关系网络算法步骤总结

2019年4月2日星期二

1. 预处理原始文件
   1. 将人-物的原始文件的人节点打上首逾超过30天的黑名单标签。输出CSV共有五列，分别是source\_id（人ID），relation\_type，destination\_id（物ID），destination\_type，source\_label
   2. 选取黑名单文件中前500个人节点ID作为考察的root\_nodes
   3. 输出relation\_type的权重映射表，重要关系为2，轻度关系为1，其他关系无
   4. 输出destination\_type的权重映射表，Identity=0.4, BankCard=0.2, PlateNum=0.2, Mobile=0.1, Device=0.1
2. 人-物关系合并
   1. 将连接人ID和物ID的多条relation边（例如，人ID-AuthMobile-物ID和人ID-BindMobile-物ID），按照relation\_type的权重映射表直接加在一起，作为人ID和物ID的唯一边的权重
   2. 删去权重为0的边
3. 转化为人-人关系
   1. 对所有的物ID，遍历其连接的所有的人ID，用组合将人ID两两相连
   2. 连接过程中除去无法连接到第二个人ID的边
   3. 将物ID两侧的边的权重用调和平均汇总在一起
4. 人-人关系同种类物品合并
   1. 将连接人ID和人ID的多条同种物品类型的边合并在一起，权重直接相加
   2. 将权重归一化
5. 人-人关系异种物品合并
   1. 将连接人ID和人ID的多条异种物品类型的边合并在一起，权重按照destination\_type的权重映射表加权平均
   2. 将整个网络的权重归一化
6. 写出人-人的边表
   1. 输出CSV共有五列，source\_id, source\_label, destination\_id, destination\_label, regulated\_total\_relation\_weight
7. 构建网络
8. 输出从root\_node出发，一度（二度，三度）（仅为该度数不含前度数）的黑节点数量（权重）和所有节点的数量（权重）
9. 加工成最终指标，输出CSV，共有6个指标，每个root\_node为一行数据。指标计算公式为：

其中，r为最终的指标，该指标综合考虑了数量和占比两个因素。n是在一度关系中（或加上二度，或再加上三度）的黑名单人数（或者是黑名单权重），m是对应度数的总人数（或者总权重）。6个指标具体为：

* 1. 一度关系中，首逾超过30天黑名单人数指标
  2. 一度及二度关系中，首逾超过30天黑名单人数指标
  3. 一度、二度及三度关系中，首逾超过30天黑名单人数指标
  4. 一度关系中，与首逾超过30天黑名单权重指标
  5. 一度及二度关系中，与首逾超过30天黑名单权重指标
  6. 一度、二度及三度关系中，与首逾超过30天黑名单权重指标