信息传播中的作用。本文分析了时间社交网络中的影响最大化问题，提出了一种基于贪婪的GLAIC算法，该算法通过基于延迟感知独立级联模型的低成本懒惰向前优化得到增强，以捕捉现实社会网络的动态方面。此外，我们还考虑幂律分布，修正了LAIC模型中影响时滞的分布。最后，我们在真实网络上进行了大量的实验，实验结果表明，我们的方案在性能上优于其他相关算法。

**1.Introduction**

最近，社交网络在信息传播中扮演着重要角色，因为越来越多的人喜欢在网络上公开他们的观点或想法。主要研究兴趣之一是了解社交网络中的影响和信息传播方式。例如，一家公司可能希望识别一小部分有影响力的用户，这样他们就可以影响他们的朋友，朋友的朋友，等等，通过使用口碑效应来采用一种产品。社会影响分析已广泛应用于许多领域，如病毒营销和推荐系统。

社会影响力的一个目标是如何找到最终会影响最大数量用户的初始用户，这被称为影响力最大化。同时，也有许多其他方法结合现实场景和问题，如主题意识的影响最大化[21]，基于社区的影响最大化[22-23]，签署网络中的影响最大化[24-25]，竞争影响最大化[26-27]等等。

上述解决方案旨在从不同方面解决影响最大化问题，并取得了不错的实验结果。然而，大多数研究假设给定的网络是静态的，忽略了时间在信息传播中的作用。事实上，几乎所有的社交网络都是动态的，并且会随着时间而变化。两个用户在下一个时间段可能互相关注也可能互相取关。在这种情况下，在解决影响最大化问题时，考虑时间因素可以恰当地反映社会网络的实际结构。

在本文中，我们放弃了静态网络的假设，将注意力转向考虑随时间变化的动态网络。我们分析了时间社会网络中的影响最大化问题，描述了网络的动态特性。为了纳入时间因素，我们考虑了刘等人提出的延迟感知独立级联模型（LAIC）[5]。此外，我们还讨论了模型中重要参数的分配，即影响时延的分布。作者方案的核心思想是基于LAIC模型，考虑影响时延的幂律时延分布，贪婪地选择最有影响的节点。因此，一个简单的基于贪婪的CELF优化算法，称为GLAIC(基于LAIC的贪婪)，被用来寻找最大的影响节点，以解决时间社会网络中的影响最大化问题。请注意，我们的建议通过采用幂律延迟分布而不是泊松和几何延迟分布等其他分布获得了更高的效率。

贡献点：

1. 在LAIC通过将幂律分布应用于影响时延，开发了一种新的时延分布，与后面的实验中所示的泊松和几何时延分布相比，该分布可以提供更好的实验结果。

2. 我们提出了一种基于贪婪和对CELF优化的GLAIC算法，基于LAIC模型捕捉现实社会网络的动态方面。

3. 我们进行了大量实验来验证我们的建议在真实社交网络上的有效性。

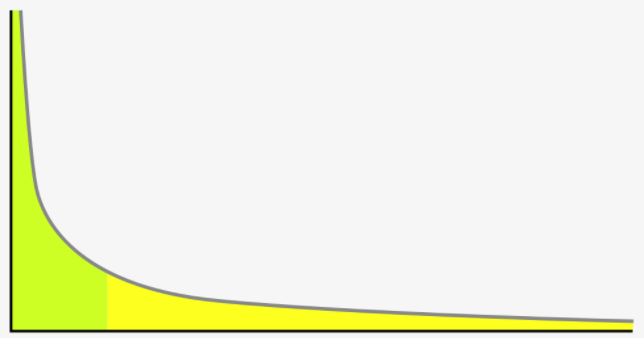
**2.Related work**

扩散模型是一个基本规则，它定义了信息如何在指定的社交网络中传播。然而，这一问题在两个经典扩散模型，即独立级联模型和线性阈值模型上被证明是NP-hard。这些模型假设每个节点在网络活动状态或非活动状态下只有两种状态。如果一个节点已经接受了消息，该节点将是活动节点。否则，该节点将是非活动节点。最经典的解决方法是贪婪算法，但是他的复杂度很高，之后Leskovec提出来了CELF算法可以降低计算复杂度。

同时，研究团体最近也研究了考虑不同真实场景的影响最大化问题。Chen等人[21]试图通过考虑主题来扩大影响力。Chen等人[22]和博佐吉等人。[23]分析了社区结构在影响传播中的作用，并利用社区内节点的影响来近似整个网络的影响。Lin等人[26]提出了一个通用模型来描述多个竞争源的影响传播，并提出了两阶段竞争影响最大化来解决部分信息下的竞争影响最大化(CIPM)问题。

上述算法从不同角度努力减少计算量，提高精度，取得了良好的实验结果。然而，这些算法假定社交网络是静态的，节点或关系不会随着时间而改变，这是不太现实的。因此，越来越多的研究从当太的角度考虑时间的作用，以下研究提出了考虑时间因素的各种时间影响传播模型，以解决影响力最大化的问题。Wang等人[15]定义了流影响最大化查询，通过采用滑动窗口模型实时跟踪有影响的用户。Li等人[16]提出传播时间意识影响最大化(PROTEUS-IM)，以解决影响传播过程中社交网络进化的局限性。刘等人。[5]提出了LAIC模型，将时间与时间期限约束下边缘的激活概率相结合。特别的，如果一个节点u在时间t被激活，那么他会以概率在时间去激活它的不活跃节点v。其中是从延迟分布中随机抽取的。与此同时，他还提出了另两个适用于这个模型的影响力最大化算法ISP和MISP。在他的工作中，考虑了两种分布，分别是泊松分布和几何分布。但是，大量最新研究显示，社交网络通常遵循幂律分布。所以我们认为使用幂律分布会比使用泊松和几何分布好一些。





**3. PRELIMINARIES**

对于一个社交网络G=（V,E），V是结点集合，E是边的集合。使用PUV来代表节点u能影响到节点v的可能性。

在带有时间限制T的影响力最大化问题有以下标识。1.目标种子集2.影响传播 3.边际收益Mg。节点u的边缘收益为。

LAIC模型通过考虑时间延迟来扩展传统IC模式，这与现实社会网络更加一致。在这个模型下，每个结点的状态分为三种，激活，延迟激活，非激活。对于一个在种子集合中的结点u，可以认为他在t=0时刻处于活跃态，如果它成功影响到了一个非活跃邻居节点v，那么节点v变为延迟活跃节点。根据之前定义的影响力延迟的分布，每条边e（u，v）可以扩展到，其中代表了一个非活跃节点v被他的活跃邻居u在一个激活延迟以概率激活。是一个延迟分布。

LAIC模型的过程如下:在任意一步t>0，在种子集S中的结点u可以有一次激活邻居节点v（v是处于非激活状态或者延迟激活状态）与传统IC模型不同的是，被激活的节点v会变味延迟激活状态而不是激活状态，在经过延迟之后，延迟激活节点v会被种子集合S激活。

在LAIC模型中，影响传播有三个特征。1.不增性2.单调性3.子模性。具体说。对于任何集合，f( )满足就叫做不增性。

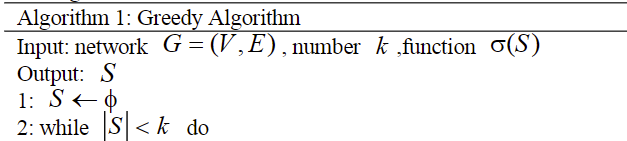
**4. the proposal**

我们首先描述我们关于影响最大化的GLAIC算法，使用了改良的CELF并且考虑到幂律时延分布。我们首先使用经过CELF优化后的贪婪算法这样可以减少计算复杂度，然后在LAIC模型中使用幂律分布代替原有的分布，最后提出GLAIC算法，基于LAIC模型上计算影响范围。

**A 被CELF优化的贪婪算法。**

1）贪婪算法

我们使用用来表示期望激活结点的数目，贪婪算法如下：





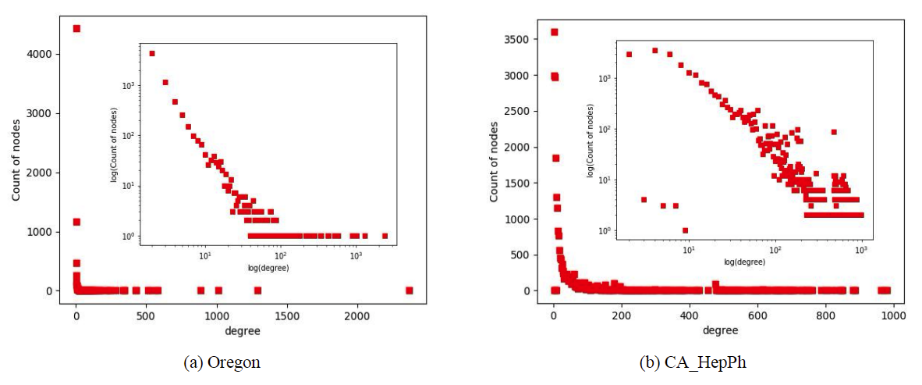
2）CELF优化

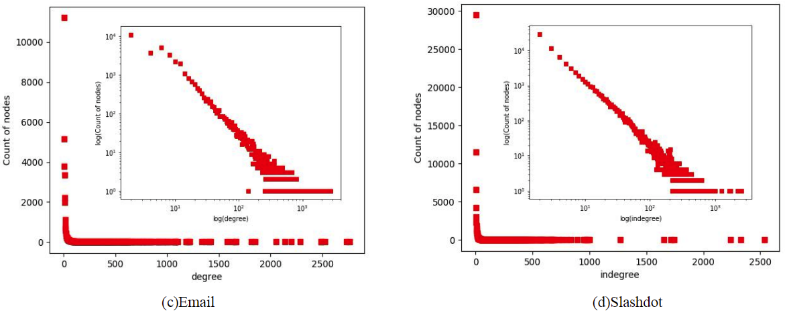
CELF利用了子模性，他维护了一个优先队列，这个队列按Mg（u）的大小降序排列，这样的话就不用每次都去迭代过程都去重新构造这个队列，每次都选择列头的节点就好。

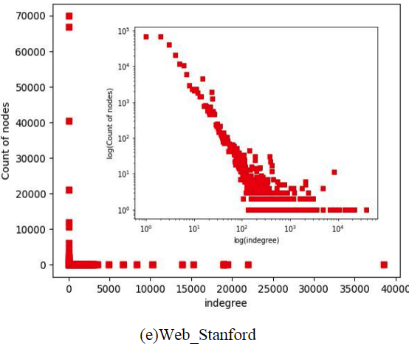
**B 幂律延迟分布**

当时用LAIC模型时，一个很重要的问题是如何选择适当的延迟分布，也就是参数，当前主要是使用泊松分布和几何分布，但是现有的研究表明很多网络遵循幂律分布，也就是说只有少数的节点有很多的边，大多数节点都只有很少的边。对于节点的度k，k的可能性分布遵循幂律，。

我们对五种社交网络中节点的度和入度的分布进行了统计。五种网络分别是Oregon, CA\_HepPH, Email, Slashdot, and Web\_Stanford。在这五个网络中中Oregon, CA\_HepPH是无向网络，我们绘制了他们节点的度的分布，剩下三个是有向网络，所以绘制了他们的入度分布图。

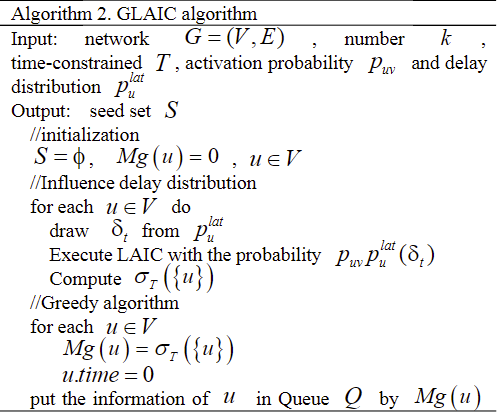


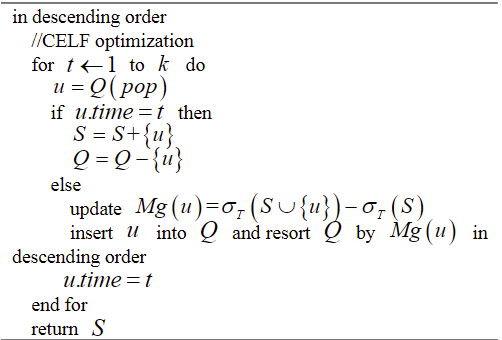




明显早直方图中可以看出度或者入度的分布跟幂律的图像很接近，在log-log图中，分布更加接近于线性。所以这给我们一个提示，如果选择幂律分布会不会是一个更好的方法。

**C. GLAIC算法**





**5.实验结果**

**A 试验建立**

**1）数据集**

我们用了五个真实的网络

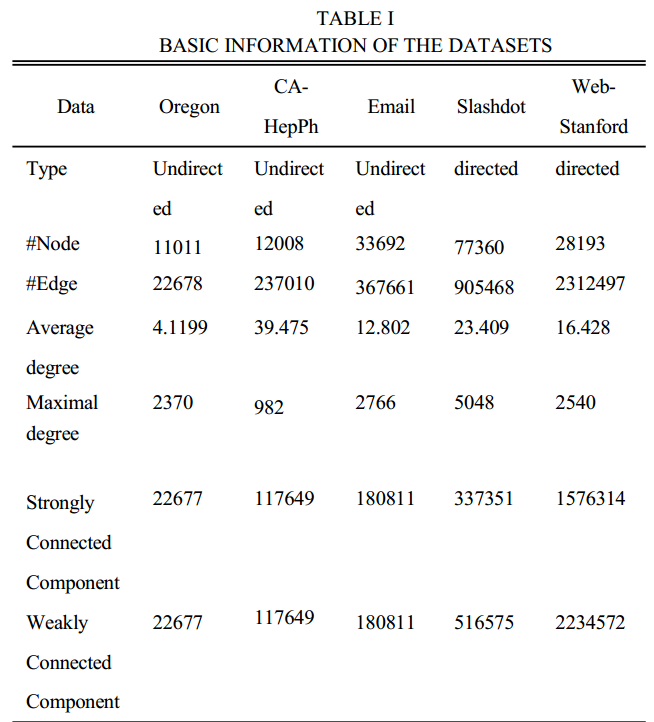
1.Oregon：数据集涵盖了九个自治图，从2001年3月31日到2001年5月26日，一个相对较小的网络，系统对从俄勒冈州收集到的信息进行监视一个小网络

2.CA\_HepPH：在高能量物理领域的引用网络

3.Email：从Enron Email中抽取的，包含1700个带标签的邮件信息

4.Slashdot：数据集来自于2009年2月的Slashdot社交网络，它在Slashdot用户的伙伴之间有朋友/敌人的连接

5.Web\_Stanford：数据集包含来自Stanford的web页面之间的超链接



**2）延迟分布**

对于每个节点u，他的延迟分布定义为：

泊松分布：泊松分布的参数从1-10中随机选择

几何分布：几何分布的参数为

幂律分布：k-a，a取1-4

**3）算法**

1.使用泊松的ISP：该算法由[5]提出。它假定一个节点只被一个影响扩散路径所激活，而这个影响扩散路径与其他路径的边缘并不相同。此外，它只考虑泊松延迟分布。

2.使用几何的ISP：与上面一个算法很相似，只不过使用了几何分布延迟分布

3.使用泊松分布的MISP：该算法由[5]提出，它是ISP的一种变体。它们之间的区别是MISP不仅考虑节点的影响扩散，而且考虑节点被激活的外部邻居的概率。

4.使用几何分布的MISP：与上述原理一样，只不过使用几何分布的延迟分布

5.MIAM与MIAC：针对时间临界影响最大化问题。

6.Degree and Discount: 该算法是一种启发式算法，具有均匀的传播概率，本文与往常一样取0.01

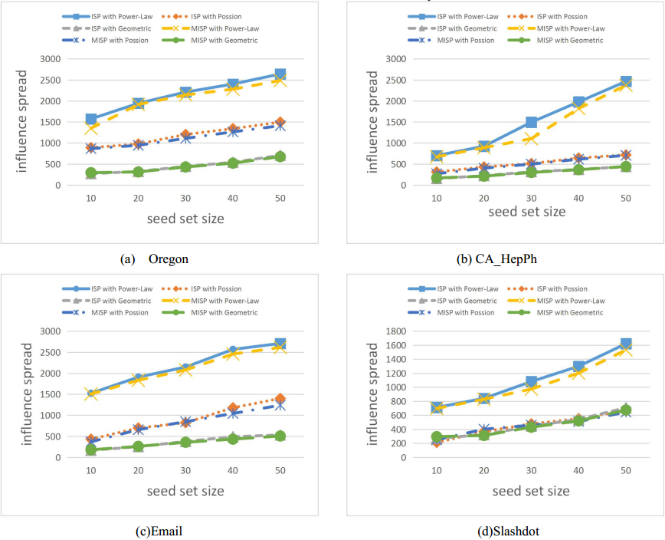
7.Random：随机选取种子

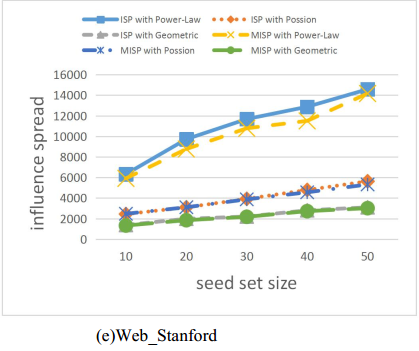
对于传统的影响最大化问题，我们按照之前的工作对每个目标集运行10000次，然后取其平均值。我们取的种子集的大小为10-50。

B.结果

1）幂律延迟分布的影响

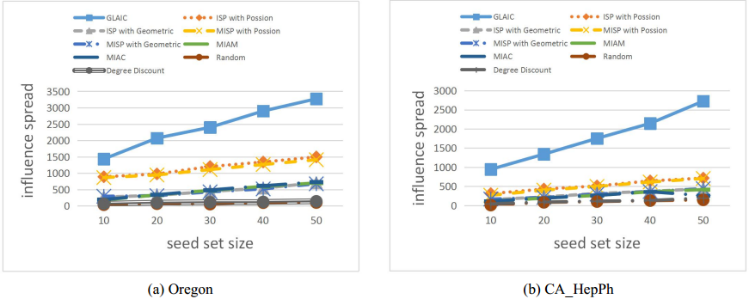
我们首先研究幂律延迟分布对影响扩散的影响并且与泊松分布和几何分布作比较，当T=10时。我们将ISP与MISP与幂律分布进行结合。结果如下。

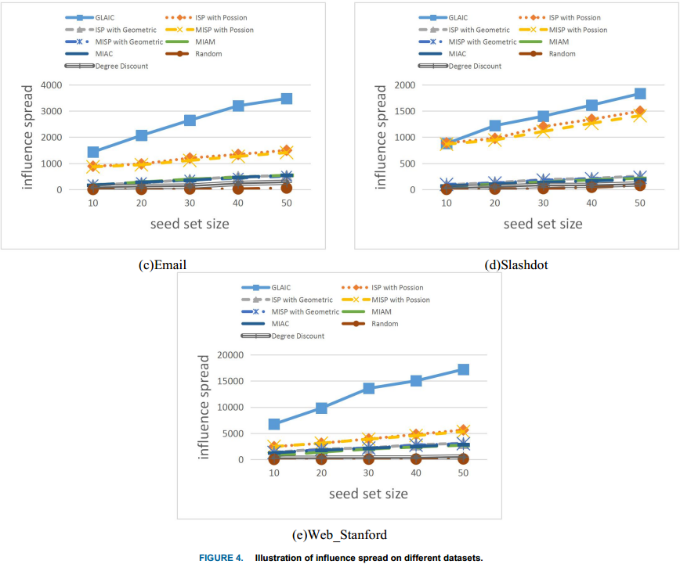




可以看出ISP-PL和MISP-PL的影响传播更大

2）影响传播





可以看出作者提出的算法的影响范围要比现有的要好得多。

**3）运行时间**

