

# 作业5：网络嵌入

- **经典node2vec算法：**
  - 1) Compute random walk probabilities
  - 2) Simulate  $r$  random walks of length  $l$  starting from each node  $u$
  - 3) Optimize the node2vec objective using Stochastic Gradient Descent
- **如何将node2vec生成的表征用于节点分类？一个可行的思路：**
  - **在Cora全图（包含测试集、训练集）上用特征 $x$ 训练出节点特征；**
  - **在训练集训练一个将节点表征映射到分类向量的MLP分类器，在测试集测试性能。**

## 作业5：网络嵌入

- **(9分)** 在Cora数据集上实现node2vec进行节点分类的算法并测试其性能。数据集采取9:1的随机训练:测试划分，参数 $q$ 和 $p$ 自由选取（可以参考node2vec原始论文）。算法实现需要完整包含随机游走概率计算、随机游走采样、node2vec优化、分类器训练等部分，模型性能应该合理（测试精度至少达到50%）。
- **(1分)** 任选其一完成：
  - 在Cora以外的更多数据集（torch\_geometric.datasets的点分类数据集）上试验不同的 $q$ 和 $p$ 参数。不同的 $q$ 和 $p$ 对不同数据集上性能有什么影响？请结合实验结果分析并给出初步结论。
  - 还有什么方法能从node2vec的特征得到最终的分类向量（直接MLP以外）？实现你选择的方法，并比较其对性能的影响。