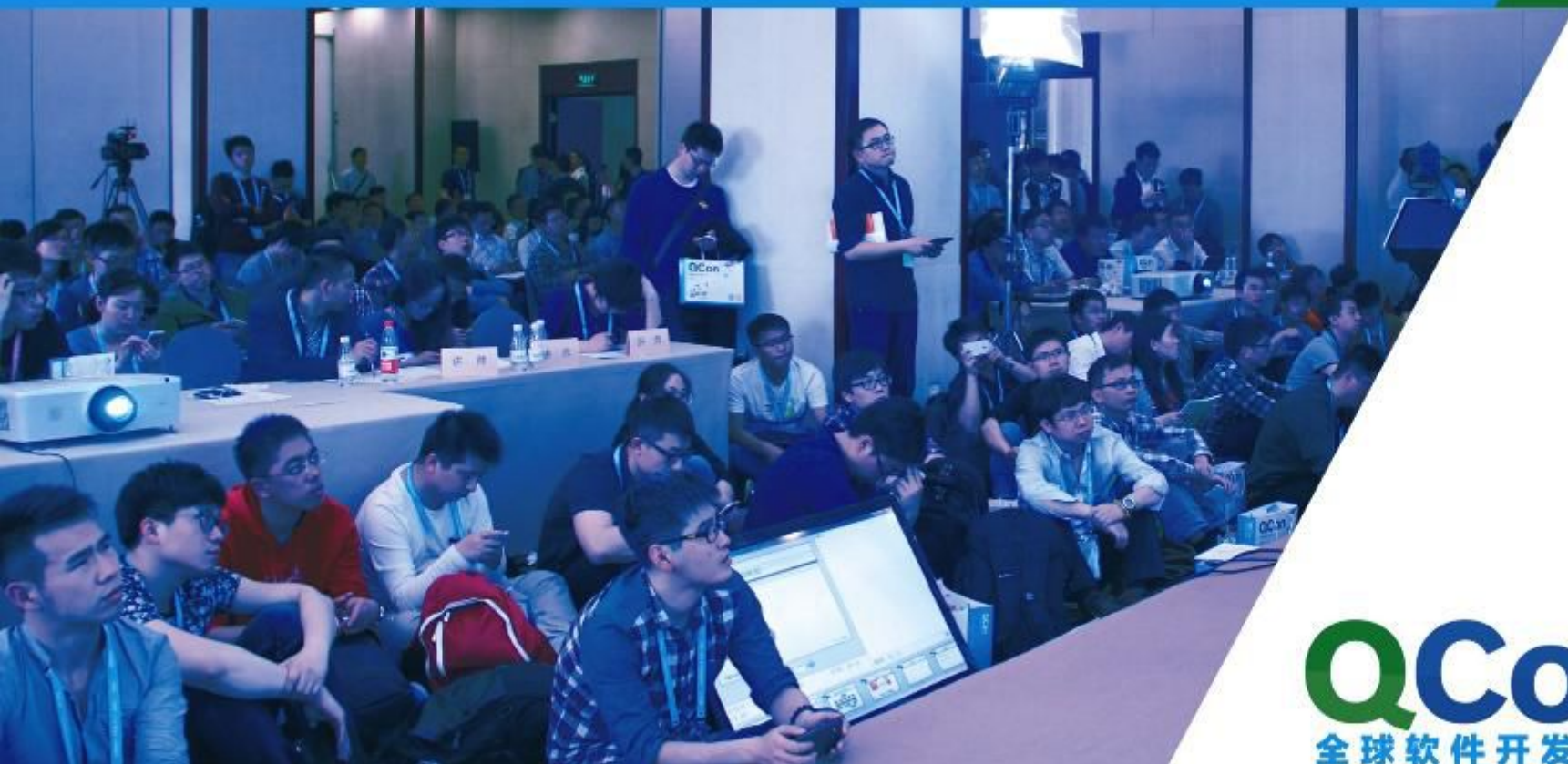


QCon全球软件开发大会

International Software Development Conference



QCon
全球软件开发大会

Geekbang>

极客邦科技

全球领先的技术人学习和交流平台

扫我，码上开启新世界



Geekbang>

InfoQ | EGO NETWORKS | StuQ

InfoQ

专注中高端技术
人员的社区媒体

EGO NETWORKS

EXTRA GEEKS' ORGANIZATION
高端技术人员
学习型社交网络

StuQ

实践驱动的IT职业
学习和服务平台



促进软件开发领域知识与创新的传播



实践第一 案例为主

时间：2015年12月18-19日 / 地点：北京·国际会议中心

欢迎您参加ArchSummit北京2015, 技术因你而不同



ArchSummit北京二维码



[北京站]

2016年04月21日-23日



关注InfoQ官方信息
及时获取QCon演讲视频信息

点融网大数据技术反欺诈实践

点融网高级数据开发工程师 程书欣

个人简介

- 点融网高级数据开发工程师
- 2013年10月份加入仍然是初创公司的点融网
- 2014年下半年开始参与研发点融网反欺诈系统与构建大数据基础架构，以应对点融网面对的风控技术挑战

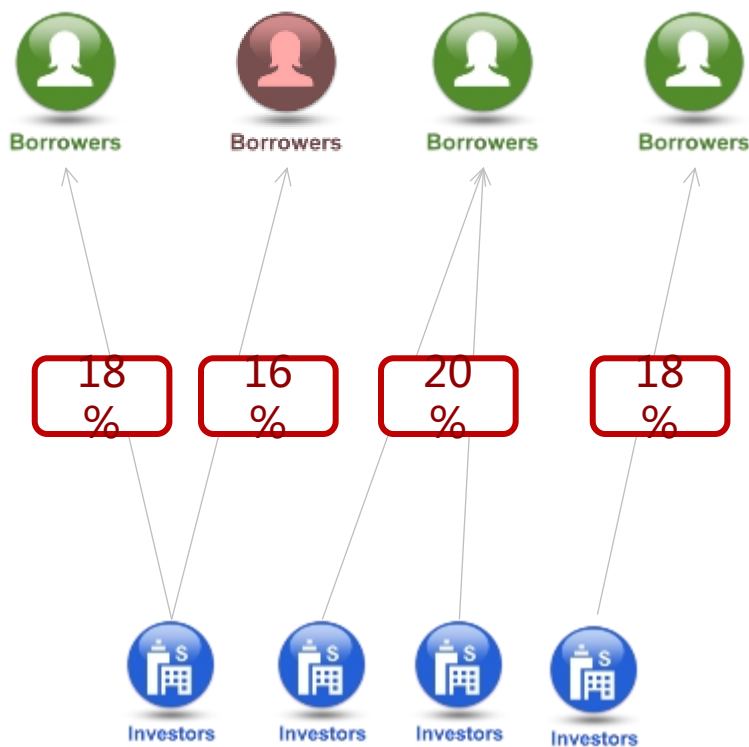
背景介绍 – 点融网

- 点融网 – dianrong.com
- 2013年成立
- 2014年6月，推出明星级产品“团团赚”
- 2015年8月，获得渣打银行和中国互联网金融科技基金领投的2.07亿美元C轮融资

Agenda

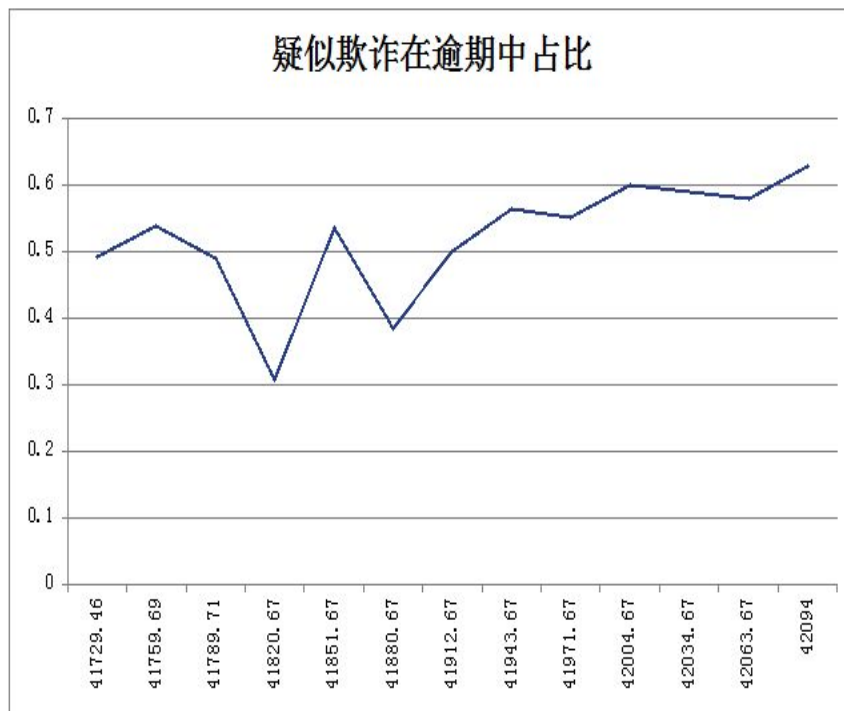
- P2P领域反欺诈的问题与挑战
- 知识图谱在反欺诈领域的应用与实践
- 反欺诈系统的技术架构

P2P存在的风险



风控和反欺诈现状

- 欺诈造成的损失



借款端欺诈频发

- 贷款资料造假
 - 伪造银行流水，收入信息
 - 工作单位涉嫌欺诈
 - 身份资料造假
- 中介代办包装
 - 中介对不符资质的借款人的资料进行包装，以获得贷款
- 组团骗贷
 - 多人联合欺诈

车贷 房贷
月息1分
15026631586

正规贷款,利息低,无抵押,当天下款 
专业包装银行房屋抵押贷款 备用房包装 贷款用途包装 30-3... (今...)
天宝投资担保  在线   1年

曹经理: 各种贷款,快速通过,快速放款

各种疑难贷款,快速审批通过,快速放款,欢迎来电咨询曹经理1861662... (06-13)
上海洲宁房地产投资咨询有限公司 

唐先生: 个人,企业贷款服务站,一站式包装

上海圣唐投资管理有限公司作为一家专业投资咨询服务公司,有着规范的专业知... (...)
上海圣唐投资管理有限公司  在线   4年

何海军: 现在贷款包装可以帮您操作了,速度快额度大

本公司专注于中小微企业贷款,额度1-55万不等,只... (01-04)
上海涌信金融信息服务有限公司 

反欺诈问题充满挑战

- 反欺诈是一个Negative goal
 - 核实身份证真假? Good to check, but not nearly enough
 - 必须检查所有欺诈方式的可能性
 - 必须不断更新反欺诈策略
- 怎样实施欺诈?
 - 伪造身份证去申请贷款
 - 伪造工作单位
 - 包装银行流水
 - 组团骗贷
 -

传统资料验证手段 - 身份证案例



传统验证手段

- 依赖于信息的人工审核，效率低
- 看
- 打电话

技术验证手段

- 复杂风控模型
 - 变量池
 - 单变量/多变量分析，筛选
 - 基于统计学建模，机器学习、逻辑回归.....
- 问题？
 - 缺乏表现力
 - 数据孤点

关联分析

- 全部的贷款中，拥有关联关系的贷款占总数的30%左右
- 约有四分之一的疑似欺诈贷款，构成相关关系

反欺诈问题是一个搜索问题

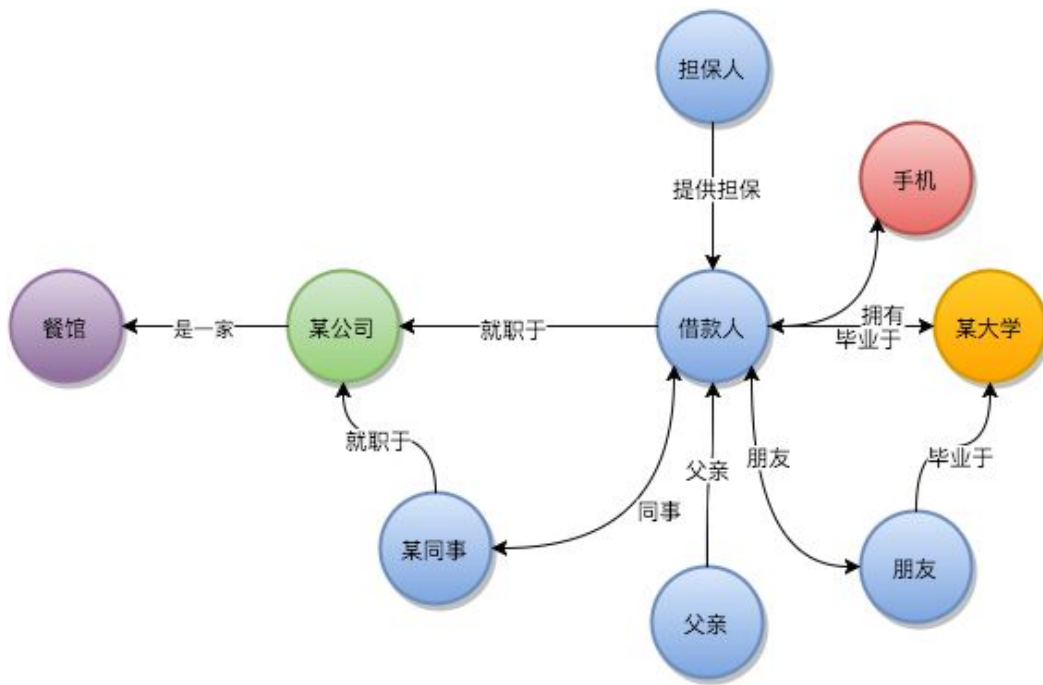
- 搜索什么？
 - 所有与借款人相关的数据
- 搜索哪些数据？
 - 点融网积累的历史数据
 - 数据提供商的数据
 - 互联网爬虫所得到的数据
- 如何更好地搜索？
 - 表结构无法很好的表达relation

为什么要用Graph

- 信息聚合， All data in one place
 - 借款人PII信息
 - 借款人的常用联系人，手机，工作单位和关系
 - 借款人的雇佣信息
 - 互联网上公开数据，例如公开黑名单、企业黄页等
 - 数据提供商的数据
- 强大的关联检索
- 机器可理解和推理

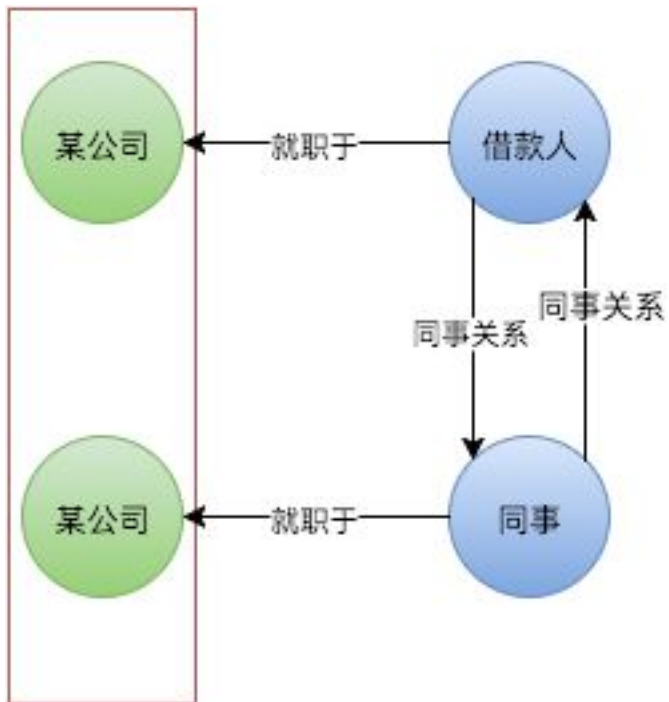
如何构建Graph

- 三元组（实体，谓词，实体）
- 实体deduction
- 关系deduction
- 推理（Inference）



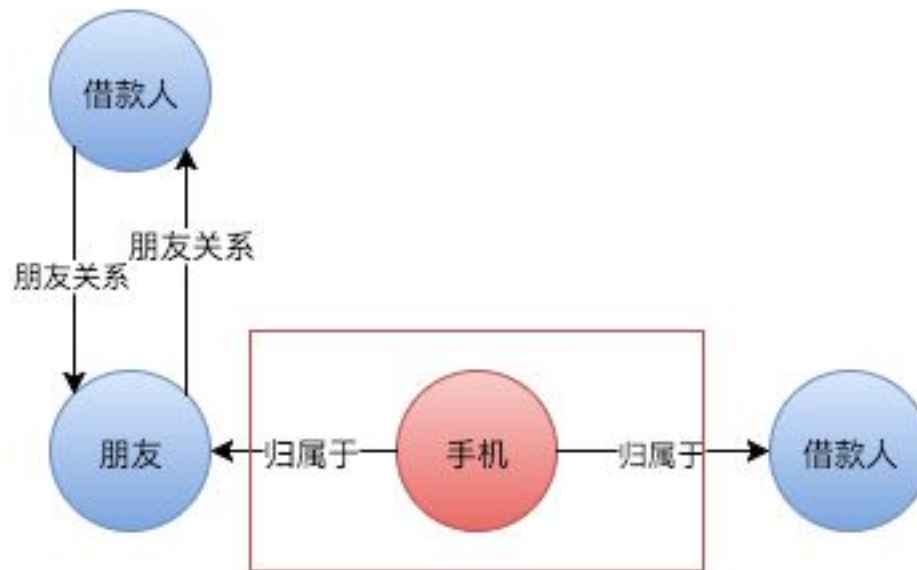
识别数据造假

- 就职公司矛盾



就职公司矛盾

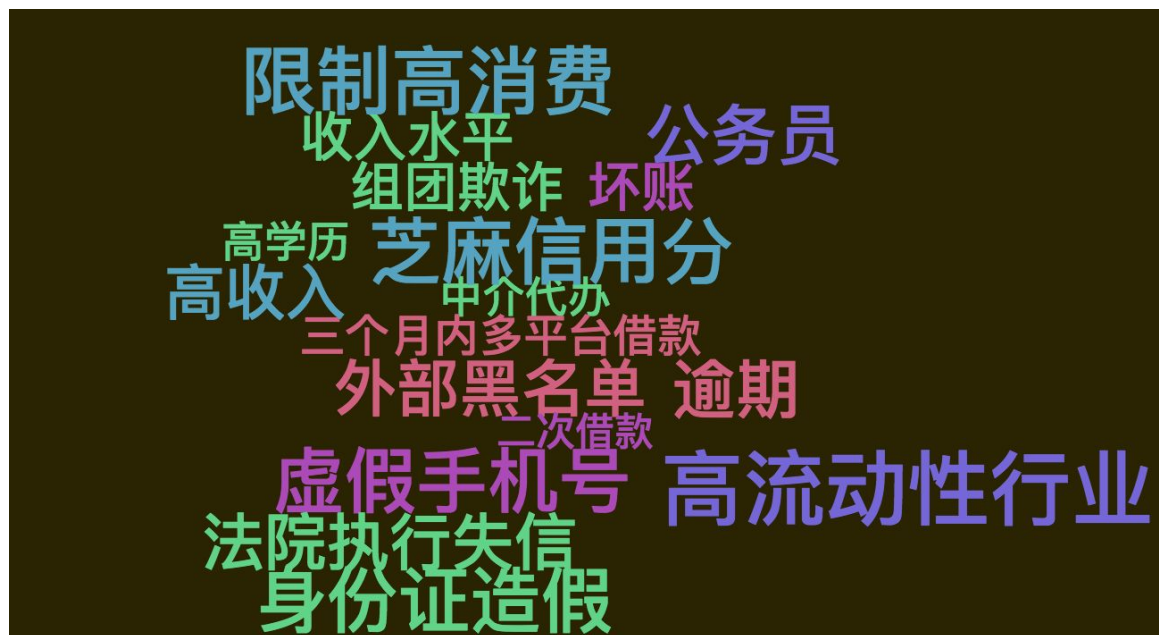
- 手机造假



同一手机归属于不同人

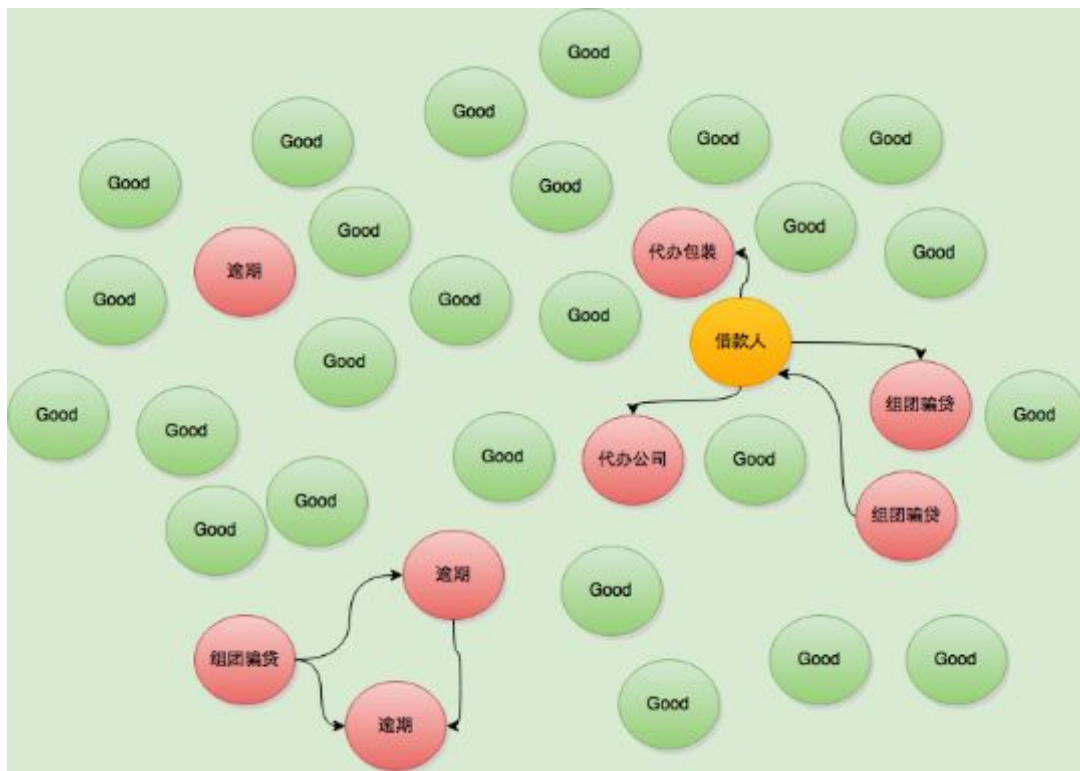
开发Tag以标记借款人特征

- 丰富数据维度
 - 对接多个数据提供商
 - 爬取互联网数据：公开黑名单，信息服务类网站，法院网
- 以Tag标记借款人特征，以便进一步分析



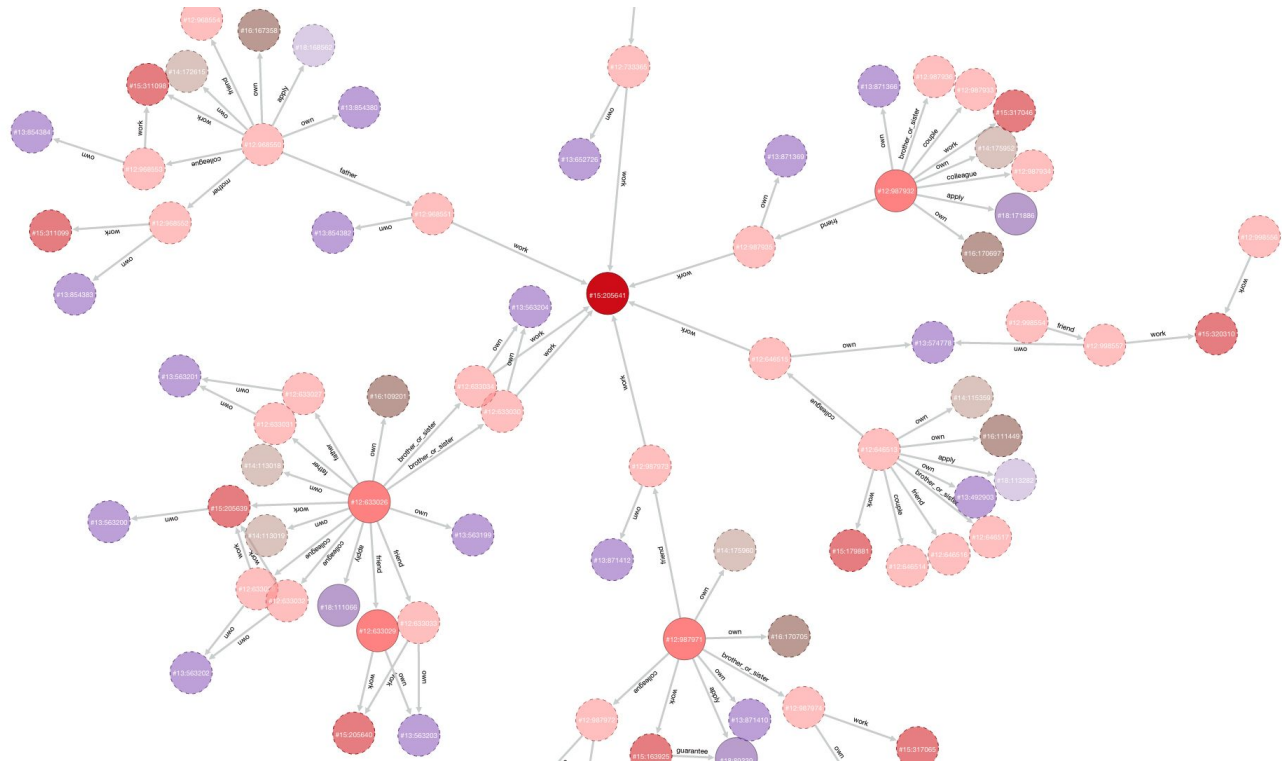
识别组团欺诈

- 关系挖掘
 - 挖掘借款人的潜在Tag的关系
 - 借款人的网络中可能存在中介
 - 借款人的网络中可能有命中诈骗Tag的人
- 挡不住第一单，可以挡住第二单、第三单



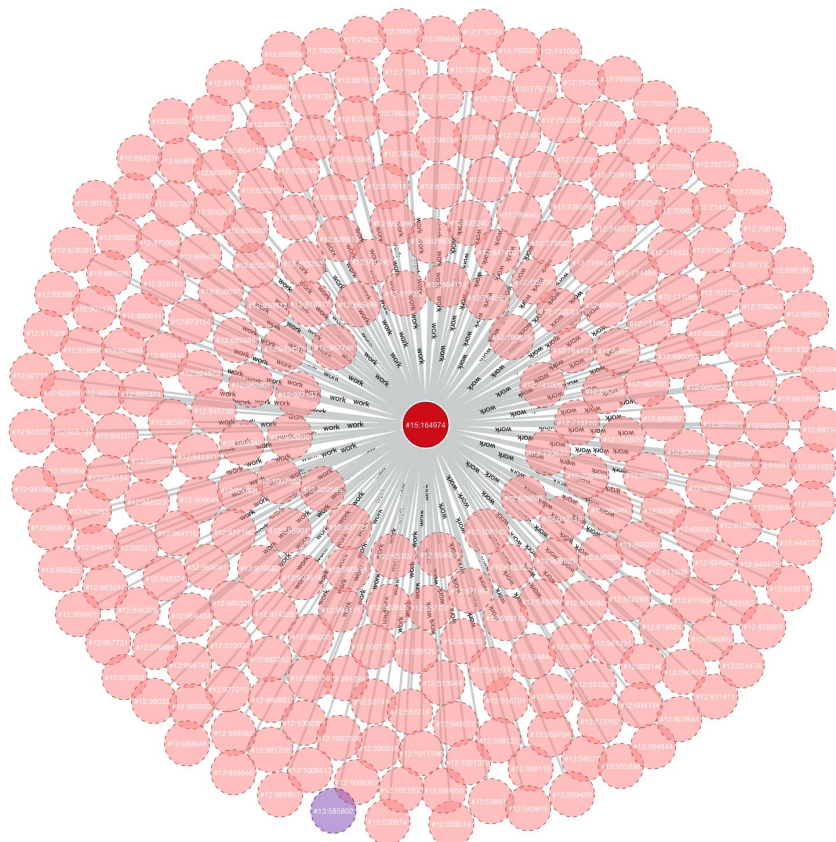
关系建模

- Loan
- Person
- Mobile
- Email
- Bankcard
- Company



问题与挑战

- 数据来源多，异构
- 数据清洗
- 节点合并
 - 实体识别
 - 相似实体合并
- 复杂的业务逻辑
- 模型的测试与部署

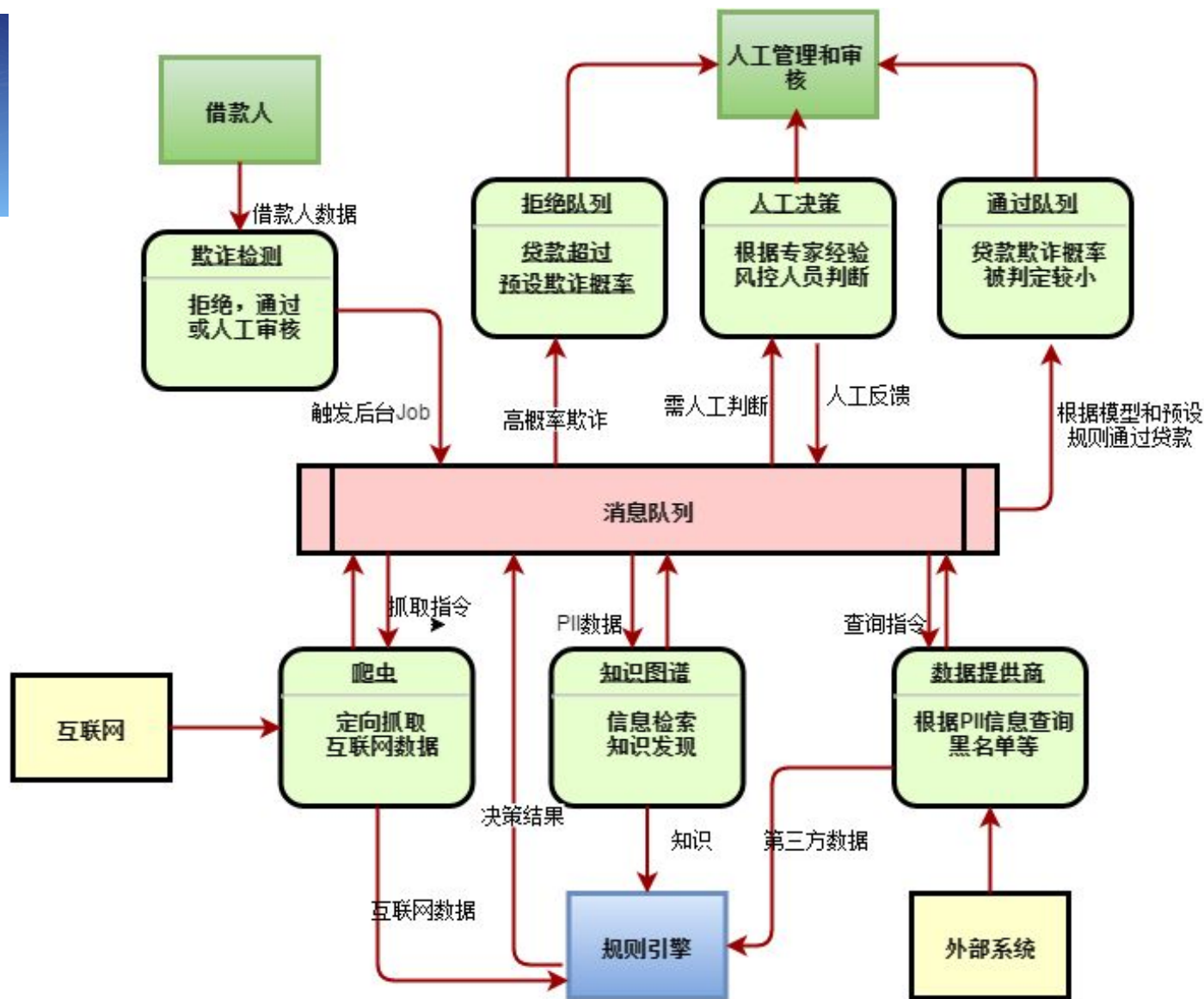


点融网反欺诈系统的技术架构

- Key use cases
 - 对接诸多第三方系统
 - 定向抓取
 - 决策模型复杂多变

点融网反欺诈系统的技术架构

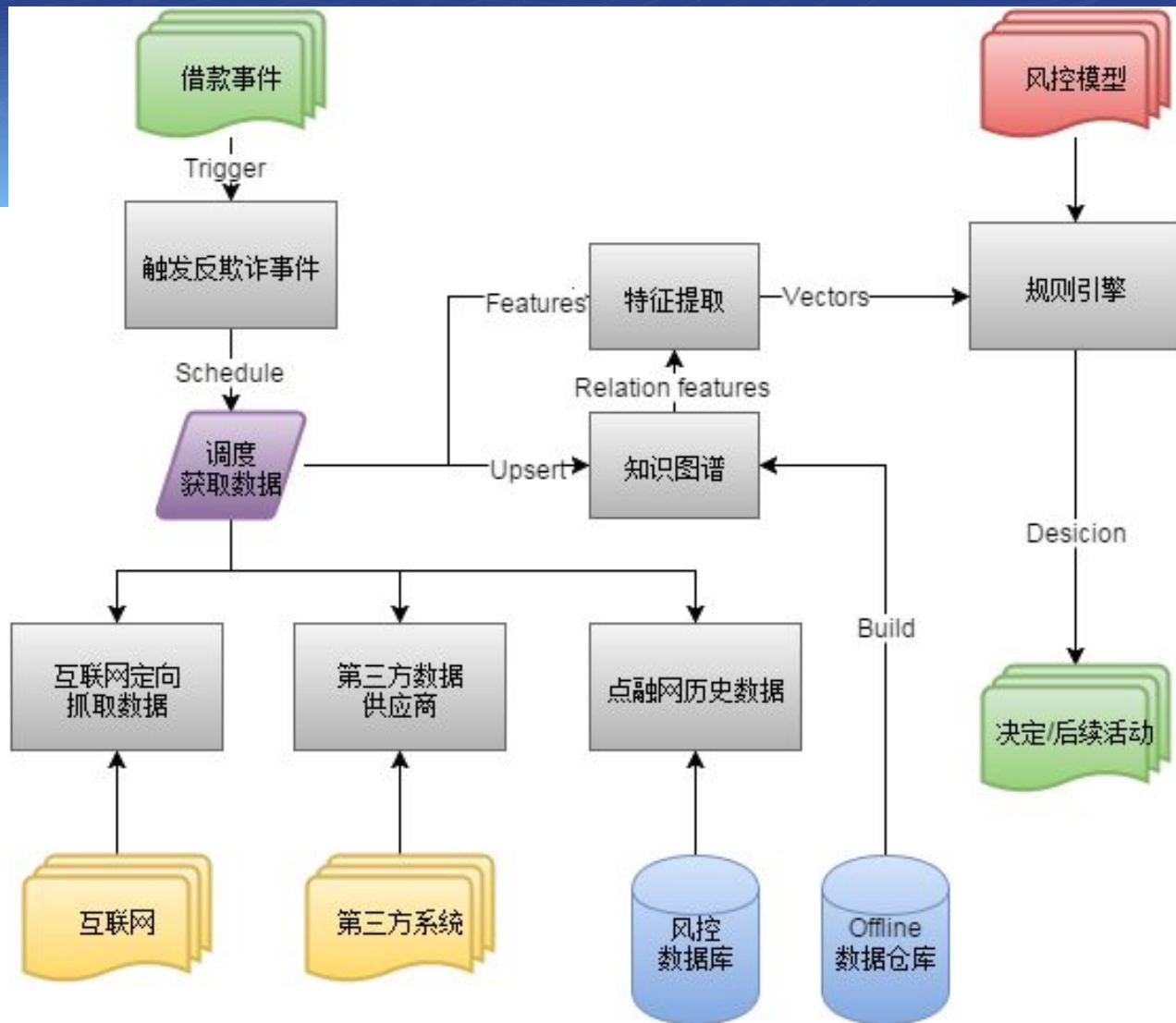
- 微服务架构
- 基于消息中间件的异步处理
 - Stateless system
 - Event idempotence
 - 监控
- 可灵活配置的规则引擎
 - Built-in策略
 - Groovy脚本支持



规则引擎

- Goal
 - 通过管理配置和扩展手段快速反应业务逻辑的变化
- Module
 - 规则元数据管理
 - 运行数据管理
 - 运行执行和规则策略
- Feature
 - 支持meta规则， built-in规则， 扩展规则
 - 支持Groovy脚本， 最大灵活程度
 - 历史验证数据， 便于审计

事件流



总结

- 知识图谱
 - 有效挖掘二维数据的潜在知识
 - 丰富数据表现力
 - 提供更有效的feature建模
 - 不仅仅是反欺诈

THANKS

Brought by **InfoQ**

International Software Development Conference