关于一些前置知识，包括什么是拉普拉斯矩阵，傅立叶变换、香农公式，以及如何安装ptorch框架，之前已经讲过了，有兴趣可以看之前整理的文档。这次时间关系就只讲如何用GCN实现反洗钱功能。

#### 数据集获取

什么是数据集？

我们训练和测试需要用现有的，带有标记的数据；这样的数据集合就是数据集。

最开始打算用IBM沃森实验室的反欺诈数据，但在IBM官网虽然能找到反欺诈业务，但并没有发现相关数据集。

但是我后面找到IBM沃森实验室的一篇论文《Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks for Financial Forensics》，采用的是Elliptic提供的比特币交易数据集，IBM在这篇论文中，使用GCN使用这个数据集进行反洗钱的研究。

由于关于反洗钱相关的数据集，我找了阿里、Kaggle等数据集，并没有找到更合适的数据集，所以本次采用Elliptic提供的数据集。

#### 2.数据集介绍

Elliptic的数据集可以通过阿里或者kaggle下载数据集

阿里下载地址：<https://tianchi.aliyun.com/dataset/dataDetail?dataId=110892>

总共23条交易数据

分为三个文件

Class、feature、edge

Feature保存每一笔交易特征属性，edge保存交易和交易之间的关系，class保存交易合法、非法或未知属性。

图卷积神经网络中需要的图的节点和边数据。

我们将以每一笔交易作为图卷积神经网络的节点，而交易和交易之间的流向作为节点之间的边。

有49个时间步长，我们采用前面34个用于训练，后面的时间步长用于测试。

我们通过下载Csv文件，然后通过读取csv加载本地数据集。

当然也可以选择接入kaggle，通过网络获取数据集。

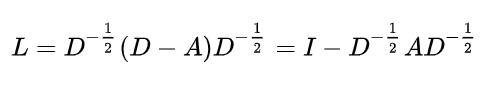
#### 3.准备工具

除了上次之前分享提到的需要安装的工具外，还需要安装Geometric

安装教程： <https://blog.csdn.net/qq_42965316/article/details/122731611>

#### 4.GCN公式推导

上次我们介绍了GCN矩阵的建立，和最终产生归一化的拉普拉斯矩阵

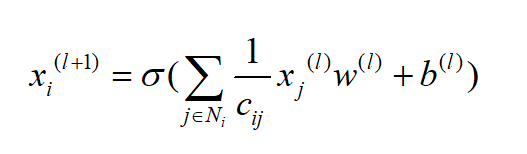


为了方便计算，我们需要进一步推导

推导过程：

<https://www.cnblogs.com/denny402/p/10917820.html>

最终推到结果



如何用代码实现GCN公式：

<https://samuel92.blog.csdn.net/article/details/105505642>

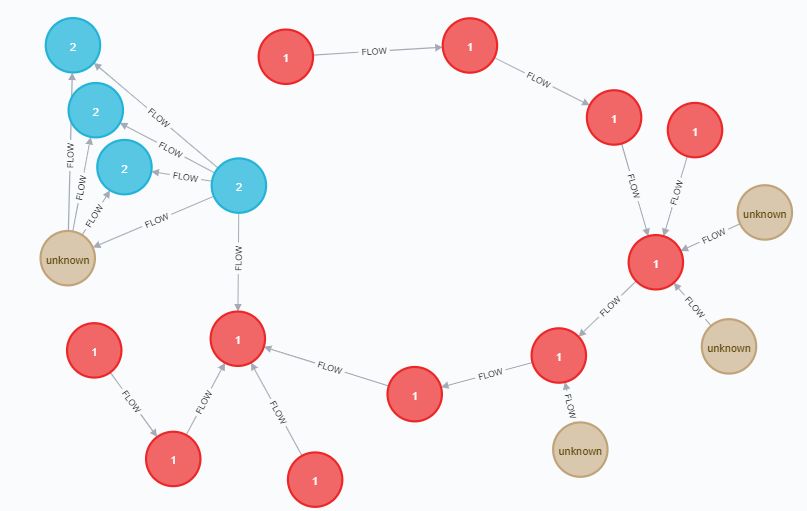
当然，我们必须要自己手写，因为官方已经封装好了，直接调用GCNconv就行了。

5.反洗钱原理

洗钱是通过处置、离析（来回转账）、归并的方式。

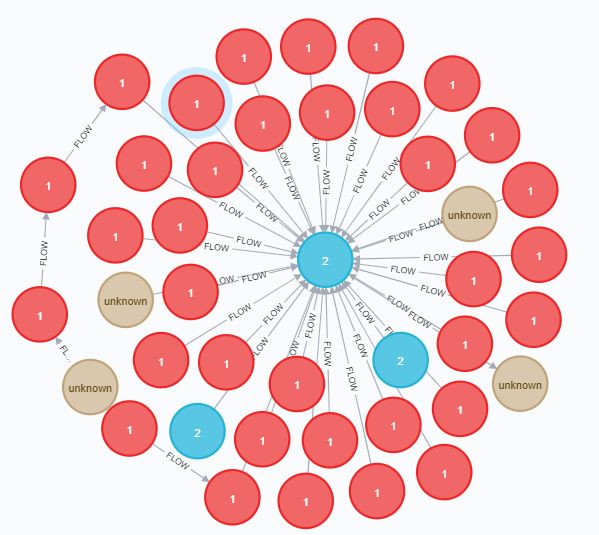
反洗钱是通过发现异常网络结构来预警

* 频繁汇入/汇出
* 链式交易结构
* 集中转入/转出
* 分散转入/集中转出
* 环状交易结构
* 其他异常交易结构

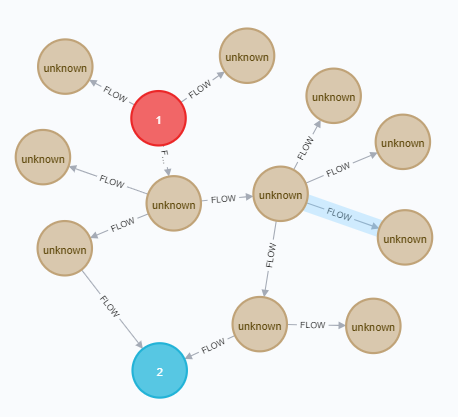


链式交易结构

节点都是单向转出，追溯交易链条发现洗钱。



非法集中到合法



非法集中转出

节点用于过度，转入的都会转出，通过计算节点折算率判断是否洗钱节点。

#### 6.代码展示

通过两个维度衡量算法

* 精确度，识别的正确率
* 漏报率，表示多少数据被遗漏

各种指标背后的意义：

<https://blog.csdn.net/Cxiazaiyu/article/details/81878556>

目前写出的最初版本，达到保证精确度90%的情况下，漏报率50%

#### 7.展望

目前实现的是初级版本，后面考虑到优化，会考虑使用时间序列等算法结合图卷积神经网络，达到更高的准确率和识别率。