

第6章：在大图上的随机（批次）训练

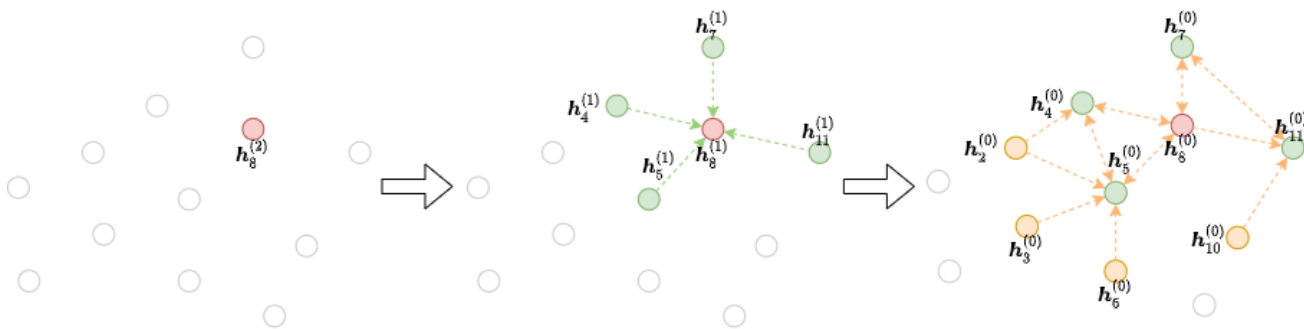
(English Version)

如果用户有包含数百万甚至数十亿个节点或边的大图，通常无法进行[第5章：训练图神经网络](#)中所述的全图训练。考虑在一个有 N 个节点的图上运行的、隐层大小为 H 的 L 层图卷积网络，存储隐层表示需要 $O(NLH)$ 的内存空间，当 N 较大时，这很容易超过一块GPU的显存限制。

本章介绍了一种在大图上进行随机小批次训练的方法，可以让用户不用一次性把所有节点特征拷贝到GPU上。

邻居采样方法概述

邻居节点采样的工作流程通常如下：[每次梯度下降，选择一个小批次的图节点，其最终表示将在神经网络的第 \$L\$ 层进行计算，然后在网络的第 \$L - 1\$ 层选择该批次节点的全部或部分邻居节点。重复这个过程，直到到达输入层。](#)这个迭代过程会构建计算的依赖关系图，从输出开始，一直到输入，如下图所示：



[该方法能节省在大图上训练图神经网络的开销和计算资源。](#)

[DGL实现了一些邻居节点采样的方法和使用邻居节点采样训练图神经网络的管道，同时也支持让用户自定义采样策略。](#)

本章路线图

本章的前半部分介绍了不同场景下如何进行随机训练的方法。

- [6.1 针对节点分类任务的邻居采样训练方法](#)
- [6.2 针对边分类任务的邻居采样训练方法](#)
- [6.3 针对链接预测任务的邻居采样训练方法](#)

本章余下的小节介绍了更多的高级主题，面向那些想要开发新的采样算法、想要实现与小批次训练兼容的图神经网络模块、以及想要了解如何在小批次数据上进行评估和推理模型的用户。

- [6.4 定制用户自己的邻居采样器](#)
- [6.5 为小批次训练实现定制化的GNN模块](#)
- [6.6 超大图上的精准离线推断](#)