

日期: /

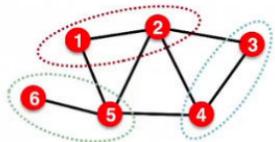
Setting - up GNN Prediction Tasks

如何对数据集进行划分?

对图数据集进行划分是特殊的, 因为节点间并不独立。
(Node, Edge level)

Node 5 will affect our prediction on node 1, because it will participate in message passing → affect node 1's embedding

Training
Validation
Test



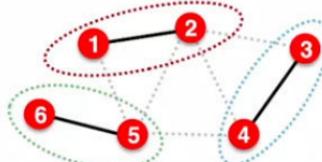
[Transductive Setting (直推式设置)] 存在部分信息泄露

- 输入图在所有分割中均可见 不适用于图级任务
- 只对节点标签进行划分而图结构不变

训练/验证时, 使用全图计算嵌入, 使用节点 1, 2 / 3, 4 的标签

[Inductive setting (归纳式设置)] 丢失部分结构信息, 适用
打破分割之间的边以获得多个独立图: 子图级任务

Training
Validation
Test



日期： /

在训练/验证/测试时，不仅使用各自的节点标签，
同时使用各自的输入图计算嵌入

[Link Prediction]

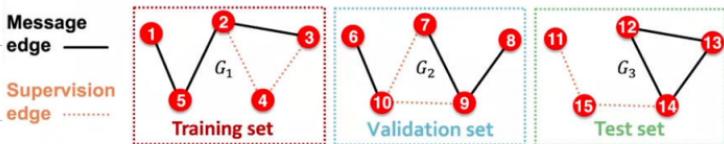
- 连接预测的目的是确定图中的缺失边。

- 这是一个自监督任务，我们需要从图中隐藏部分边并由GNN产生预测。

Step 1：向原始图分配两类边(message edge/supervision edge)
(只有 message edge 会在图中保留)

Step 2：划分数据集：

Option 1: Inductive link prediction split



在每个分割中使用不同图，不同边划分

日期: /

Option 2 : Transductive link prediction split:

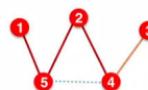
④



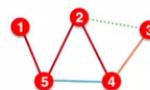
The original graph



(1) At training time:
Use **training message edges** to predict **training supervision edges**



(2) At validation time:
Use **training message edges & training supervision edges** to predict **validation edges**



(3) At test time:
Use **training message edges & training supervision edges & validation edges** to predict **test edges**

稀疏 → 組密