反洗钱社团发现模型：样本详细需求

（注：本模型主要用于发现地下钱庄或非法集资！）

**一、构造金融交易网络**

1. 选取5个地下钱庄和5个非法集资银行卡（见第二部分）：

**风控中心给的15个可疑地下钱庄或非法集资银行卡**

**账号1：账号2：账号3：账号4：账号5：账号6：账号7：账号8：账号9：账号10：**

2. 选取100个未出现在可疑样本中的正常银行卡；

**注：能否判断这张卡是个人持有还是商户持有的？**

3. 根据以上银行卡，采用关联交易（只要有相互转账，就有关联交易），拓展出其它银行卡；

4. 满足以下两个条件中的一个条件，就停止拓展：

（1）满足1年内关联交易的银行卡都提取出来了，此时银行卡总的数目为m；

（2）银行卡总数目达到了1000个，假设此时计算的总天数为D’；

5. 将任何一张与它有交易关联的银行卡标记为一个类，记录下来；

6. 假设达到以上条件时，银行卡总数目为N，总天数为D；现在随机打乱这N张银行卡，然后进行排序：序号从1到N；根据这N张卡之间的相互交易关系，我们可以得到一张金融交易网络；对于交易天数，从实际发生的时间先后关系进行排序：序号从1到D；

7. 需要记录的数据：

（1）银行卡号与编号的一一对应文件

对这N张银行卡号进行编号，然后需要指出它们之间的一一对应关系；

（2）交易时间与编号的一一对应文件

对这D天按照时间先后顺序进行排序，然后需要指出它们之间的一一对应关系；

（3）记录任意一张银行卡每天的详细交易数据

以1天为一个时间单位，记录任意一张银行卡i的交易数据：

转入数据：

1. 银行卡k1向i转入的每一笔金额、转入日期、交易地点；

2. 银行卡k2向i转入的每一笔金额、转入日期、交易地点；

如此，等等

转出数据：

1. 银行卡i向s1转出的每一笔金额、转入日期、交易地点；

2. 银行卡i向s2转出的每一笔金额、转入日期、交易地点；

如此，等等

分别记录D天的数据；

8. 记录格式：

总共四个文件：

（1）记录银行卡号与编号的一一对应文件

它由两个字段组成：“编号”和“银行卡号”，每一行代表一个完整记录；

（2）记录银行卡号与编号的一一对应文件

它由两个字段组成：“编号”和“日期”，每一行代表一个完整记录；

（3）记录任意一张银行卡的每天转入转出交易数据文件

**文件1记录转入交易：**

它由六个字段组成：“银行卡编号”、“银行卡编号”、“转入金额”、“转入时间编号”、“交易地点”。一行记录一笔交易，形式如下：

银行卡编号（接受转账） 银行卡编号（发起） 转入金额 转入日期编号 交易地点

I1 k1 x1 D1 add1

I1 k2 x2 D1 add2

I2 k3 x3 D1 add

I2 k4 x4 D1 add

注：第一列和第二列都按照从小到大排序；而且，第二列向第一列转账，也就是说任意一张银行卡，记录有哪些银行卡向该银行卡转账了，以及转账的金额、日期；如果有多笔转账，则分别记录。

**文件2记录转出记录：**

它由六个字段组成：“银行卡编号”、“银行卡编号”、“转出金额”、“转出时间编号”、“交易地点”； 一行记录一笔交易，形式如下：

银行卡编号（发起转账） 银行卡编号（接受） 转出金额 转出日期编号 交易地点

I1 k1 x1 D1 add1

I1 k2 x2 D1 add2

I2 k3 x3 D1 add

I2 k4 x4 D1 add

注：第一列和第二列都按照从小到大排序；而且，第一列向第二列转账，也就是说任意一张银行卡，记录该银行卡向哪些银行卡转账了，以及转账金额、日期；如果有多笔转账，则分别记录。