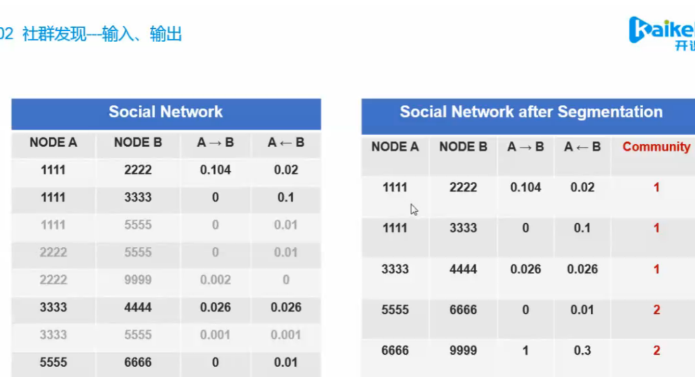
有向加权，从a→b，b→a



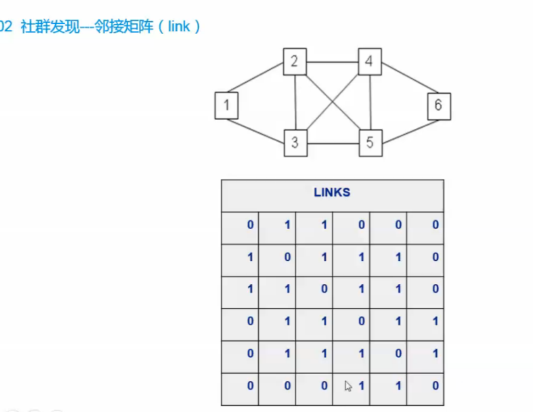
左边是输入，右边是输出

Q：社团划分是不是每一条边是一个社团？

答：每一条边不一定是一个社团

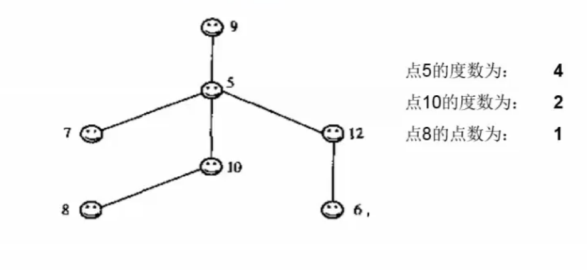


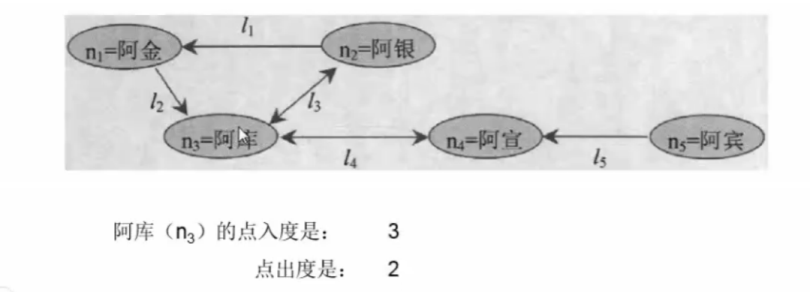
图论、离散数学、数据结构有相关概念

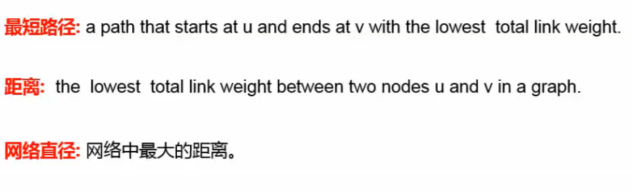


有边为1，没有边为0

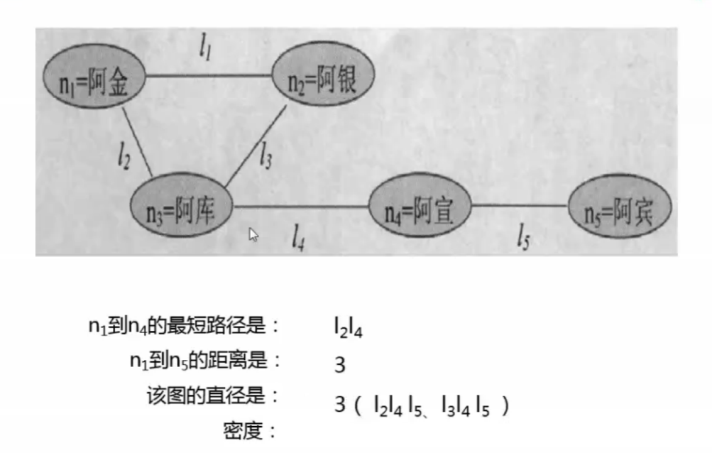
相邻的边数







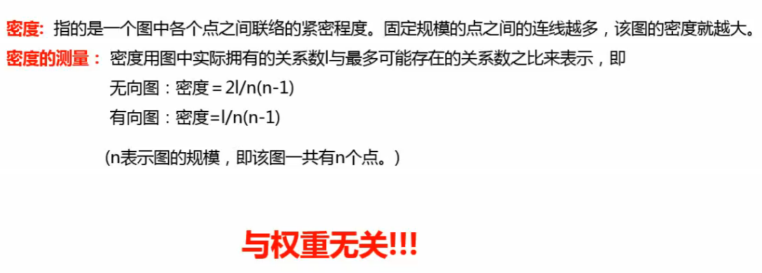
网络直径：两两距离挑最大的

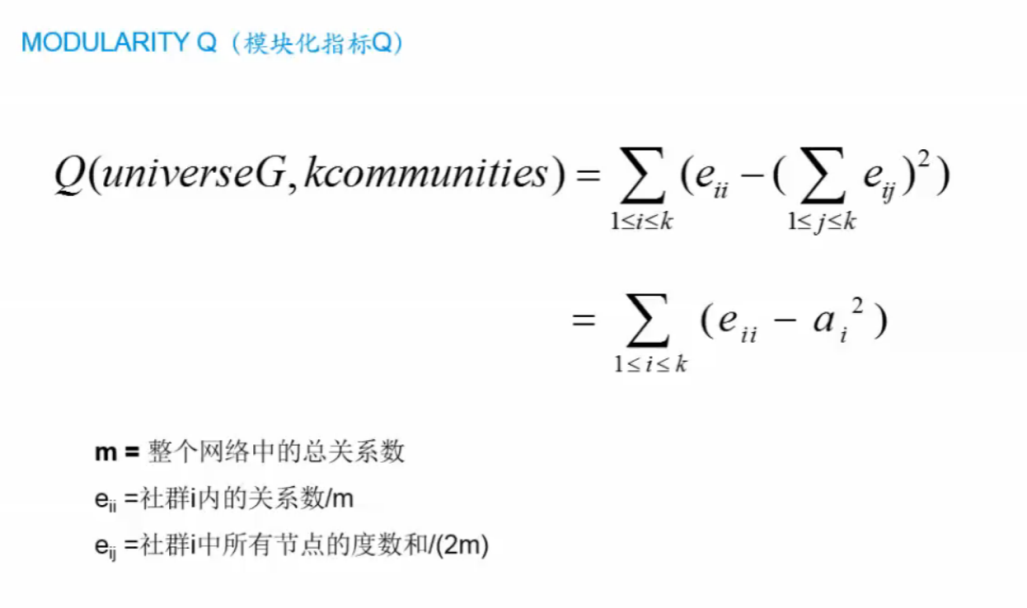


密度：0.5

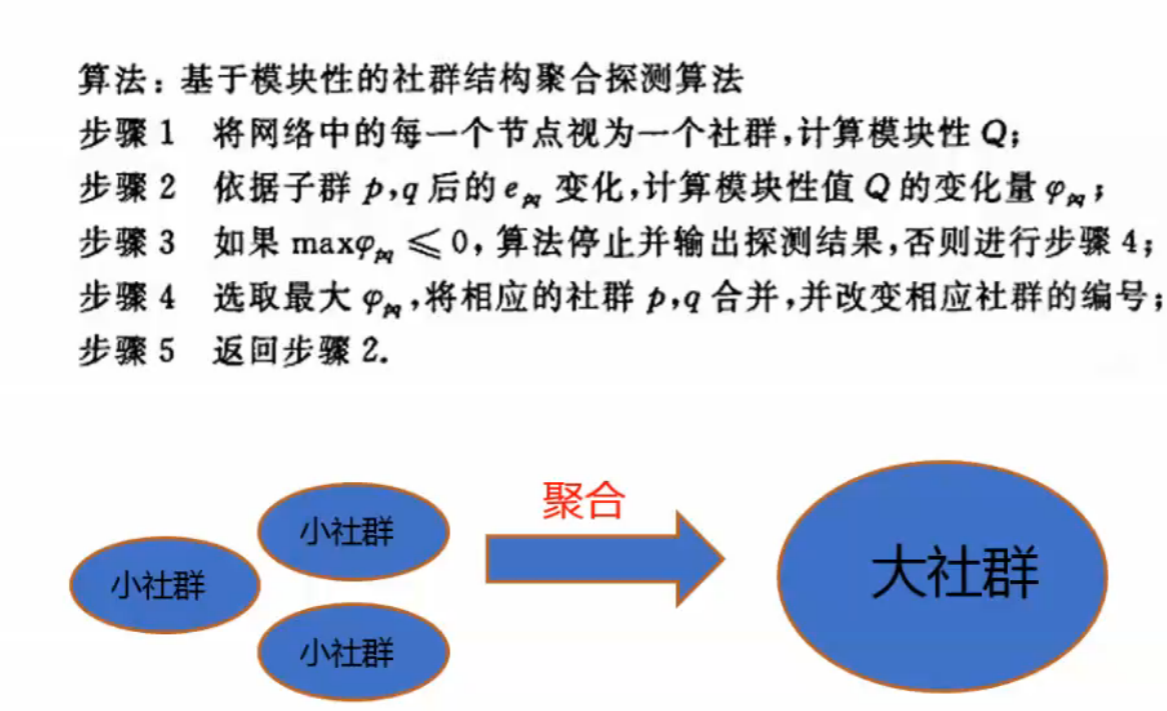
实际有的边数/可有的边数，5/10=0.5

5/（5x4÷2）





Q表示越大好，表示划分的结构越清楚了



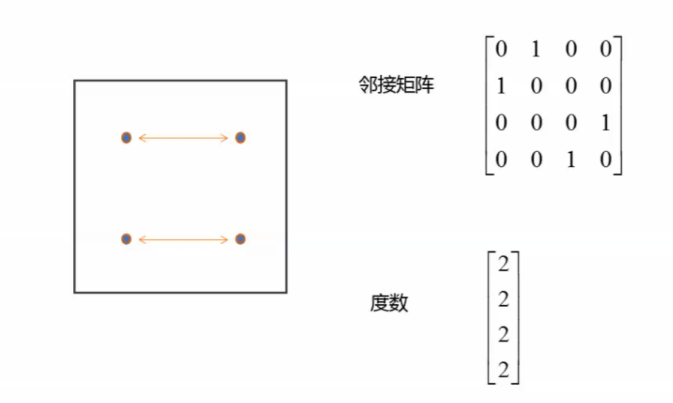
聚类算法跟凝聚法相似

Q值小于0.3表示没有明显的社群

Q值0.5可以分但是有散户

Q值大于0.7，可以分成明显社团

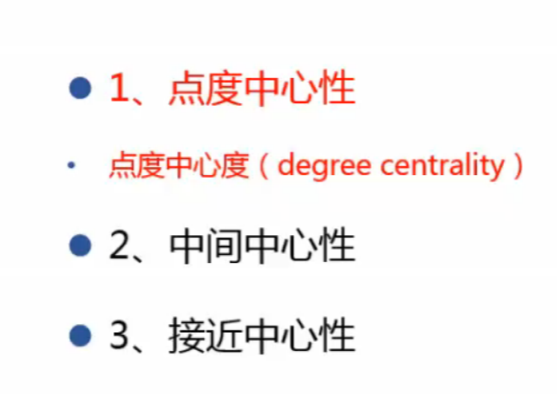
交给：pyhon和r， gehip

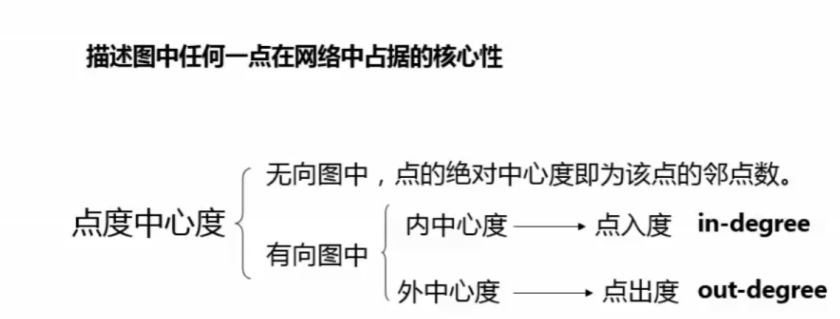


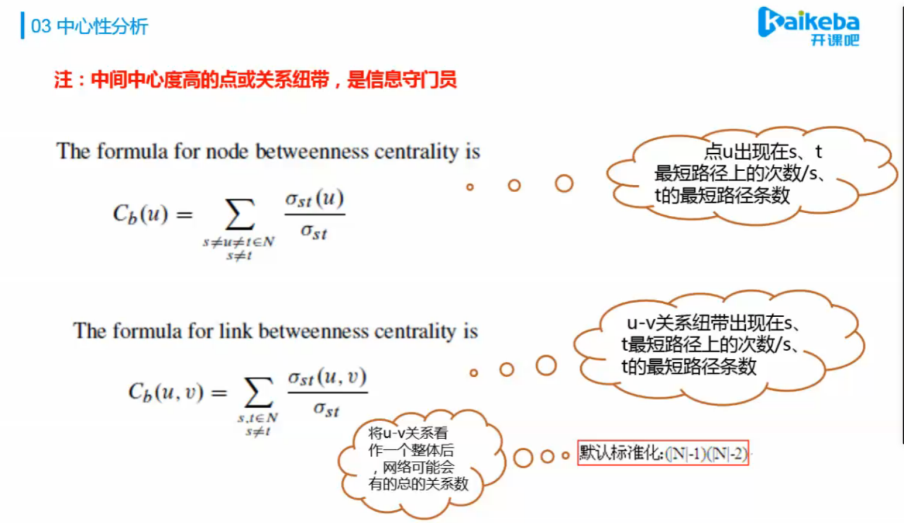
知识图谱是可解释AI的利器

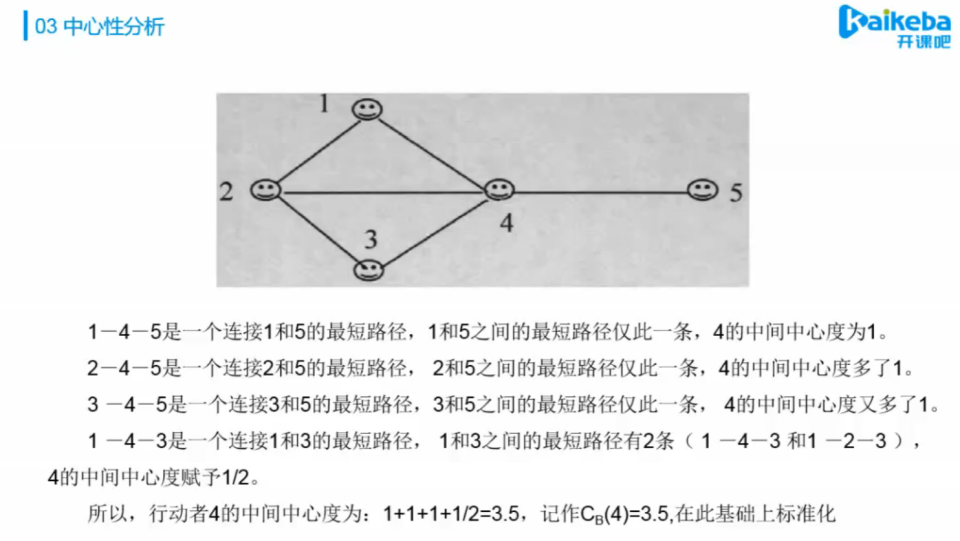
NLP和推荐系统

神经网络可以解释，但是不可解释。









Kaggle入门

Lettcode里面去练习代码

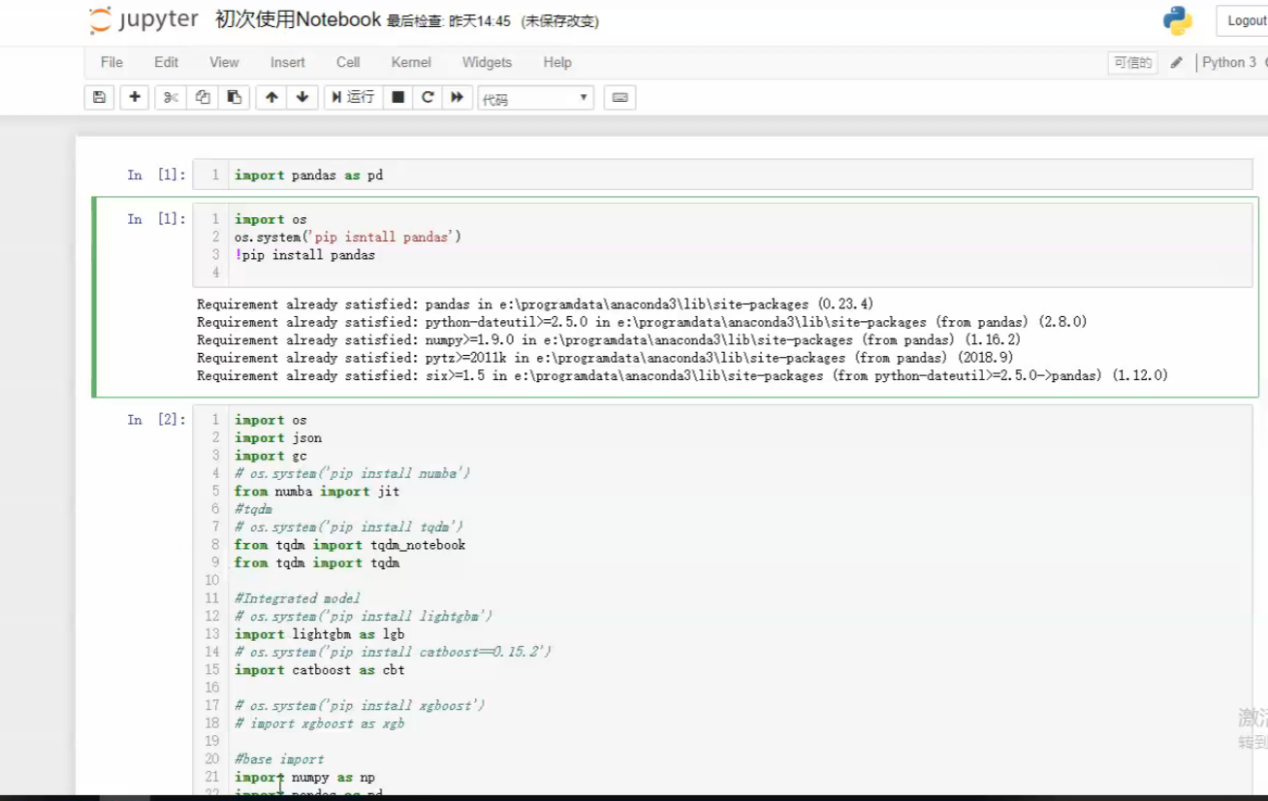
剑指offer:

Python基础： notebook

怎么用python在国际精彩平台

怎么封装函数？

怎么提交

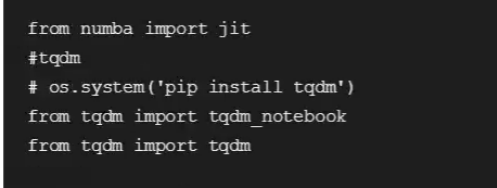


Os:系统包

Json:文件

Gc:回收机制的包

tqdm:显示进度条的时间



Numpy,pandas包

Sklearn包

%% time 显示代码显示的视角

NLP: 多用 import re包

怎么快速写一个baseline?

阿里天池的包：

Tianchi.aliyu.com/home

需要注册

技术圈

拿到赛题，把这些数据集保存数据集。

哪些是类别特征、数值特征

类别特征（onehot）

类别特征与数值特征

回归题目：

天池不是特别严

Datafountain,cn/competitions

Ai指北

Kesai.com/home/competition

没有严格反作弊

Kaggle老师的网址：Github.com/PandasCute

个别比较喜欢dcjingsai.com

少参加付费赛

Dc平台很坑：只显示第一和第二

用excel提交的记录，慎重，不能用同一台机器登录几个号

Tione/百度

新手：二分类

手写字体识别，花的识别

用linux跑，是机器学习，数据挖掘。

Windows可以做linux

传统机器学习做时序

自己装一个linux虚拟机？？？？

XGBoost.

买阿里云？

推荐三本书

1西瓜书 《机器学习》，周志华

3数据挖掘导论，很适合新手

3统计学习方法 第二版 ，李航

服务费、薅羊毛？

SVM 适合小数据集。

链接：https://pan.baidu.com/s/10zSEzjfrb58qaaKDQsohhA&shfl=sharepset 提取码：l975 复制这段内容后打开百度网盘手机App，操作更方便哦

印象笔记

神经网络：加了隐藏层，容易过拟合

传统机器学习的方法？

Xgboost 的安装路径

<https://www.jianshu.com/p/d9df1b3bd259>

conda3

**利用conda新建环境**

conda create -n kaggle\_env python=3.6

**在conda创建的环境中安装xgboost**

pip install xgboost

**在jupyter notebook 添加新的虚拟环境**

conda install nb\_conda // jupyter notebook 可以检测到 新的虚拟环境