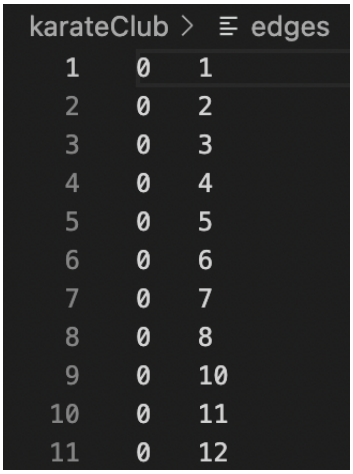
第一次作业

数据集介绍

边数据

Cora和KarateClub目录下分别是Cora的边数据和KarateClub的边数据，形式如下：

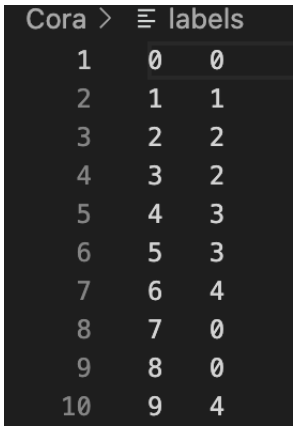


edges中有E行（E为边的数目），每一行表示一条边，包含两个整数，分别是源结点的ID（ID编号从0开始），目标结点的ID。

提示：Cora的边数据为**有向边数据**，而KarateClub的边数据为**无向边数据**。即在KarateClub的edges里面，若存在边(i,j)那么一定存在边(j,i)，但是Cora里面不是这样。

标签数据

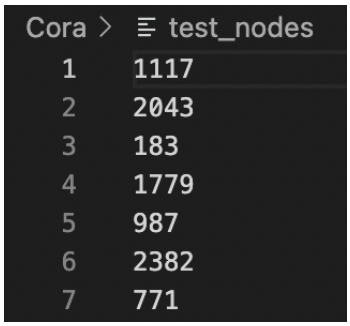
Cora目录下的labels表示结点的标签数据，形式如下：



labels有N行(N为结点的数目)，每一行的第一个整数表示结点的ID，第二个整数表示标签的ID（标签ID也从0开始）

训练结点数据和测试结点数据

Cora数据集下的train_nodes和test_nodes分别表示训练结点数据和测试结点数据，形式如下：



Cora >	test_nodes
1	1117
2	2043
3	183
4	1779
5	987
6	2382
7	771

train_nodes / test_nodes 里面分别是：训练结点的ID / 测试结点的ID。

报告内容介绍

作业报告需要包括以下内容：

1. 将图数据视为**无向图**，计算图数据的：

- 平均结点度数，以数值的形式给出
- 度分布，以直方图的形式给出，横轴k代表度的取值，纵轴P(k)代表任取结点度数为k的概率
- 平均结点聚集系数，以数值的形式给出。

数据的计算参考第一次课程的PPT。

2. 简单比较分析社交网络数据（KarateClub）和引文网络数据（Cora）的不同。

3. 从Relational classification，Iterative classification，Belief propagation（参考第三次课程的PPT）任选一种方法，在Cora数据集里，根据训练结点的标签对测试结点的标签进行预测，报告预测的准确率。

提交说明

第一次作业提交报告即可，无需提交代码。

作业以邮件形式发送到：2001213110@stu.pku.edu.cn，邮件主题和报告名均为：学号+姓名+第一次作业报告。

作业截止时间为：2021年11月10日晚24:00。