

# 关联关系图谱技术在团伙欺诈识别中的应用

陈俊, 乔辉

江苏银行股份有限公司风险管理部, 江苏 南京  
Email: xaviercj@163.com, qiaohui@jsbchina.cn

收稿日期: 2020年10月6日; 录用日期: 2020年10月21日; 发布日期: 2020年10月28日

---

## 摘要

银行现有的反欺诈体系主要以专家规则为主, 在防范新型的欺诈手段和欺诈形势时, 存在两大痛点: 一是无法识别个体正常但属于团伙欺诈的情形; 二是缺乏对风险传导的防控。基于此, 本文开展基于关联图谱的反欺诈体系研究, 从关联图谱的底层逻辑、构建原理、方法路径以及应用价值等方面阐述关联图谱反欺诈体系的底层逻辑和应用价值, 并进一步探索在营销精准化的应用前景。

## 关键词

关联图谱, 团伙欺诈, 团伙聚类, 精准营销

---

# Application of Association Relation Graph Technology in Group Fraud Recognition

Jun Chen, Hui Qiao

Risk Management Department of Bank of Jiangsu, Nanjing Jiangsu  
Email: xaviercj@163.com, qiaohui@jsbchina.cn

Received: Oct. 6<sup>th</sup>, 2020; accepted: Oct. 21<sup>st</sup>, 2020; published: Oct. 28<sup>th</sup>, 2020

---

## Abstract

The bank's existing anti-fraud system is mainly based on expert rules. There are two major pain points in preventing new fraud methods and fraud situations: one is the failure to identify individuals that are normal but belong to group fraud; the other is the lack of risk transmission in prevention and control. Based on this, this article carries out research on the anti-fraud system based on the correlation graph, expounds the underlying logic and application logic value of the anti-fraud system of the correlation graph from the aspects of the underlying logic, construction principle, method path and application value of the correlation graph, and further explores the precision in marketing the application prospects of modernization.

## Keywords

Association Graph, Group Fraud, Group Clustering, Precision Marketing

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

互联网、物联网、大数据的兴起和普及,正深刻改变着当前的金融生态和金融格局。在此宏观背景下,银行为了提高自身竞争力,降低运营成本,提高客户满意度,逐步向互联网业务转型。同时,当下的新冠肺炎疫情限制了原有的线下业务渠道,这也进一步加快了银行网上业务的转型步伐。虽然银行互联网业务具有收益高、成本低、体验好、方便快捷等优势,但是该业务因具有“无接触”的特点也导致客户欺诈风险的激增,这无疑对银行互联网业务的发展带来巨大的挑战。

互联网背景下的金融欺诈形势逐渐以“黑产”的组织形式开展,威胁着金融安全,对金融主客体造成巨大的损失。数据显示,国内因倒卖、遗失等原因而游离在市场上的身份证约 1000 万张,“三件套”、“四件套”(身份证、手机卡、银行卡、网银盾)市场倒卖报价百元到千元不等,“黑产”每年给社会造成的损失超过 1000 亿。除此之外,“黑产”组织还掌握了超过 5000 万的虚拟号码、6000 万的代理 IP、以及市场上 83% 的设备信息,这些都是“黑产”组织进行欺诈的重要基础[1]。

面对如此威胁,以银行为首的金融机构并未坐以待毙,为了打击欺诈行为,他们积极推进以规则为主的反欺诈体系,并在反欺诈防范中起到一定的成效,以 X 银行为例,2019 年 X 银行基于反欺诈规则,有效拦截了 3.4% 的异常信息申请,发挥了有效初筛准入作用。但是,在防范新型的欺诈手段和欺诈形势时,以规则为主的反欺诈体系存在两大痛点:一是无法识别个体正常但属于团伙欺诈的情形;二是缺乏对风险传导的防控。而关联关系图谱(以下简称“关联图谱”)则可以通过图分析技术深度挖掘客户潜在风险,能够有效解决现有反欺诈规则的不足。

## 2. 关联图谱技术识别欺诈风险的底层逻辑

关联图谱的本质是整合客户的申请、交易、还款等信息,构建手机号、设备指纹、工作单位、pos 机编号、账号等要素共现的拓补网络。在反欺诈场景中,除了考虑单一信息点的属性,客户间的隐藏关联往往包含更多未知的潜在价值信息。因此,客户信贷欺诈行为的识别问题可以转化为关联图谱挖掘或社交网络分析问题。基于关联图谱的反欺诈体系旨在将多源异构的数据整合成机器可以理解的知识,将单点的信息转化为平面的相互关联的图谱,进行异常风险子图的监测,从而实现欺诈的识别与防御[2]。

通过关联图谱进行反欺诈的最大优势在于:识别“个体正常”但属于欺诈的团伙,这也是目前银行进行反欺诈防范的最大痛点。基于关联关系分析,可以将相同特征的人聚成群组,从而做批量分析和特征学习。在此基础上,进一步根据群组特征分析与挖掘,可以对特定人群制定特定的反欺诈策略[3]。

## 3. 关联图谱技术识别欺诈团伙的方法路径

### 3.1. 关联图谱的构建

关联图谱建设的关键是基于手机号、设备指纹、工作单位、pos 机编号、账号等节点,确立关系类型,

并进一步实现要素共现的拓扑网络图。关系类型包括但不限于：

- 1) 亲属关系：两个客户家庭地址一致或申请信息直接明确为申请人家属的即可认定为亲属关系；
- 2) 同事关系：两个客户通过公司名称构成关联，如果公司名称一致即可认定为同事关系；
- 3) 手机号关联：通过手机号将客户关联，包括共用手机号码、紧急联系人号码等；
- 4) 设备、IP 关联：通过相同的设备指纹或 IP 关联不同的申请客户，形成同设备、同 IP 的关联关系；
- 5) 资金交易关系：通过转账行为关联两个客户；
- 6) 其他关联：社交关系、朋友关系等，甄别手段包括邮件往来、通讯聊天等。

构建的关联图谱示意图如图 1 所示。



Figure 1. Schematic diagram of association graph  
图 1. 关联图谱示意图

### 3.2. 基于强联通子图和 LPA 算法的团伙聚类

关联图谱分析的核心是从图的角度识别网络中的各个团体。本文基于构建的零售关联全局网络图谱，并通过强联通子图和 LPA 两种社团分割算法，聚类网络中的各个团体，实现路径如图 2 所示。



Figure 2. Schematic diagram of group clustering  
图 2. 团伙聚类示意图

本研究使用的社团分割算法包括强联通子图和 LPA 两种算法, 具体如下:

**强联通子图算法:** 如果两个顶点间至少存在一条路径, 即可称为两个顶点强联通。进一步的, 如果对于网络中的一个子图, 每两个顶点都至少有一条路径, 即认为这个子图为强联通子图。如图 3 所示, 节点 1、2、3、4 构成强联通子图, 但节点 5、6 并未包含在该联通子图中。

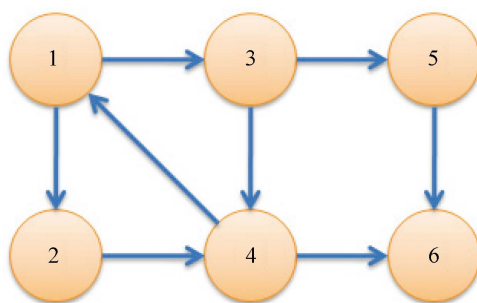


Figure 3. Schematic diagram of unicom submap  
图 3. 联通子图示意图

**LPA 算法:** 即标签传播算法, 该算法首先对关联图谱中每个节点定义自己特有的标签, 在迭代过程中通过标签不断扩散传播, 标签传播的逻辑为: 某一节点选择自己邻居中出现次数最多的标签, 如果每个邻居标签的次数一样多, 就会随机选择一个邻居标签作为自己的标签。如此往复, 直至满足收敛条件为止, 即每个节点的标签不再发生变化, 那么此时将相同标签的节点归为一个社区, 具体传播过程如图 4 所示。

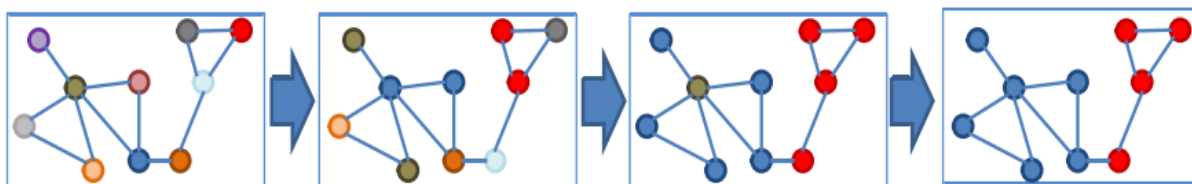


Figure 4. Schematic diagram of LPA  
图 4. LPA 示意图

**基于强联通子图和 LPA 算法识别团伙的逻辑:** 基于大零售客户数据构建的全局关联图谱, 首先通过强联通子图算法排除孤立点或者弱联通关系的节点, 再通过 LPA 标签传播算法对特征相同的节点进行聚类, 最终形成各个团伙。

### 3.3. 申请团伙风险特征挖掘

在上一步完成团伙聚类的基础上, 进一步对团伙的风险特征进行深度挖掘, 从而识别欺诈风险高的团伙, 其分析步骤如下:

- 1) 梳理能够反映团伙特征的长清单变量, 变量至少包括: 设备数量、团伙规模、贷款余额、城市分布、手机号码数量、公司数量、逾期金额等。
- 2) 进行单因素分析, 主要通过 IV 值筛选出对团伙欺诈具有预测能力的团伙特征指标。
- 3) 将通过 IV 值筛选出的特征指标, 进一步利用决策树算法生产规则, 挖掘风险团伙。
- 4) 在挖掘出风险团伙的基础上, 进一步利用评分卡模型, 对挖掘的团伙进行风险评分, 预测团伙欺诈的概率, 将欺诈概率高的团伙识别为高危风险团伙。

## 4. 关联图谱技术的应用价值

关联图谱技术充分发挥其直观化、效率化的图技术优势, 在反欺诈尤其是团伙反欺诈方面的应用成效显著, 具体表现为以下四个方面:

### 4.1. 抓取高危风险团伙, 避免潜在损失

当前银行以规则反欺诈体系为主, 主要用于识别单一客户的真实性, 但对单一客户正常但属于团伙欺诈情形的识别性不足。而基于关联关系图谱, 通过风险团伙识别、传播、评价等模型能够有效挖掘异常模式子图, 有效识别高危风险团伙, 避免潜在损失。

### 4.2. 以图的方式直观展示客户的关联关系, 提高决策效率

基于底层构建的关联图谱, 以直观的方式在可视化工作平台展示客户的关联关系, 便于审批人员快速处理复杂网络, 提高决策效率。

### 4.3. 实现了全行零售业务风险客户名单、风险策略与模型的共享机制

关联图谱在实现业务数据打通的基础上, 进一步实现了各部门的反欺诈策略与模型的共享, 打破了原先反欺诈风险防范时部门间的信息壁垒, 丰富了欺诈防范的手段与技术。其次, 对不同条线确定的风险客户名单, 同样实现共享, 有利于跨部门、跨产品识别客户的风险水平。

### 4.4. 弥补现有规则策略和有监督学习在欺诈识别中的不足

图数据挖掘算法有大量分析方法和图算法, 这些算法并不依赖标签数据, 从图的自身结构中学习图的特征, 应用于异常结构子图的检测。同时图的半监督学习算法可以利用少量欺诈节点标签, 结合图的关系结构信息, 概率推断其他节点实体的欺诈概率。因此, 基于关联图谱的反欺诈手段作为规则策略和有监督学习的补足方案, 有效弥补了原有反欺诈方法的不足。

## 5. 关联图谱技术的应用展望

目前, 项目成果主要应用在防范欺诈风险方面, 并且初显成效。而结合关联图谱的特点和优势, 下一步拓展关联图谱在精准化营销方面的应用。

### 5.1. 深度挖掘营销客群的特征

欺诈是为了识别“坏客户”, 而营销是为了识别“好客户”, 其本质是相通的。实现精准化营销的重点突破口在于围绕“好用户”或者“高价值用户”的数据采集和特征工程开发领域, 在各类不同营销场景下挖掘典型目标客户的各类基本属性和行为特征, 并以此作为关联分析、聚类的依据, 找到和他们相似的客群, 为精准化的营销活动提供依据, 最大化营销费用的投入产出比。

### 5.2. 发展核心人物的自营营销

利用关联图谱技术的特点和优势, 识别团伙中的核心人物。在此基础上, 向识别出的核心人物推送团体成员名单, 由团体核心人物在团体成员间自行营销, 根据营销成功的客户数量等指标给予相应的奖励。

### 5.3. 次优人群的精准营销

目前银行获客成本越来越高, 获客难度也越来越大, 而在风险可控的前提下, 发展存量客群中的次

优人群则具有可行性。利用关联图谱的特点和优势,可以对符合次优特点的客户群体进行精准识别,其方式是:基于行内的优质种子客群,识别出与优质客群具有强关联关系的次优客群,对这部分次优人群进行进一步营销。

### 参考文献

- [1] 王成,舒鹏飞. WEB:一种基于网络嵌入的互联网借贷欺诈预测方法[J]. 大数据, 2019, 5(6): 1-16.
- [2] 金磐石,万光明,沈丽忠. 基于知识图谱的小微企业贷款申请反欺诈方案[J]. 大数据, 2019, 5(4): 100-112.
- [3] 谭波,滕光进,王浩. 基于大数据的客户关联关系及风险预警研究[J]. 清华金融评论, 2017(8): 19-21.