

因唯安全
所以信赖
唯与同行,智御未来

基于金融知识图谱的反欺诈

主讲人: 卢铮

唯品金融主任算法工程师

2018 唯品会第三届互联网电商安全峰会

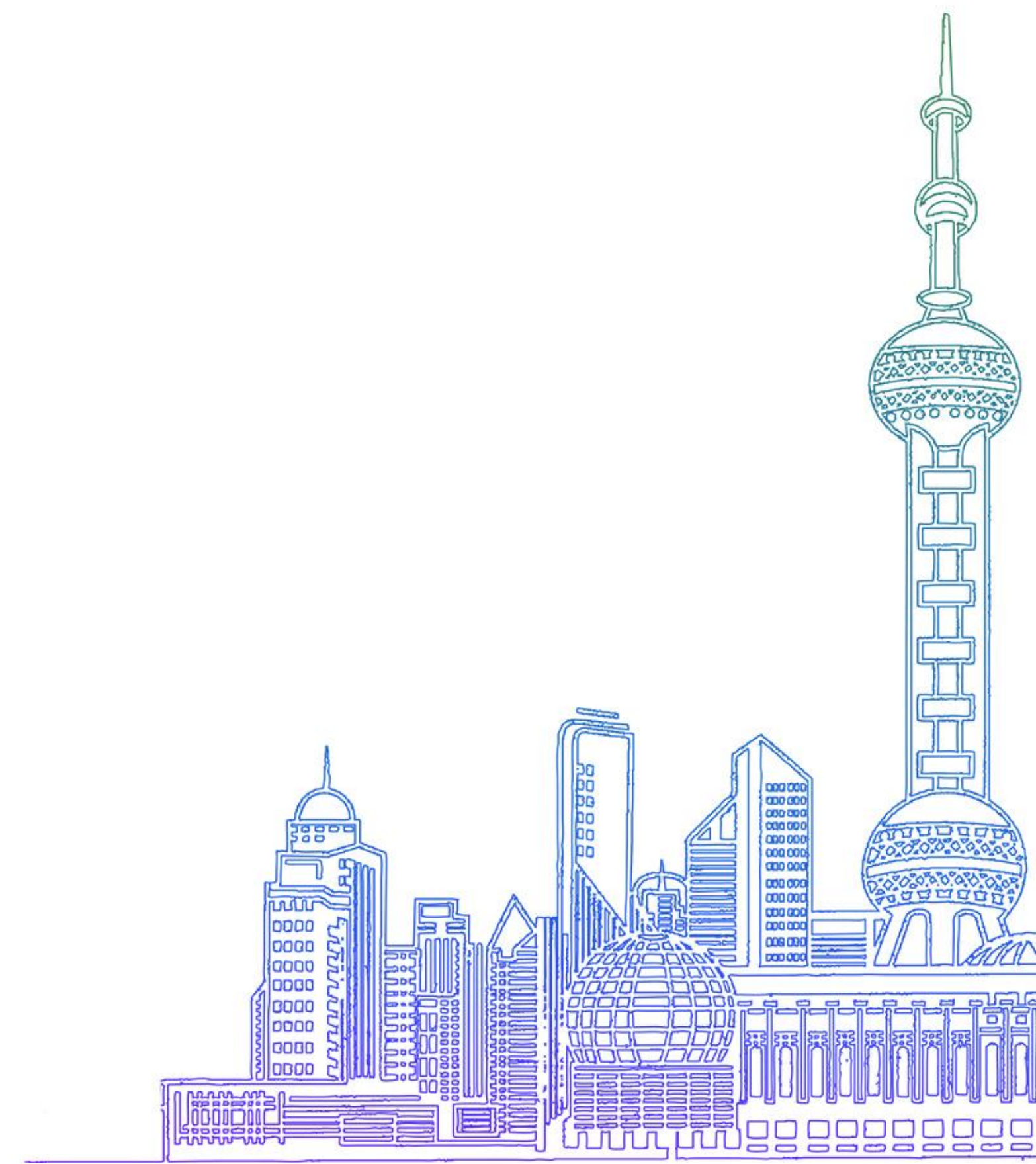
2018 vip.com third Internet ecommerce Security Summit

2018-5-5 上海



概述

- 背景介绍
- 建图
- 分群
- 图特征提取
- 模型训练与效果



背景介绍

- 业务场景
 - 唯品花（消费贷）在线申请时的欺诈识别
- 主要防范对象：团伙欺诈
 - 中介自己实施欺诈
 - 中介教大家实施欺诈
- 金融知识图谱
 - 连接欺诈账号——发现背后的人
 - 连接欺诈者——发现背后的团伙



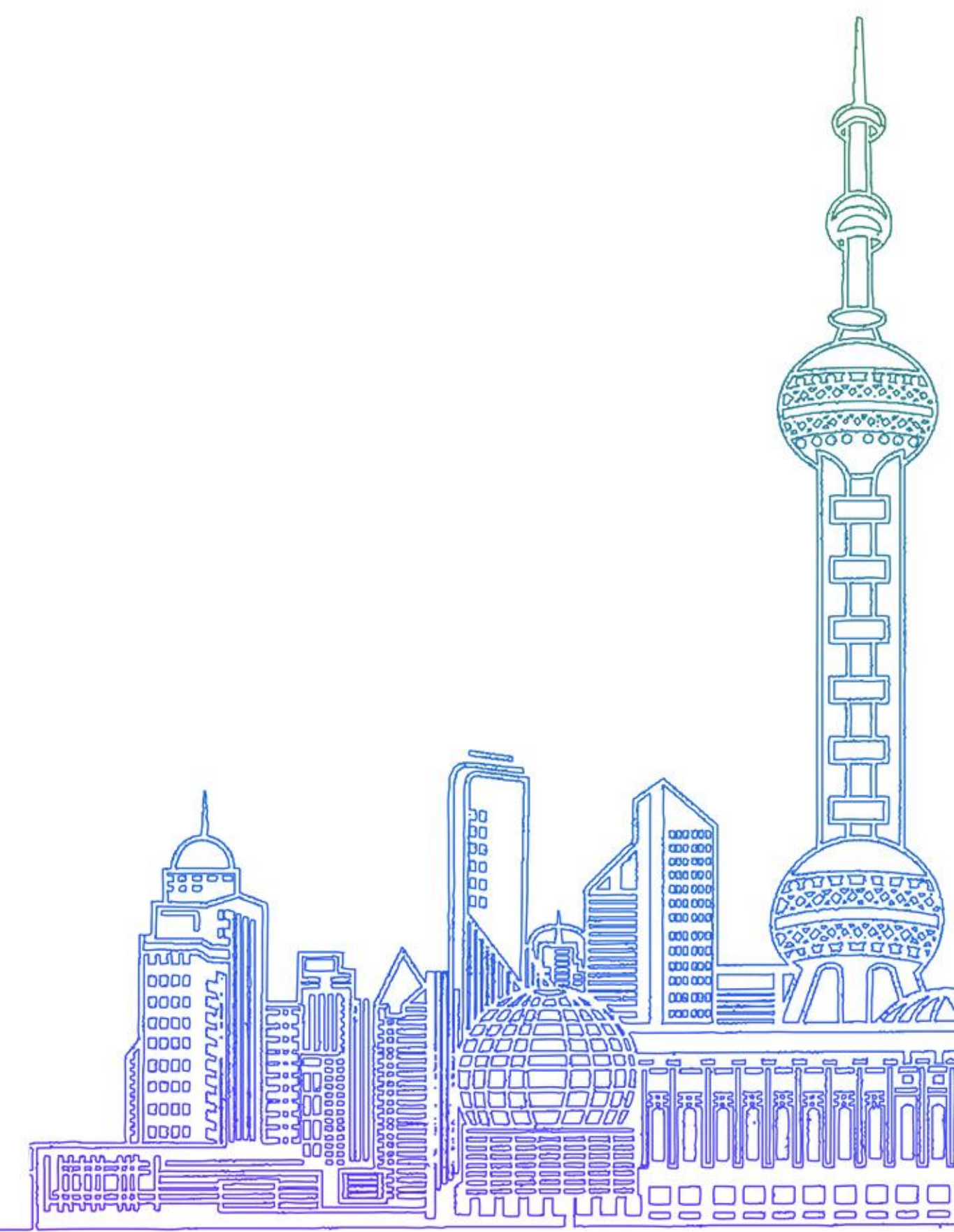
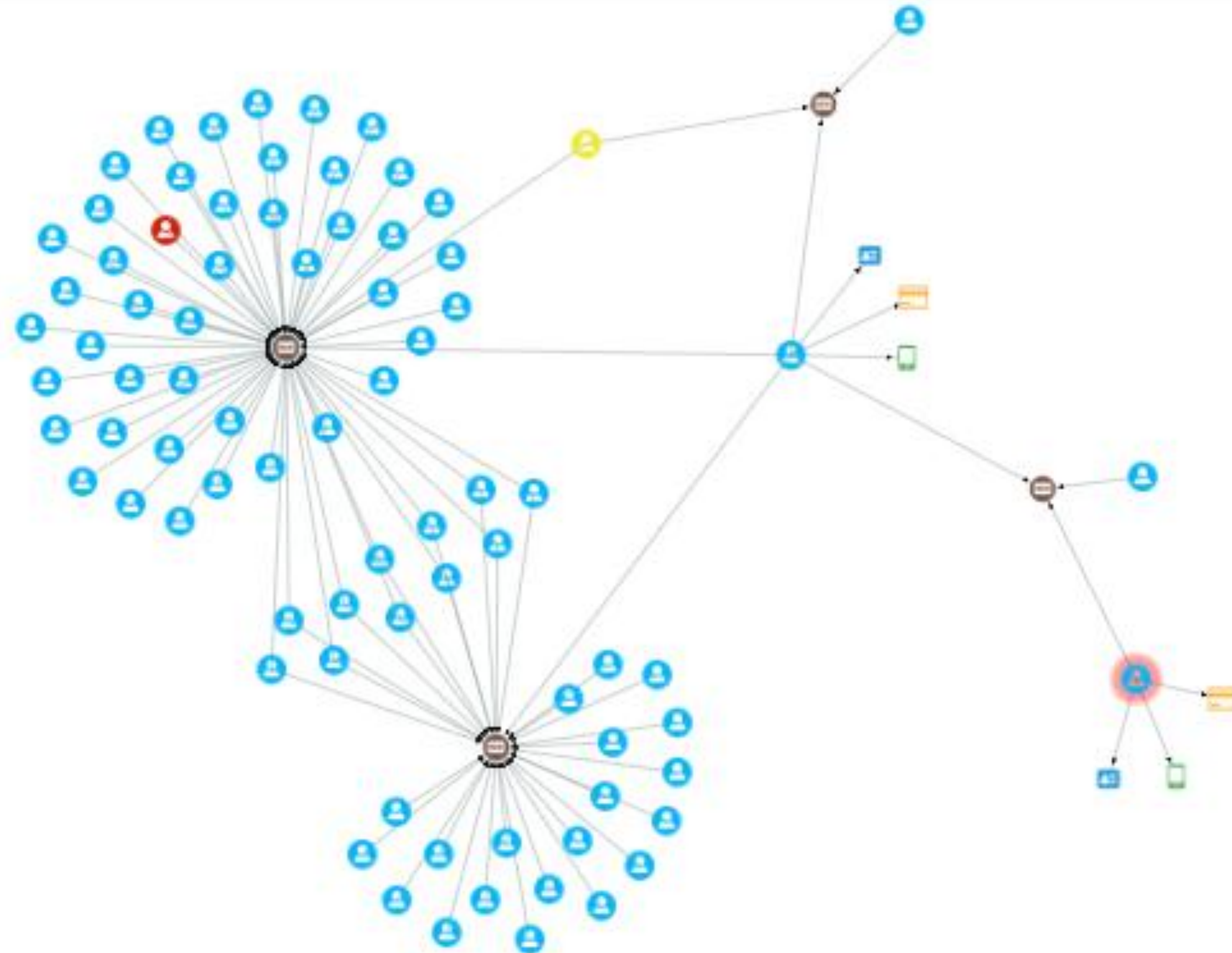
金融知识图谱示意图

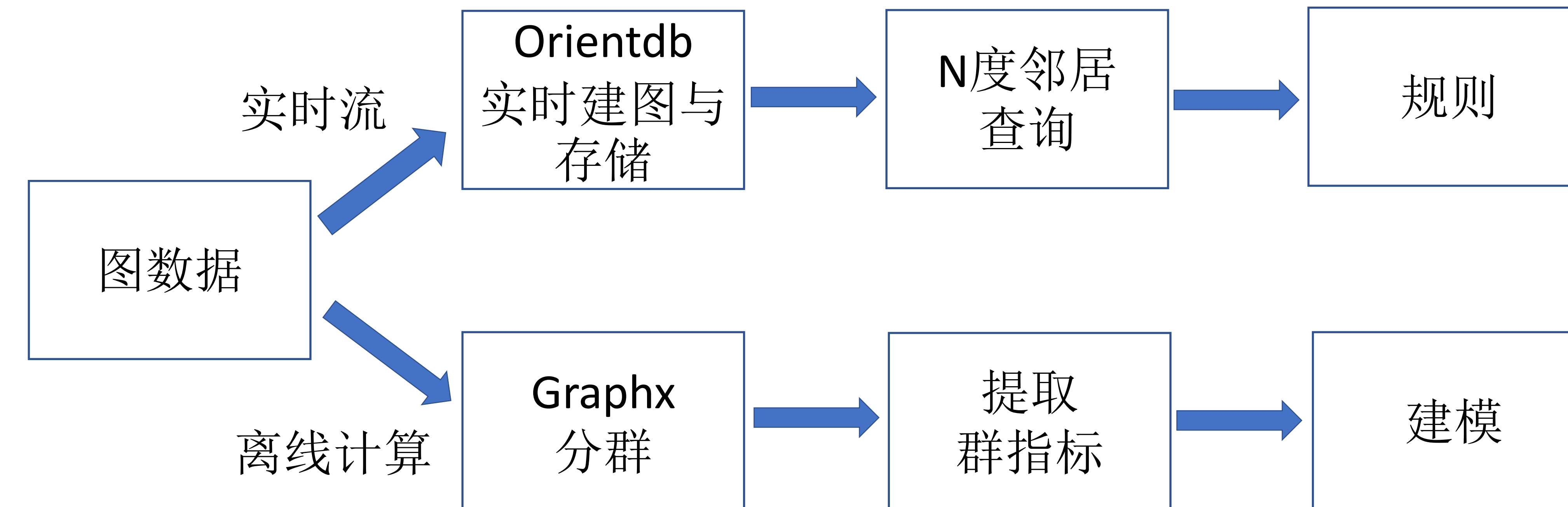
指标图

- 关注账号
- 风险账号
- 账号
- 证件号
- 手机物理串号
- 银行卡号
- 手机号

节点类型	user_id
节点内容	138564195
群编号	
距离	0

一度关联节点二度关联节点



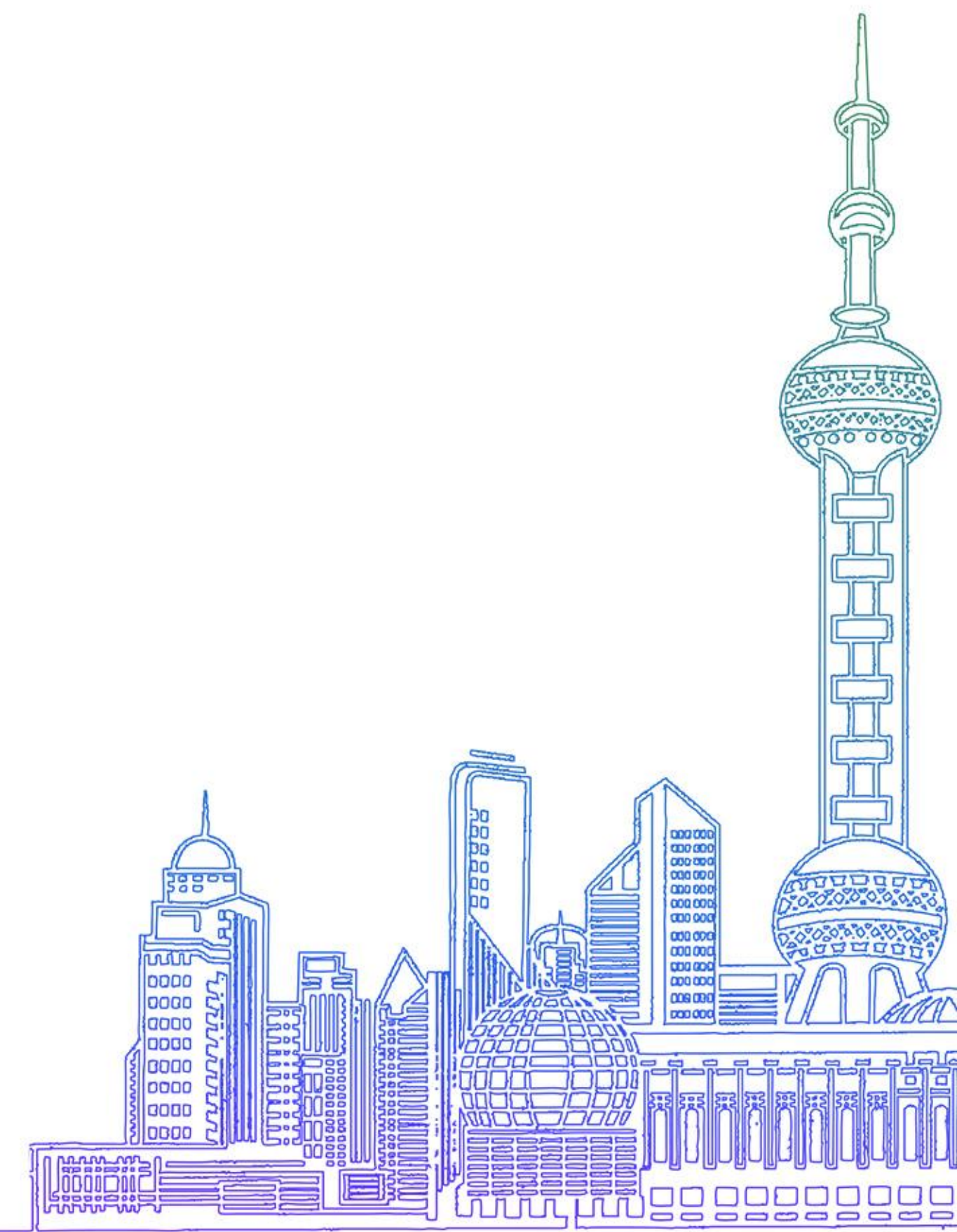
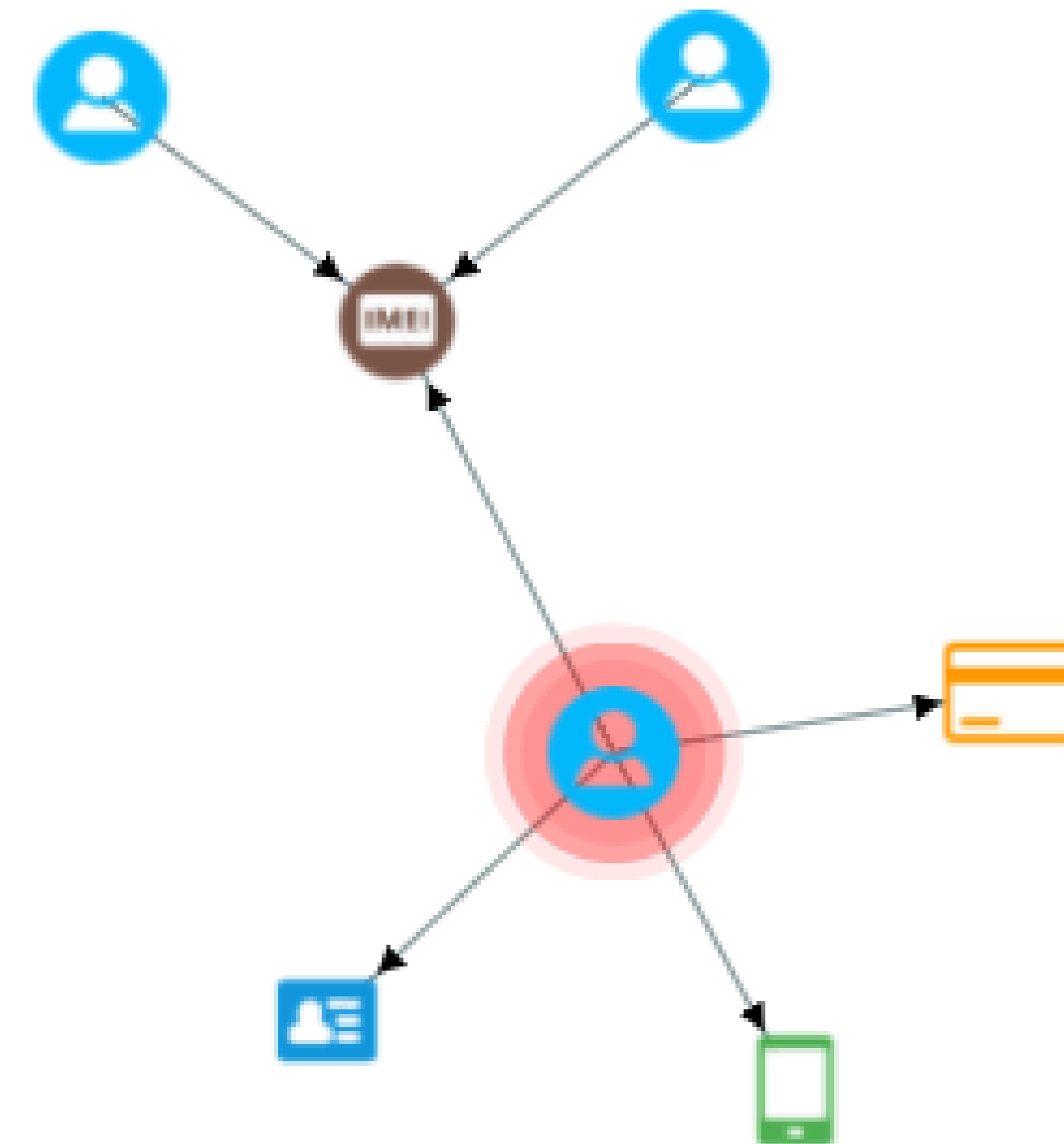


建图

- 节点
 - 用户账号
 - 手机号，证件号，银行卡号，设备号
 - 申请时点GPS、WIFI，收货地址
- 边
 - 账号与其属性节点的连接
 - 账号与账号通过属性节点建立关系
 - 边权重
- 唯品金融图规模
 - 10亿节点
 - 9亿边

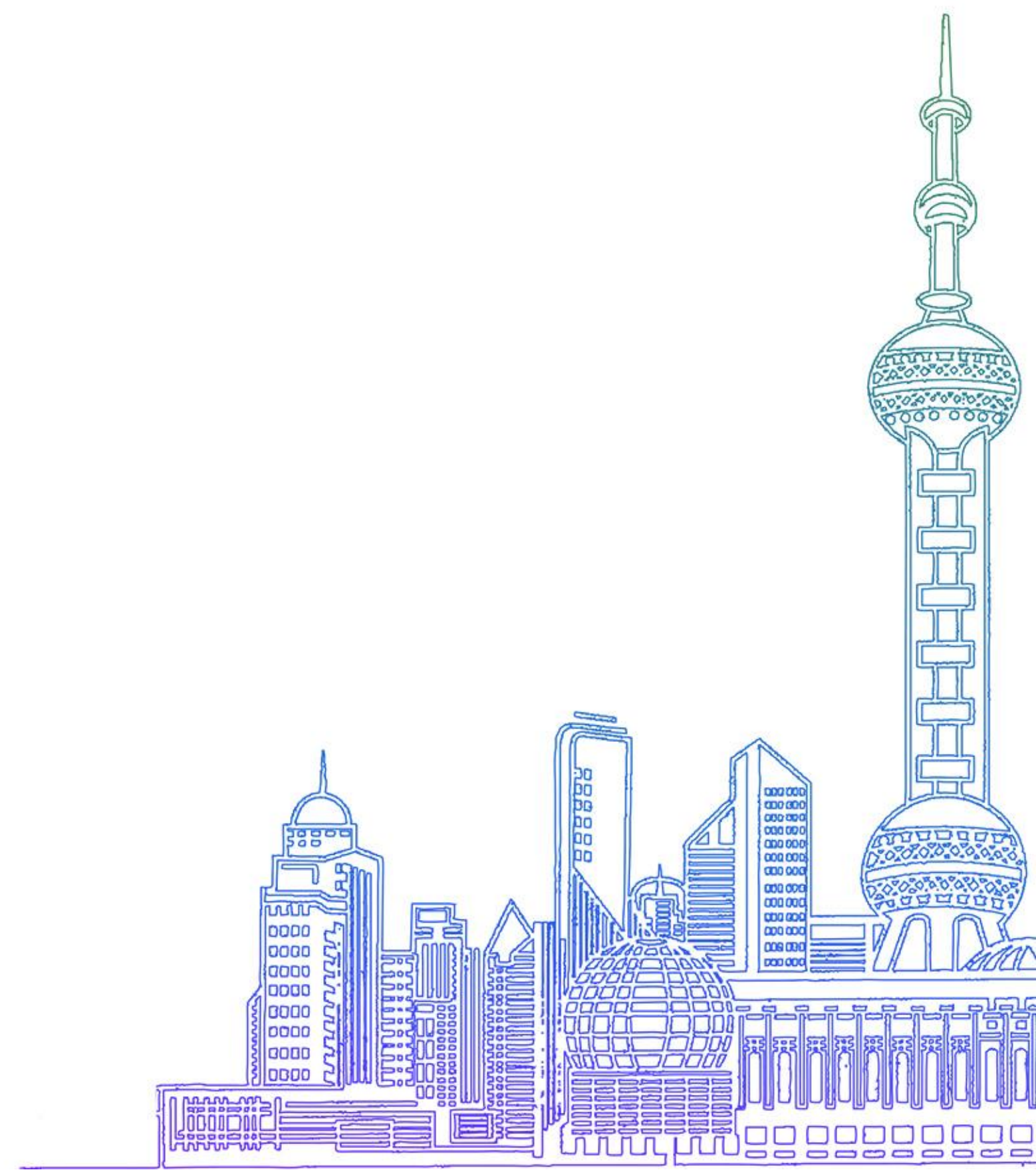
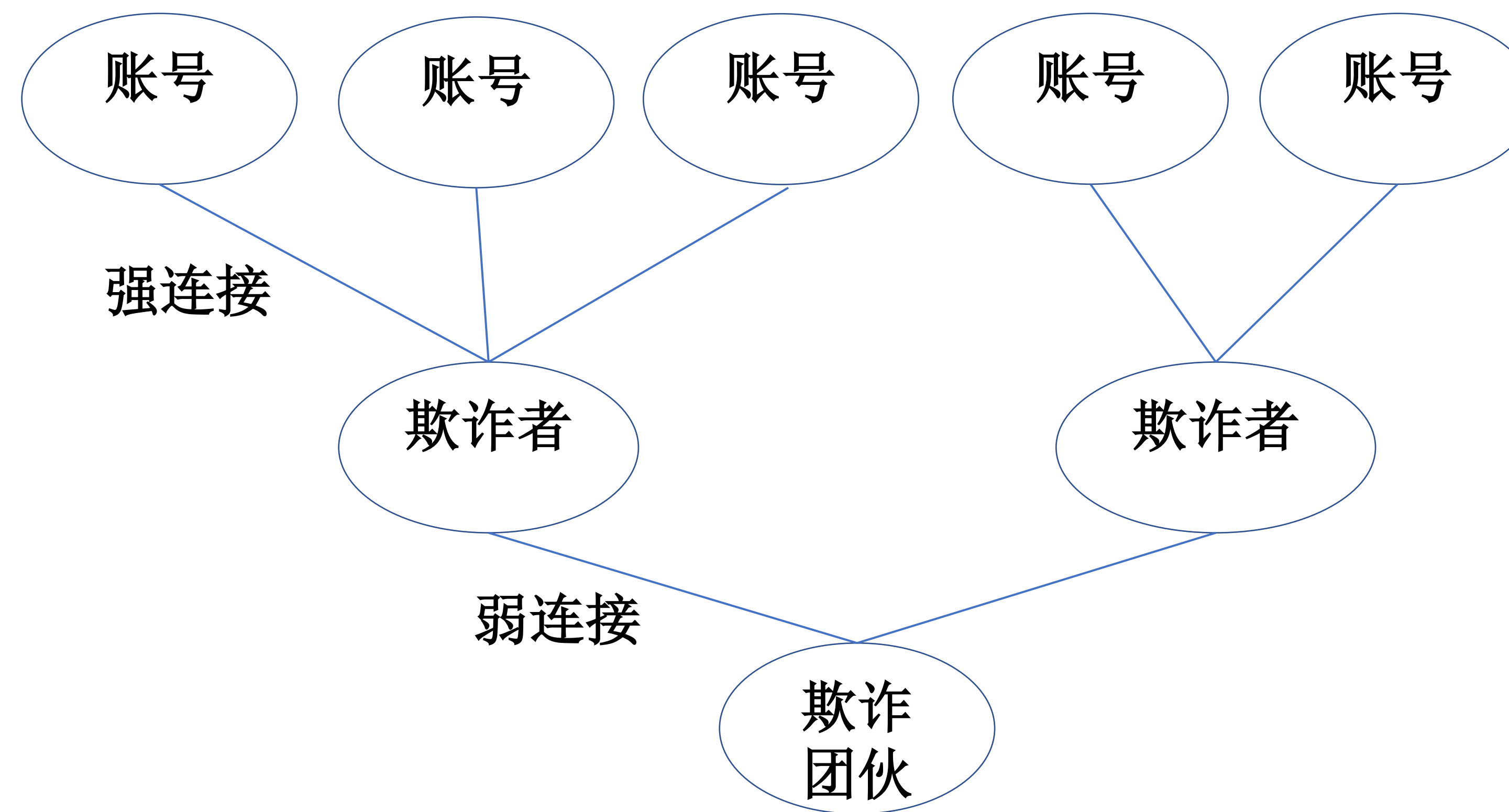
- 账号
- 证件号
- 手机物理串号
- 银行卡号
- 手机号

节点类型	user_id
节点内容	138564195
群编号	
距离	0



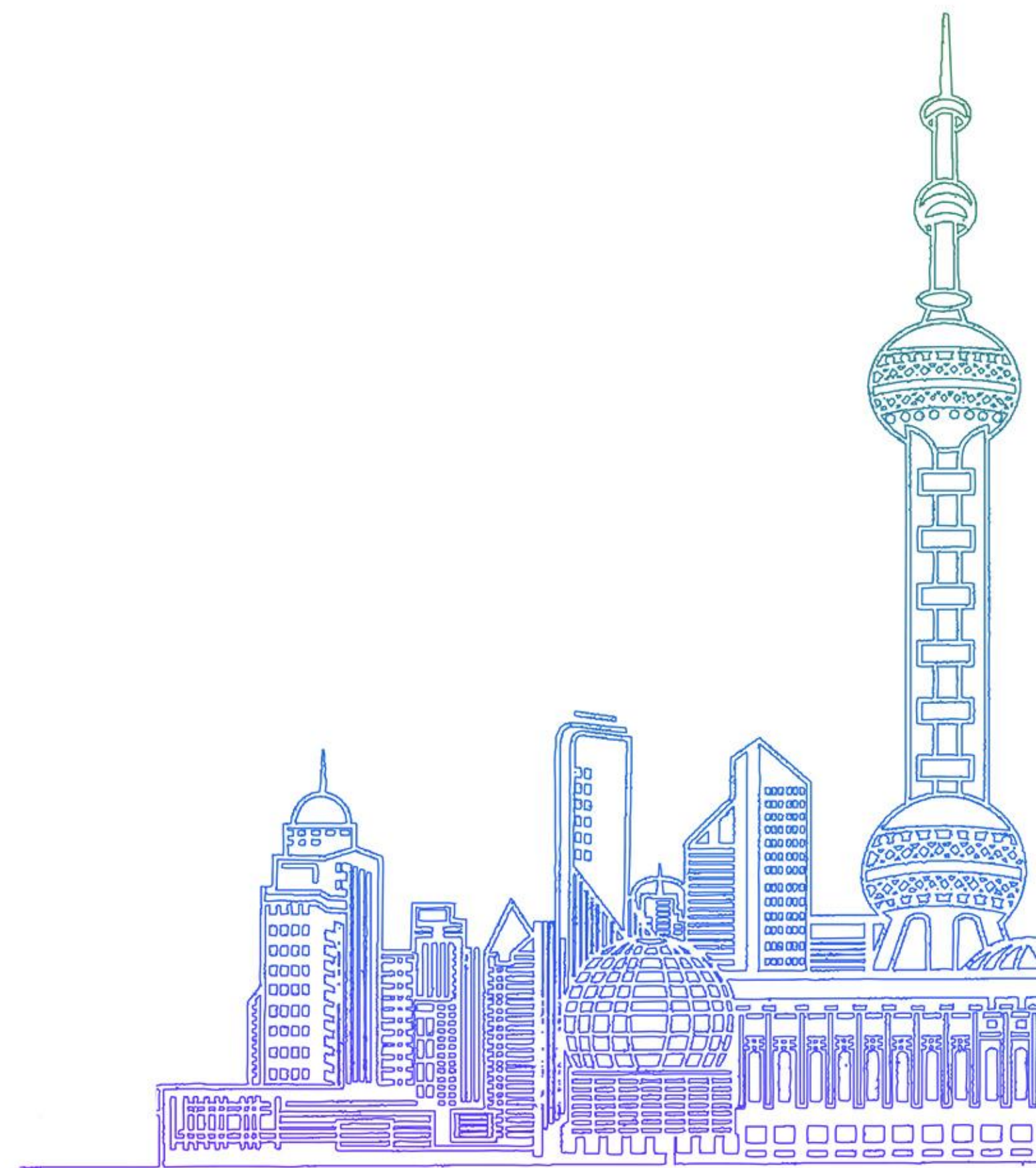
建图

- 分层级的关联挖掘
 - 挖掘账号背后的欺诈者
 - 挖掘人背后的欺诈团伙



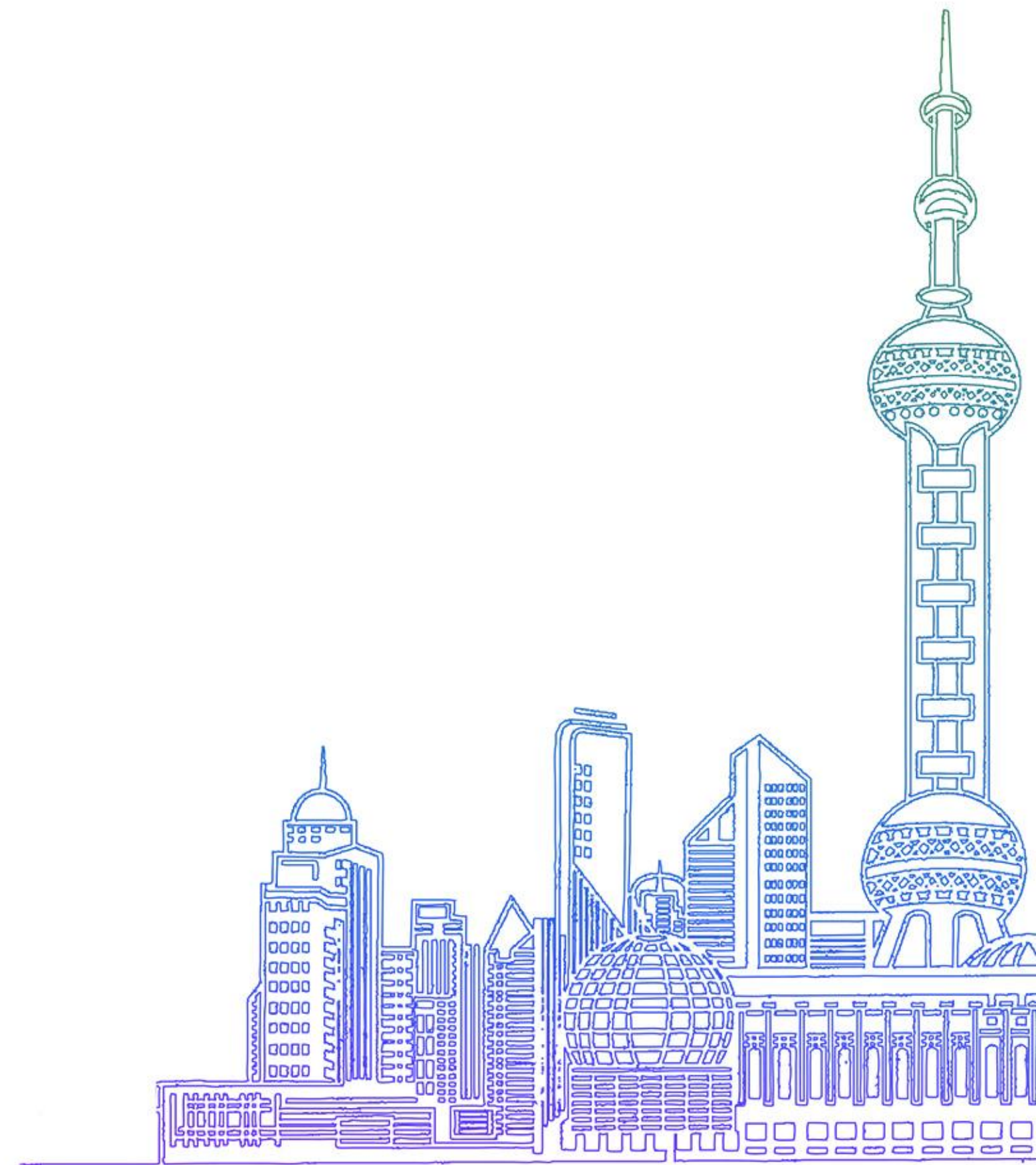
分群

- 物以类聚，人以群分
- 群定义
 - N度邻居
 - 连通图
 - 社区（community）



分群

- N度邻居
 - 批量获取计算量大
 - Orientdb实时查询
 - 一般用于规则

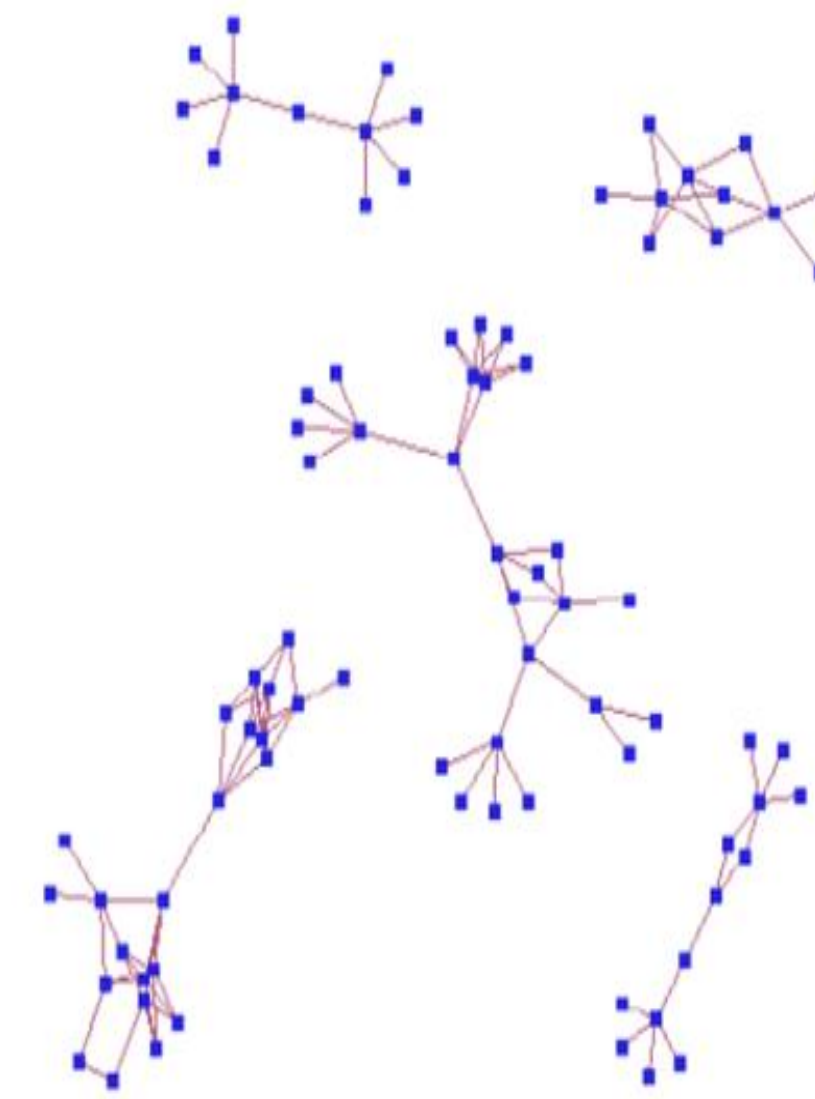


分群

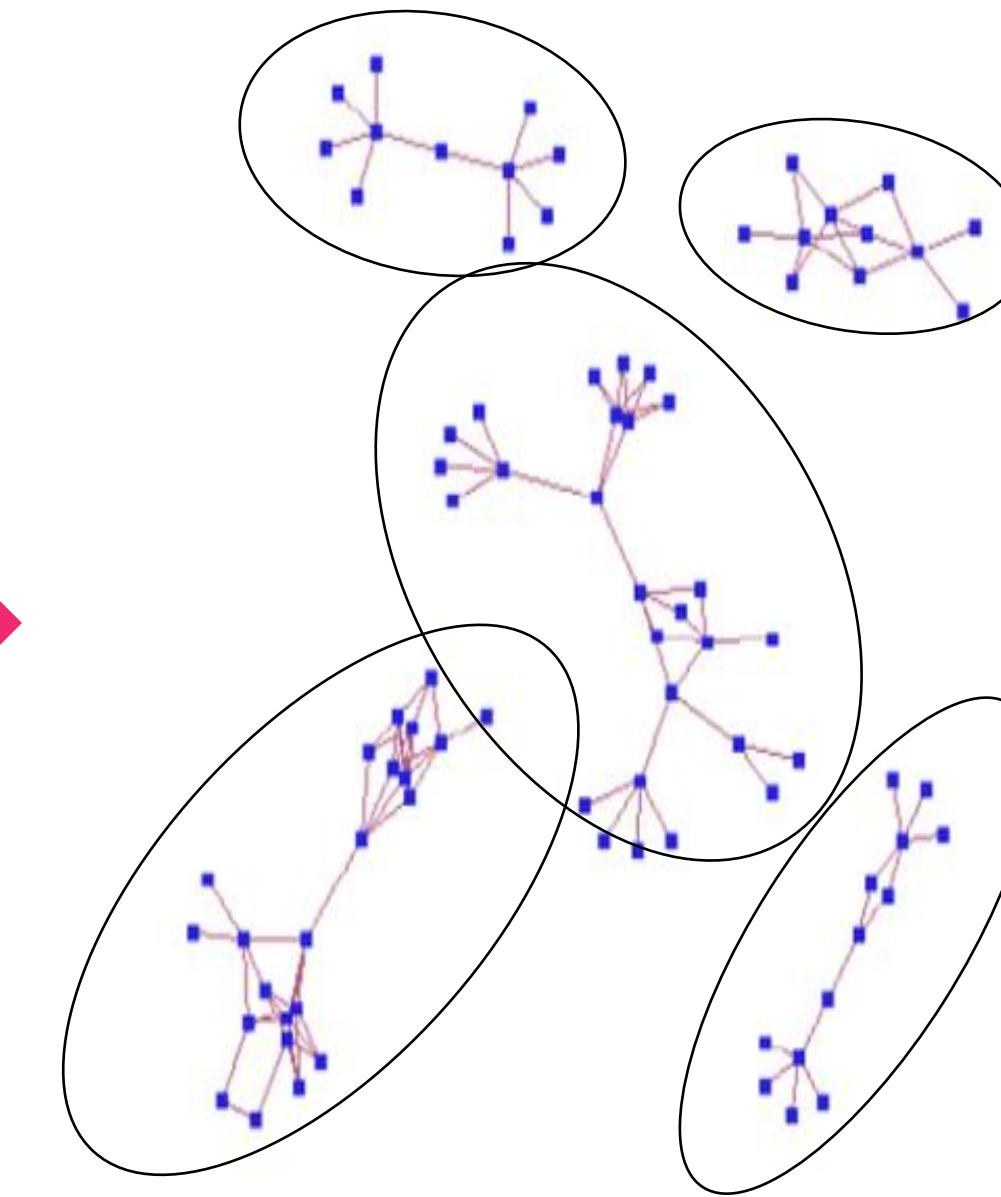
- Graphx离线分群

- 连通图适用于强连接图
- 社区发现适用于弱连接图
- 先划分连通图，再对大群社区发现

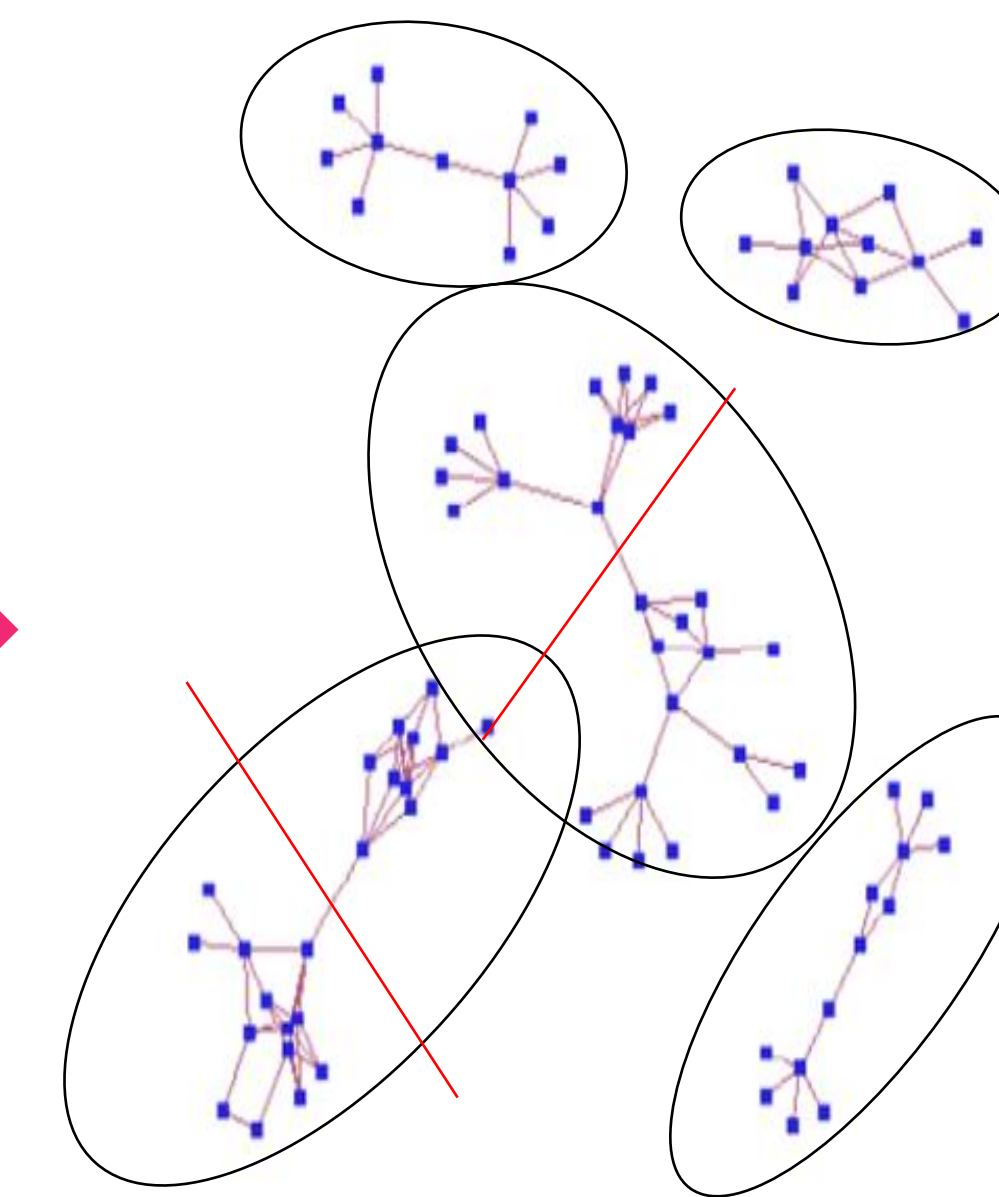
图 $G=(V,E)$



先连通图分群



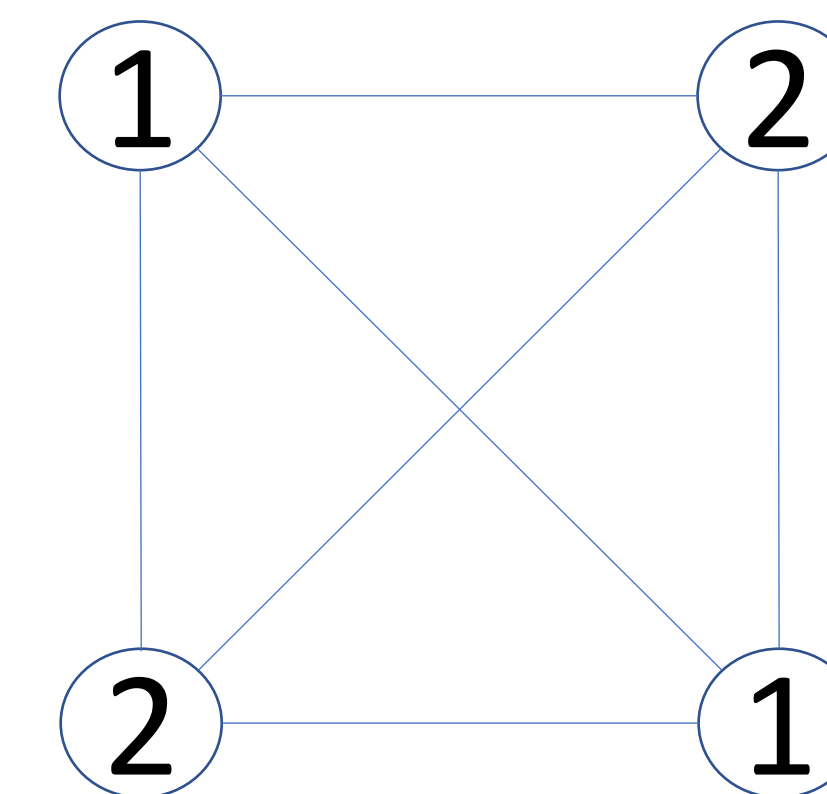
再对大群社区发现



分群

- 社区发现
 - LPA(Label Propagation Algorithm)
 - 算法原理：邻居属于哪个社区最多，节点就属于哪个社区
 - 不收敛问题
- 唯品金融群规模
 - 1.6亿个群
 - 平均每个群7个节点
 - 最大群4万节点

1->2->1->2..... 2->1->2->1.....



LPA不收敛示例



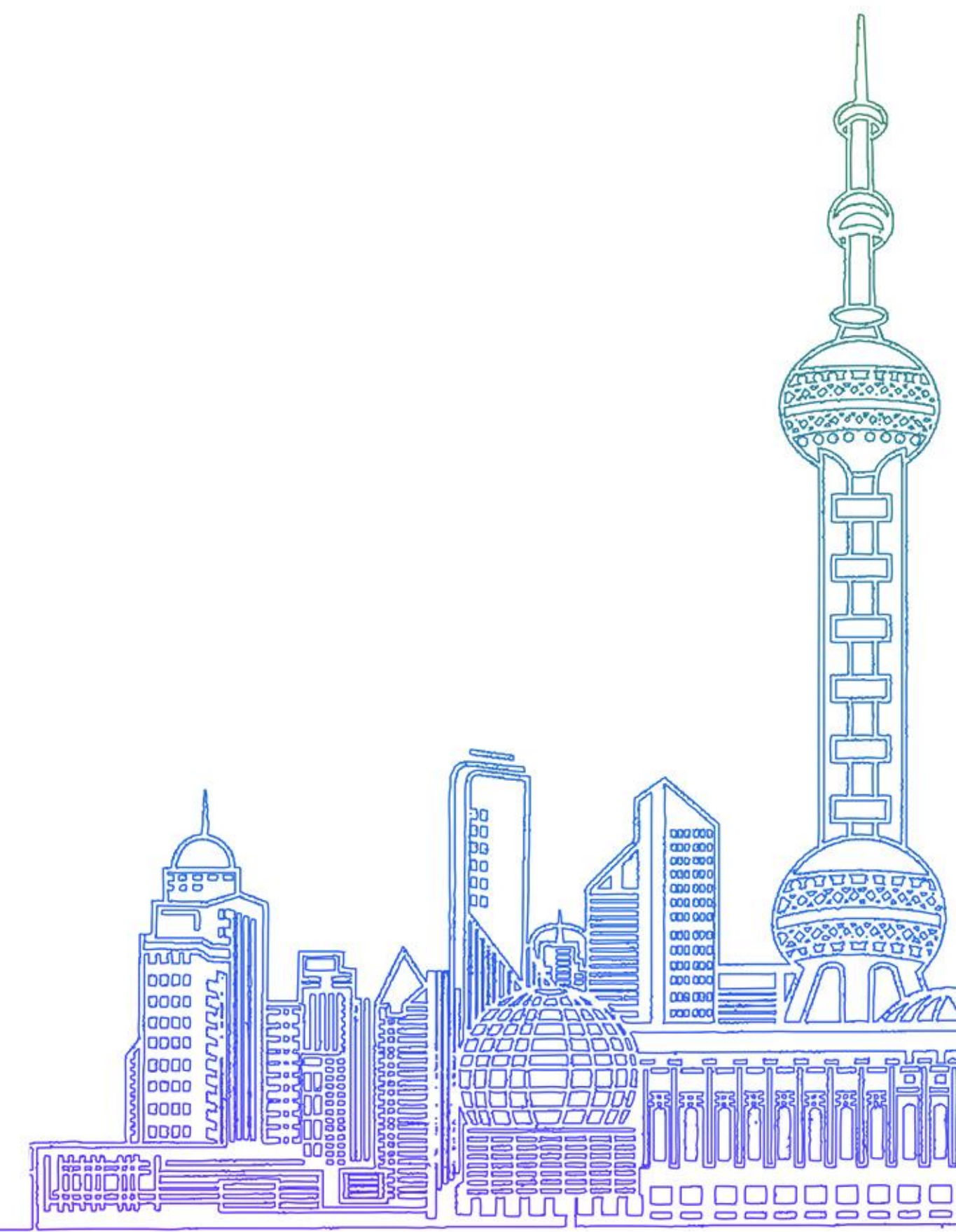
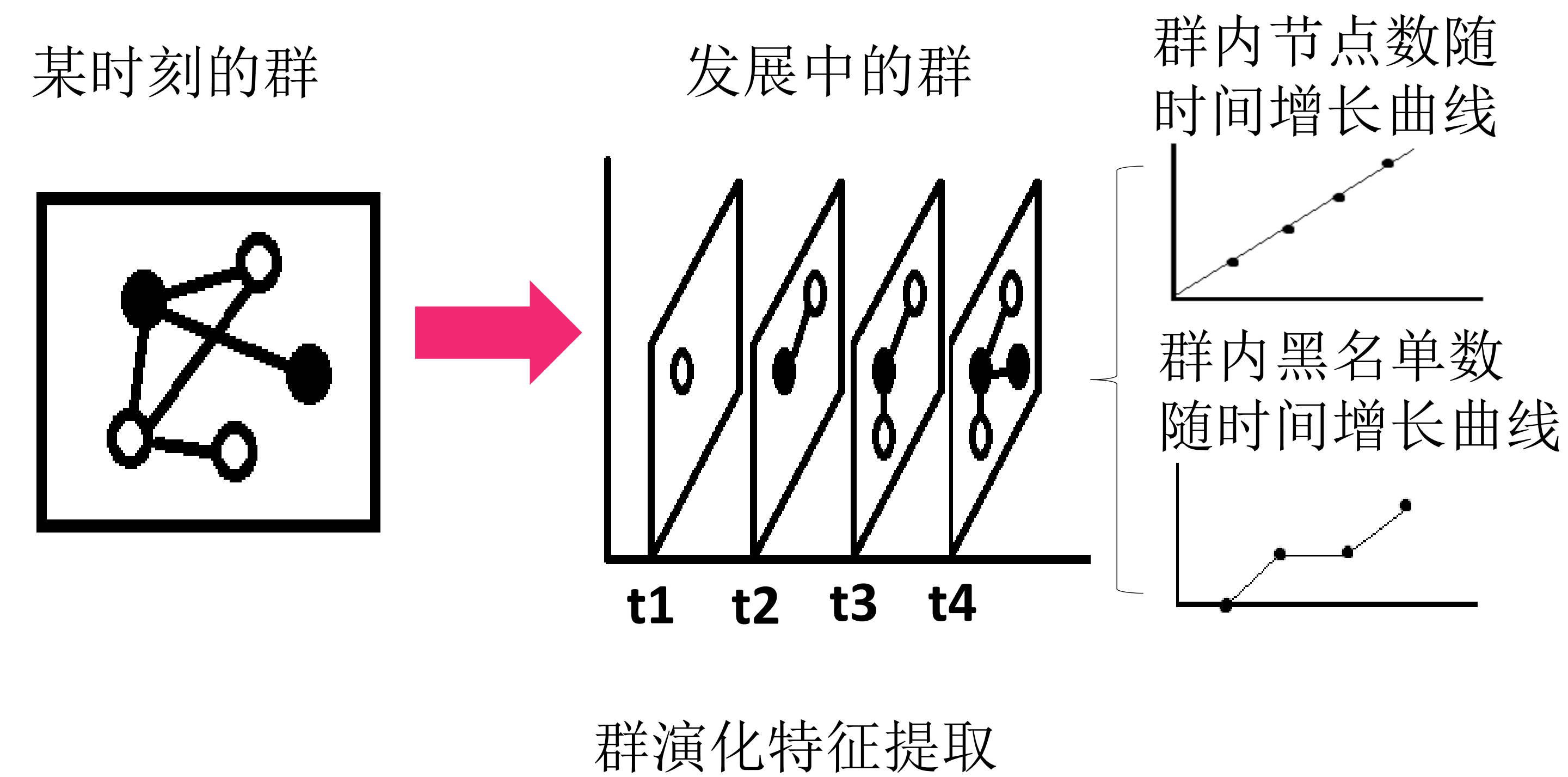
图特征提取

- 群指标
 - 群内各类型节点数量、比例
 - 群内黑名单、灰名单数量、比例
- 群内用户特征分布
 - 男女比例、年龄分布、平均消费水平
 - 平均额度使用率、平均逾期天数



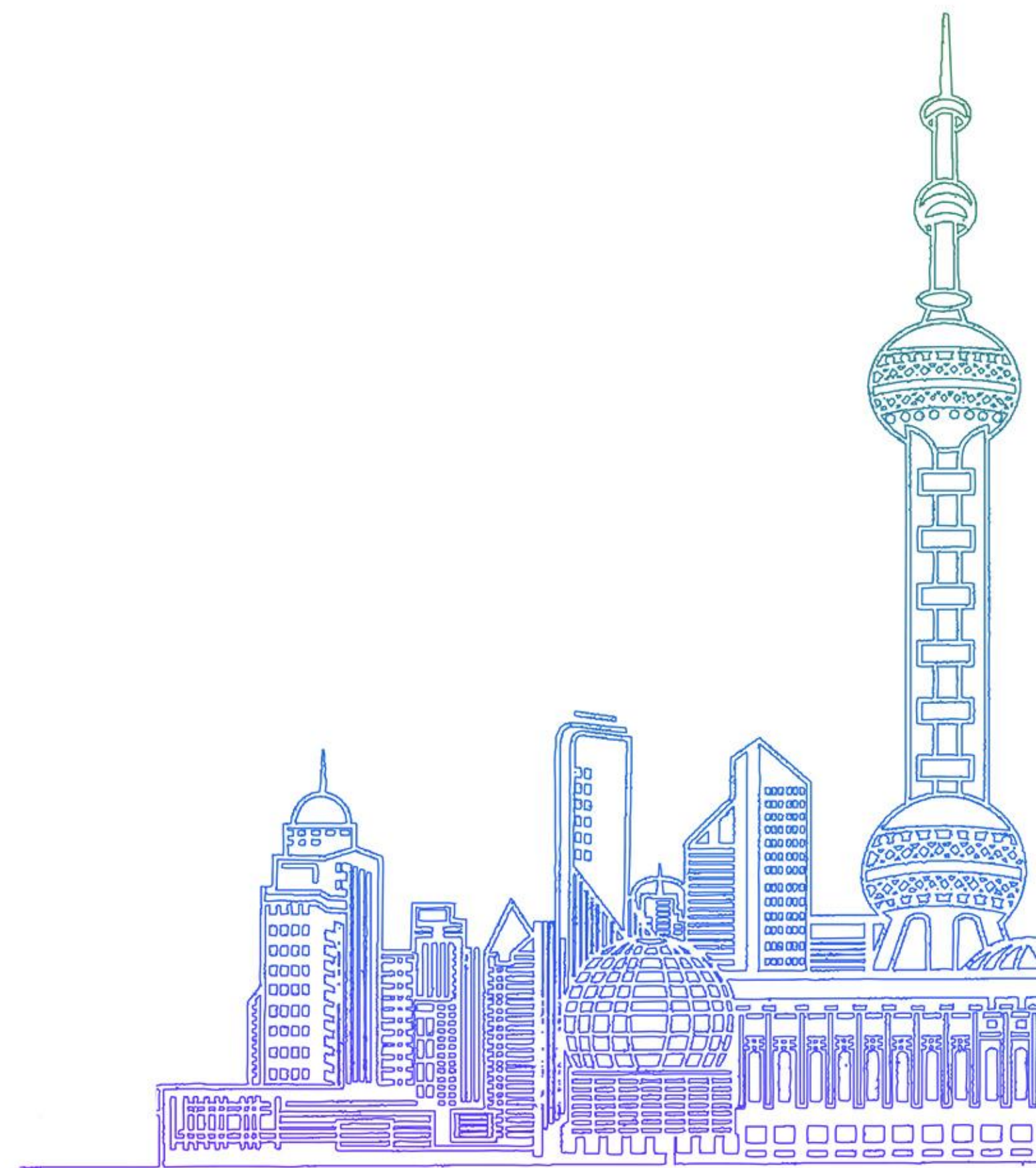
图特征提取

- 群演化特征
 - 群发展过程中的特征



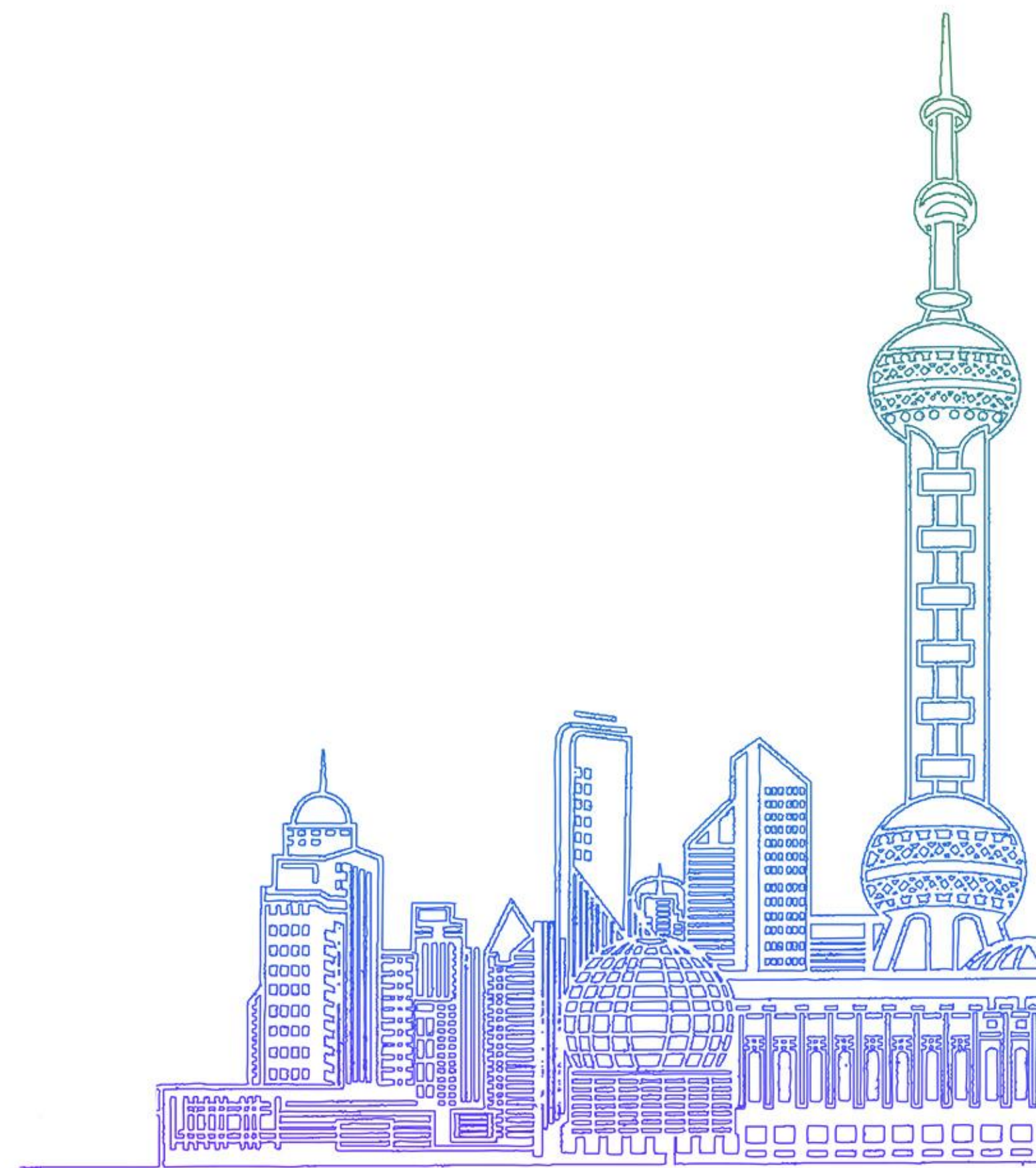
图特征提取

- 节点指标
 - 度数
 - 是否核心节点
 - node2vector



模型训练与效果

- 反欺诈模型训练
 - 标注：使用唯品花逾期90天以上不还
 - 特征：用户画像类特征+图特征
- 结果
 - 图特征为模型AUC带来约3%提升
 - 模型整体效果：唯品花申请通过率降1%，坏账降20%



谢谢观看！

2018 唯品会第三届互联网电商安全峰会

2018 vip.com third Internet ecommerce Security Summit

2018-5-5 上海

