Author = Song Jian

#### 阅读论文：HARP: Hierarchical Representation Learning for Networks

|  |  |  |
| --- | --- | --- |
| @article{Chen2018,  author = {Chen, Haochen and Hu, Yifan and Perozzi, Bryan and Skiena, Steven},  journal = {32nd AAAI Conference on Artificial Intelligence, AAAI 2018},  pages = {2127--2134},  title = {{HARP: Hierarchical representation learning for networks}},  year = {2018}  } | Chen, H., Hu, Y., Perozzi, B., & Skiena, S. (2018). HARP: Hierarchical representation learning for networks. 32nd AAAI Conference on Artificial Intelligence, AAAI 2018, 2127–2134.  [文章下载链接](https://arxiv.org/pdf/1706.07845.pdf) | 美国石溪大学  雅虎研究员  谷歌研究员 |

#### 文档由以下几个部分组成：

* 文章背景
* 文章贡献
* 问题定义
* 算法流程
* 实验

#### 文章背景

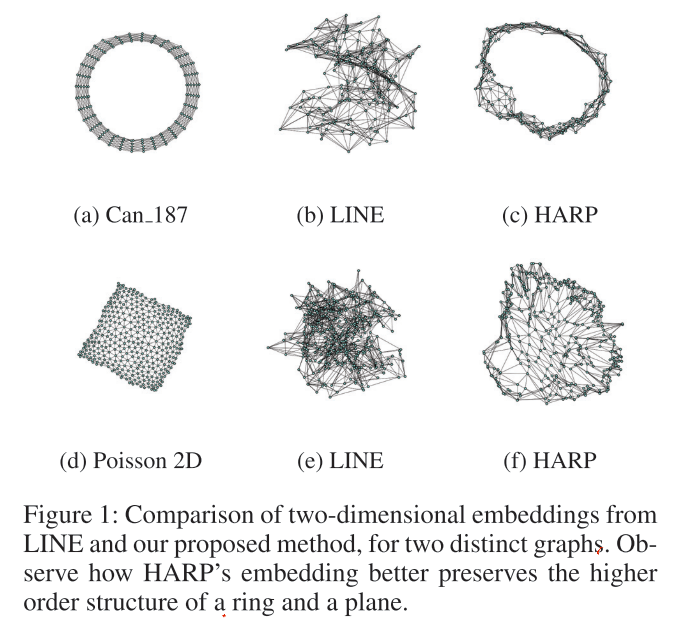
Network embedding领域在近几年发展迅猛，出现了许多效果公认不错的算法，比如：DeepWalk、Node2Vec和LINE这三个算法。但是这三个算法都有两个共同的问题：

1. 基本都把考虑的重点放在了网络的局部结构关系（如：一阶相似性和二阶相似性），而忽略了距离较长的全局关系。
2. 它们都是通过随机梯度下降法对一个非凸的目标函数进行优化，这样如果初始化不好就很容易陷入局部最优。

#### 文章贡献

作者提出的HARP算法希望通过递归地粗粒化方式，将原网络图的节点和边通过折叠划分成一系列分层的结构更小的网络图，然后再利用现有的算法进行不断的特征提取，从而实现最终的network embedding特征提取，这种方法很好的保留了图的高阶结构特征。所以，在这篇文章中，作者的贡献可以总结为以下几点：

1. 提出新的表示学习范式(通用框架，可直接用于上述的三种embedding算法)
2. 改进优化过程的效果
3. 更好的embedding效果，有利于接下来的任务



#### 问题定义

图表示学习的定义是寻找一个映射函数(mapping function)。这个函数定义了每个节点的潜在表示(latent representation)或者(embedding)。学习Φ的参数的常见模型有DeepWalk,LINE,Node2Vec。

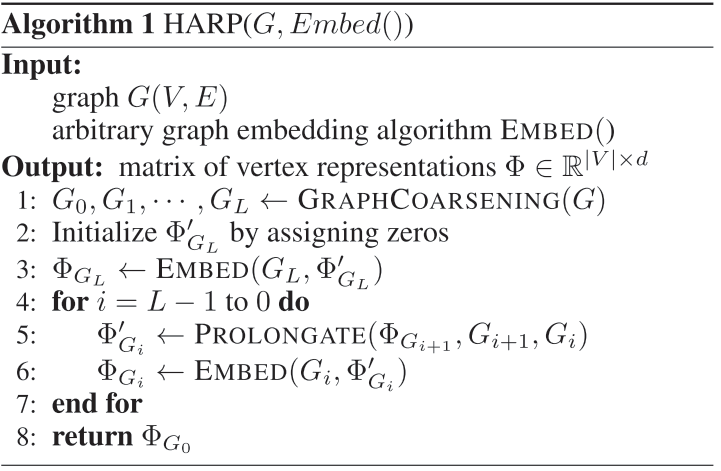
在这篇文章中，作者希望找到比原图更小的图，其中

而且，我们得到的图会更加有利于做embedding，原因有以下两个：

1. 相比于原图, 的节点对之间关系更少,

2. 图 的直径要比图小，当小到一定程度时，基于局部结构的算法也能利用到图的全局结构。

#### 算法流程



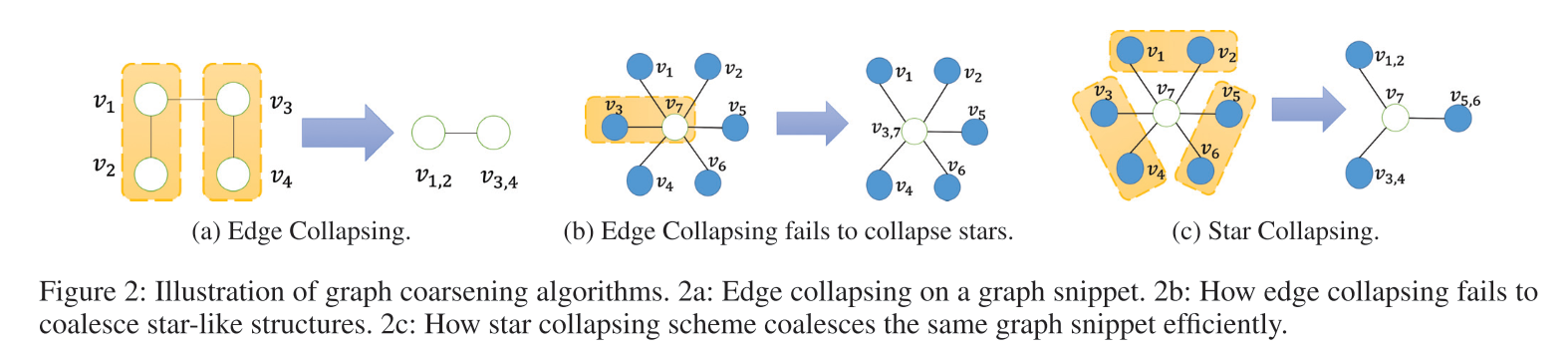
第1行：将原图划分为多层次的小图(折叠，粗粒度化)，具体算法后面会介绍；

第2-3行：初始化参数 是划分到最后最小的图,直接利用现有算法(DeepWalk,LINE,Node2Vec)对其求embedding；

第4-7行：迭代地对比较高层次的（较大的）图和对较低层次的（较小的）图的embedding结果进行结合延伸，作为较大图的embedding参数输入进行embedding；

第8行，返回原图的embedding结果。

#### 图折叠(粗粒度化)



* 边折叠

边合并是为了保留一阶相似度，如图2a

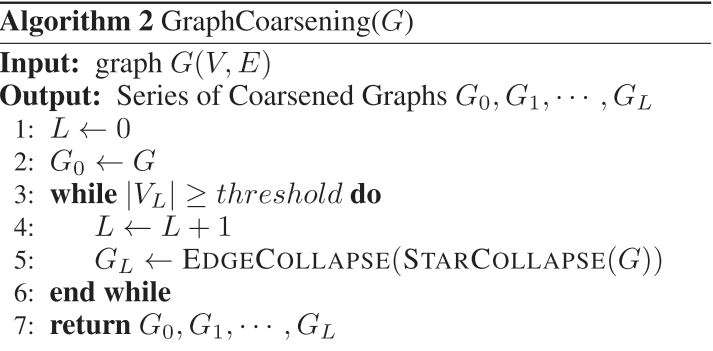
* 星折叠

星折叠的原理是**利用了二阶相似度**。如果图2b的星型结构按照边折叠的方法，会出现k个节点的图折叠为k-1个节点的图，折叠次数会增加。

因此考虑二阶相似度的关系，可以如图2c的方法折叠，共享邻居，因此可以考虑将他们两两折叠为一个节点。

* 混合折叠

先按星折叠，再按边折叠，直到节点小于一个阈值，停止折叠。每一层

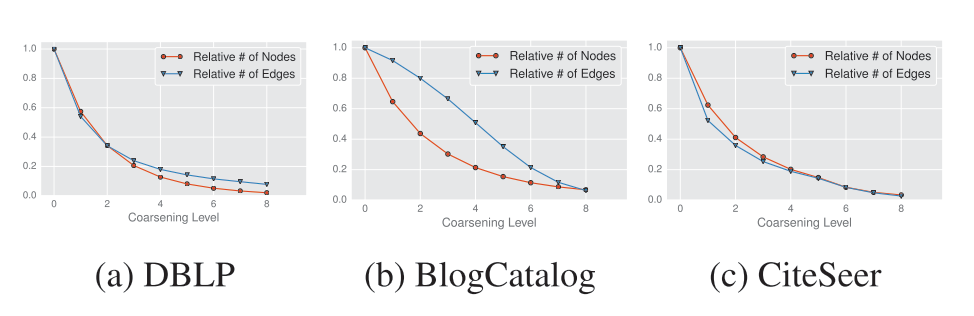


第1-2行：初始化L为0，输入的原图为。

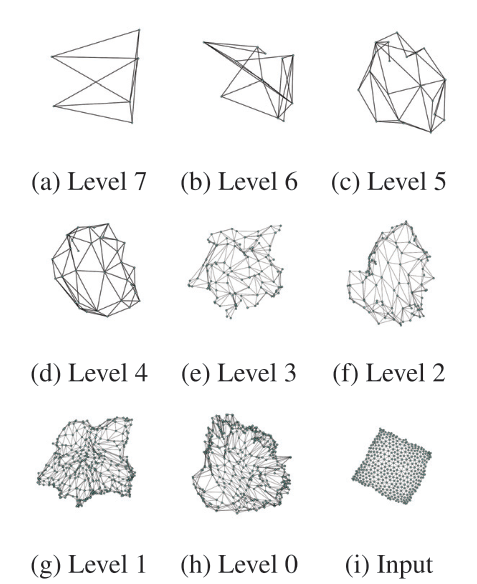
第3-6行：判断当前图的是否小于阈值(文中为100)，递增L，星折叠，然后对结果进行边折叠，之后将结果赋值为

第7行：输出

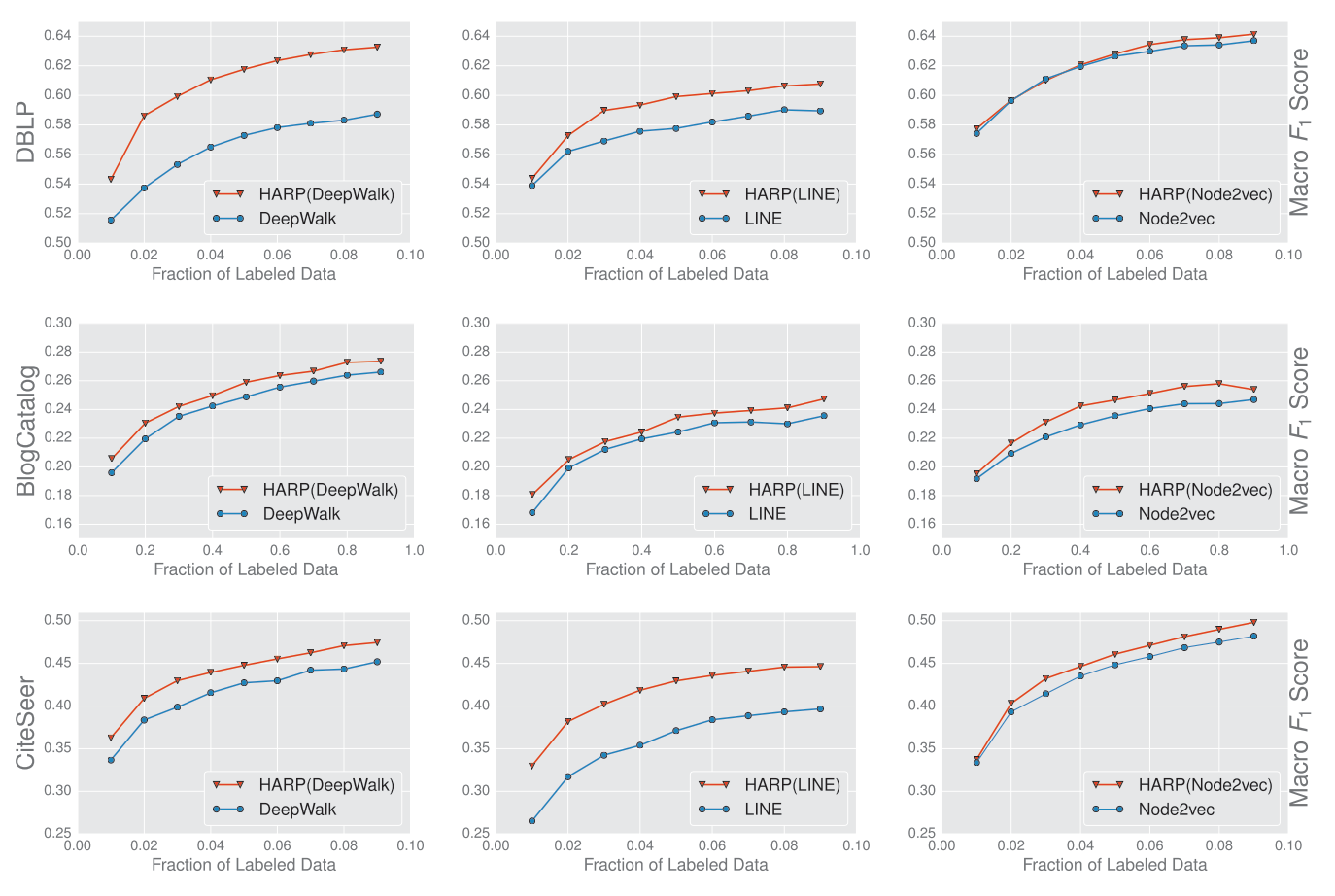
#### 实验



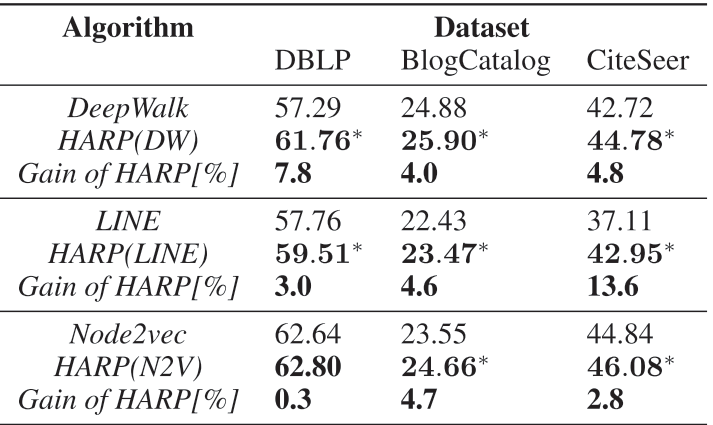
不同数据集图粗粒度化层级与原图的节点比例



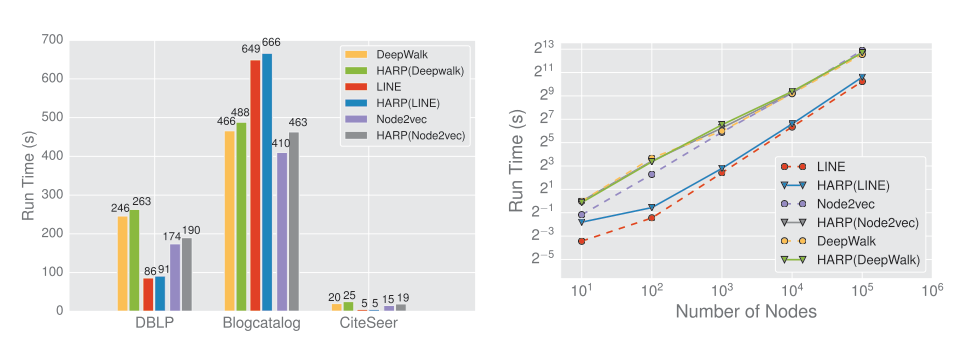
Possion 2D 在HARP(LINE)不同层级上的表示



多标签分类，使用HARP框架与原方法的性能对比



使用框架后的性能提升



运行时间对比 1