01



该文章是 Google research 发表的，被 KDD 2019收录。

GCN 成功地应用于很多基于图的任务，但是训练大规模图仍然很有挑战性。最近基于 SGD的算法要么需要计算消耗，尤其是随着 GCN 层数的增加消耗也逐渐增大，要么就是需要大量的存储空间来保持整个图以及存储每个节点嵌入。在这篇论文中，作者提出了

Cluster-GCN ，一个新的 GCN 算法，适用于利用图聚类结构进行基于 SGD 的训练。

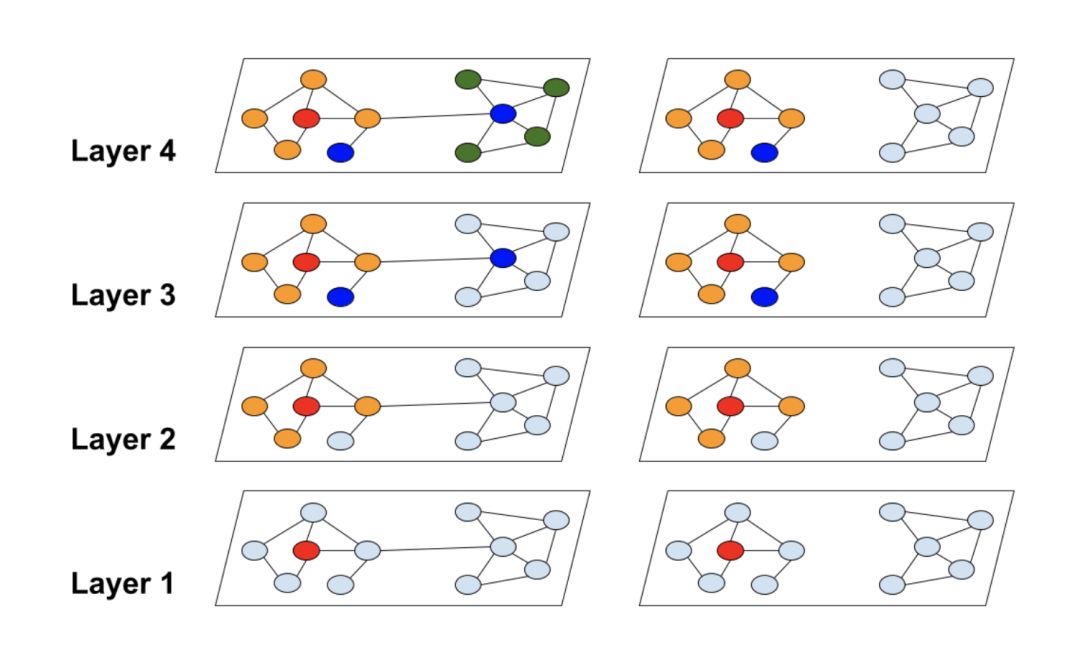
Cluster-GCN 的主要工作原理如下：在每一步，

通过图聚类算法识别高密子图，

然后采样与该子图相关联的节点块，并限制该子图中的邻域搜索。

这看起来很简单，但是对存储和计算效率的提升确是直接有效的，并且和之前的算法比起来，准确率也有所提升。

为了测试这个新的算法，作者还创建了一个新的 Amazon2M 数据集，包含 200 万节点和 6100 万边，是之前的数据集的 5 倍。



**论文链接：https://arxiv.org/pdf/1905.07953.pdf**