

# 互联网金融风控中的数据科学

# 背景

- 互联网金融行业在解决什么问题？面临怎样的风险？用什么样的技术手段来应对？
  - 传统金融 V.S. 互联网金融
  - 信用风险 V.S. 欺诈风险
  - 传统风控 V.S. 数据科学技术风控



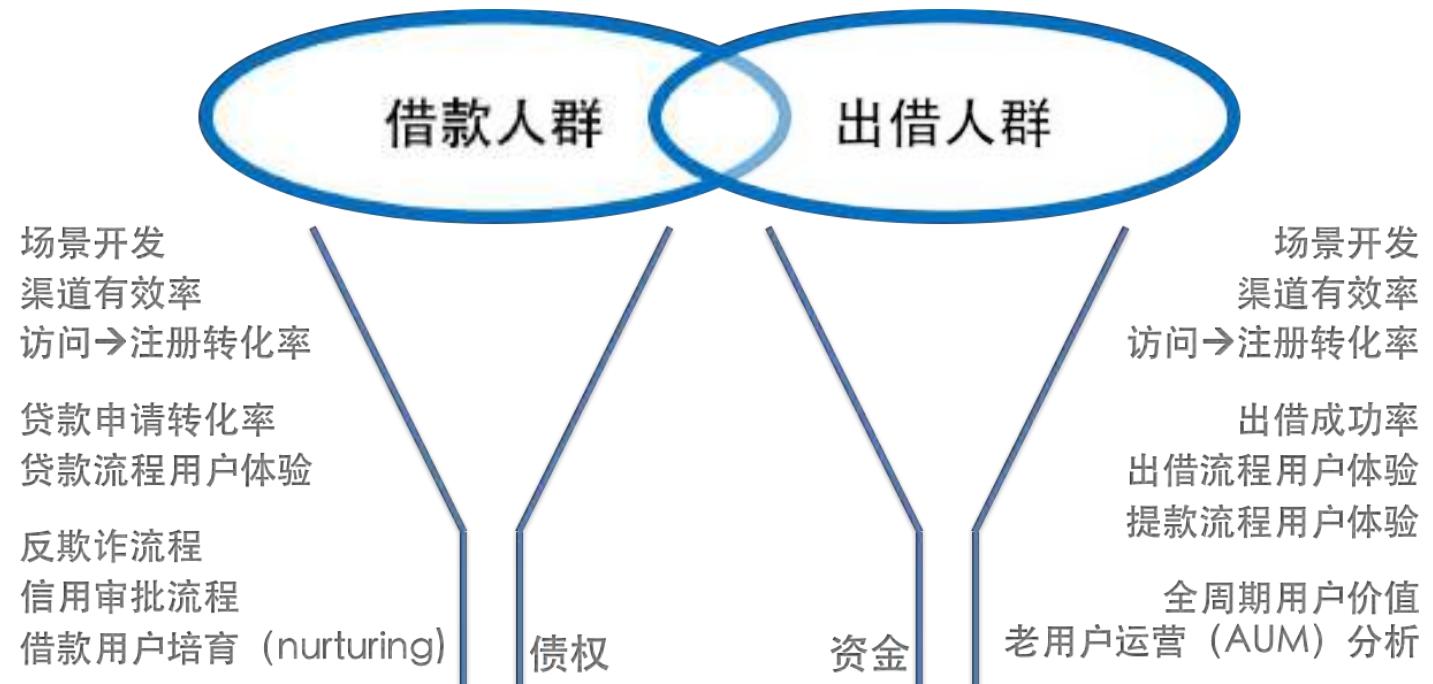
# 互联网金融风控中的数据科学

- ① 互联网金融服务面临的风险
- ② 知识图谱在金融风控中的应用场景
- ③ 大规模金融服务中的实时架构

# 互联网金融风控中的数据科学

- ① 互联网金融服务面临的风险
- ② 知识图谱在金融风控中的应用场景
- ③ 大规模金融服务中的实时架构

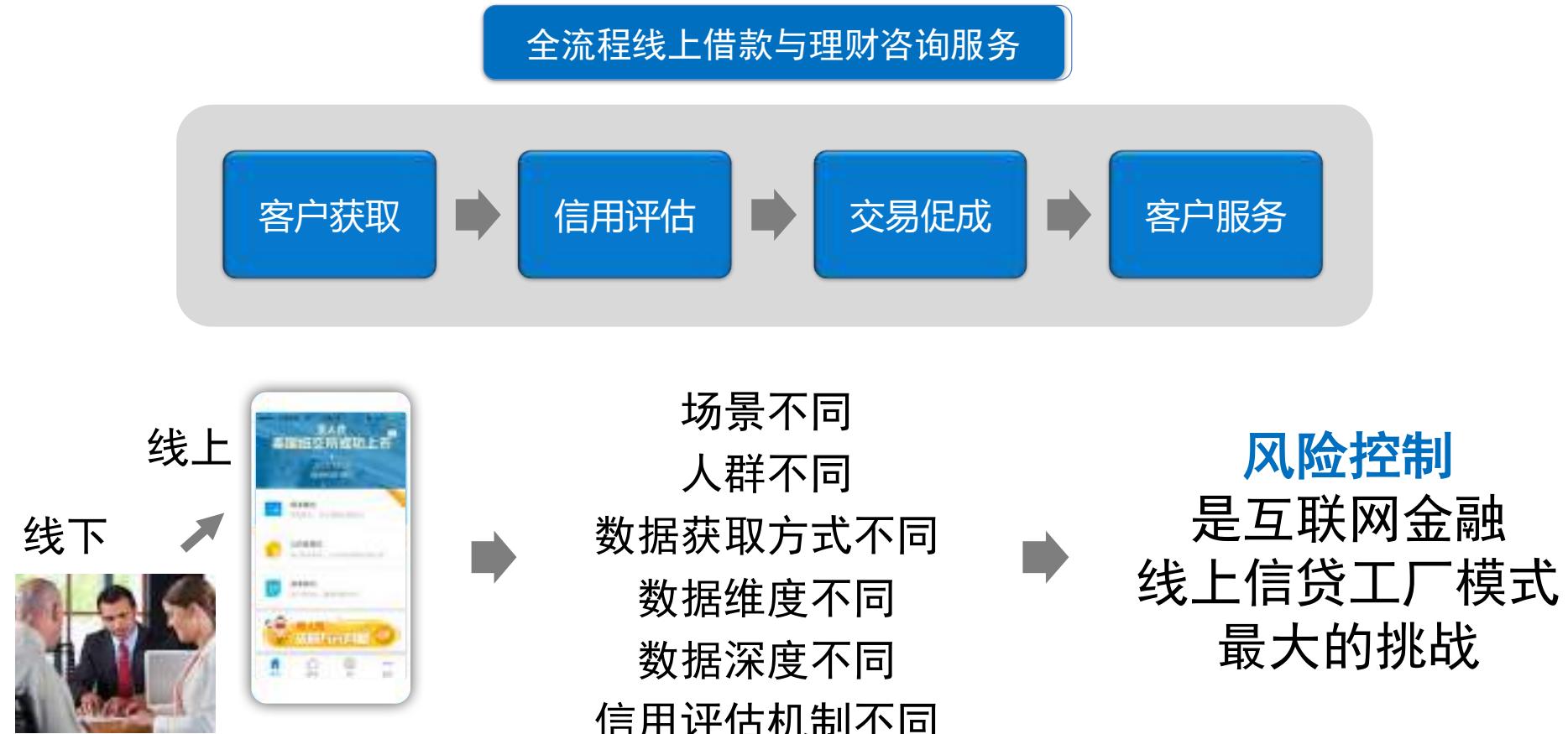
# 互联网金融-个人对个人的信用贷款



宜人贷：P2P 借款与理财咨询服务平台



# 宜人贷：极速信任 - 自动化信用评估



# 互联网金融行业中的风险

风险	遇到的问题	业界通常解决方法	业界的方法为什么无效
信用风险	还款能力	收集收入水平、消费水平、负债情况等对用户进行风险评分	无权威数据、数据收集难度大、传统评分卡有效特征挖掘难度大
欺诈风险	伪冒申请和欺诈交易	人工审查、信用黑名单、基于规则	人工效率低、无权威黑名单、无法自动发现异常、欺诈手段更新快

人群团体化



地区集中化



方式多样化



工具智能化



# 互联网金融风控中的数据科学

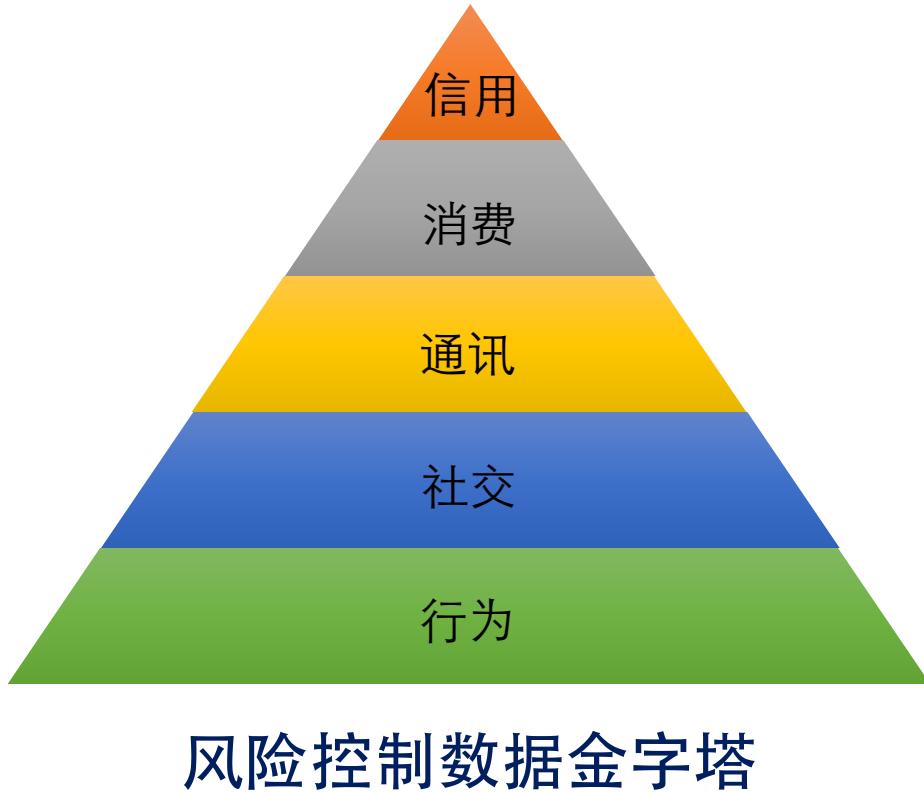
① 互联网金融服务面临的风险

② 知识图谱在金融风控中的应用场景

③ 大规模金融服务中的实时架构

# 互联网金融中的风控是一种机器学习过程

- Y目标： Benchmark选取
  - 好、坏用户定义
  - 训练、测试和跨时间验证样本
- X变量： 特征工程
  - 人工特征工程
  - 图谱特征挖掘技术
    - ✓ 知识图谱技术
    - ✓ 图挖掘技术



# 知识图谱在金融风控中的优势

## 传统风控：

- 客户是否触黑
- 客户的消费记录是否异常

点



## 利用知识图谱风控：

- 客户一度、二度关系是否触黑
- 客户消费关联商家是否异常
- 一机多人
- 识别组团欺诈

面

# 构建金融知识图谱：FinGraph

- 10种实体
  - 电话、身份证、银行卡、信用卡、IP、设备号、地理位置等
- 约1.3亿节点
- 约10亿边关系
- 预计到2017年增长20倍

## FinGraph 平台系统

### 应用场景层面

智能搜索、反欺诈、贷后管理、营销分析、运营支撑 等

### 系统支持层面

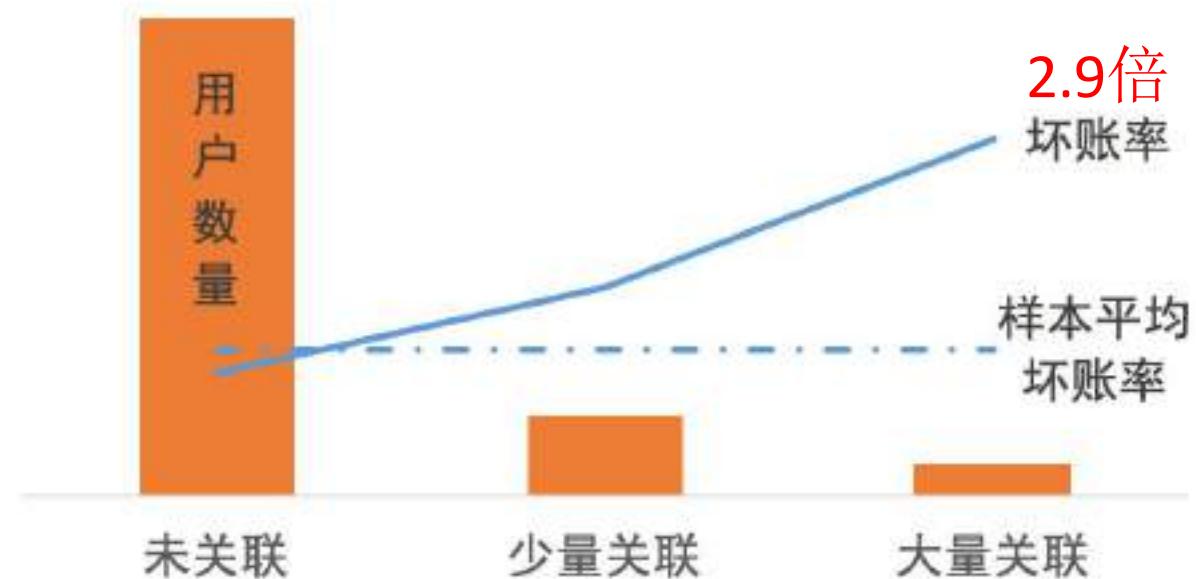
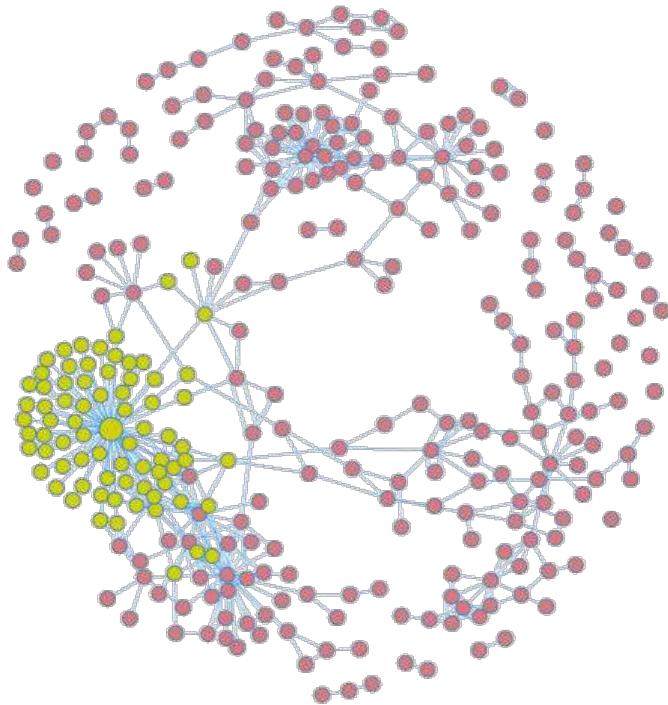
特征工程、模型开发、异常监控、推荐系统 等  
Spark+GraphX+MLlib+Streaming+TensorFlow

### 数据整合层面

信用数据、金融消费数据、行为数据、社交数据、  
网络安全、第三方数据 等  
图数据库neo4j

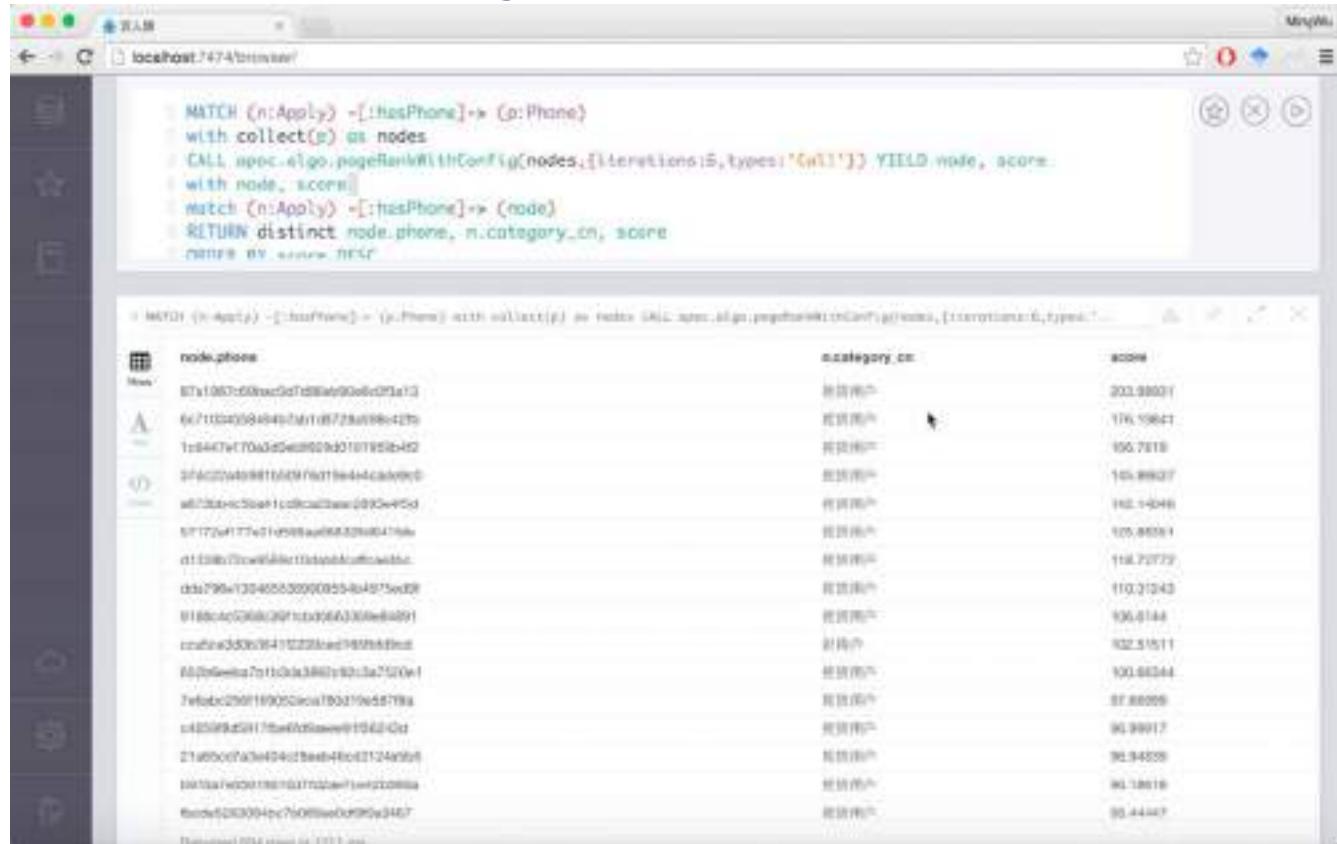
# 风险分析实践：人以群分(1)

- 借款用户通信社交网络与欺诈风险
  - 结论：与坏用户有大量关联的借款用户的坏账率是未关联用户的2.9倍



# 风险分析实践：人以群分(2)

- 从整体借款群体的角度，用PageRank算法探索哪些用户与大量借款用户有  
关联关系



The screenshot shows the Neo4j browser interface. The top part displays a Cypher query:

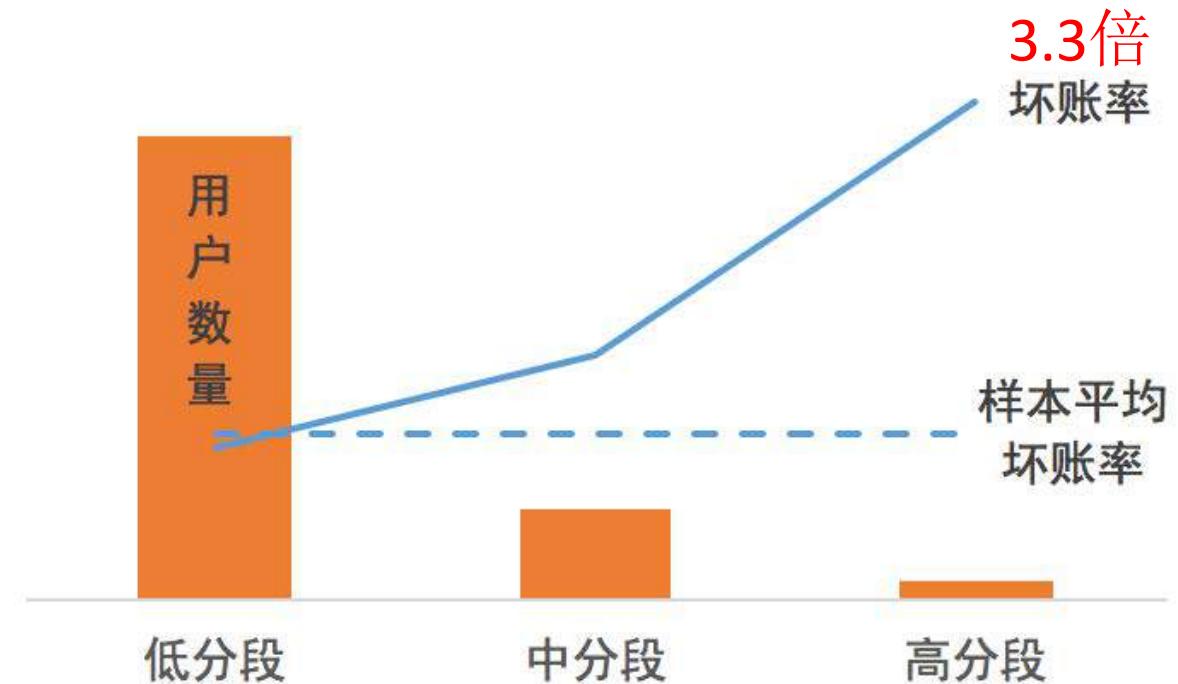
```
MATCH (n:Apply) -[:HasPhone]-> (p:Phone)
WITH collect(p) AS nodes
CALL apocalgo.pageRankWithConfig(nodes,[:iterations:5,:types:'Call']) YIELD node, score
WITH node, score
MATCH (n:Apply) -[:hasPhone]-> (node)
RETURN DISTINCT node.phone, n.category_cn, score
ORDER BY score DESC
```

The bottom part shows the results of the query, which is a table with three columns: node.phone, category\_cn, and score. The data is as follows:

node.phone	category_cn	score
671087569ac5d7e88ab90e6c29a13	投资类	203.38031
667103459440756108728a698e42b	投资类	176.19841
1c8447970d3d5e029d3010198e442	投资类	156.7818
3f4c22498816497ed1fe44ca99y6	投资类	145.88627
af723a9c35ef1c09ca23ee2990e45d	投资类	142.54046
57172941774014988ae94329e04716	投资类	129.86584
c1128a772ce854de7f5a9a4e4e0a	投资类	118.72779
1251799e13046553290005558a4975e0f	投资类	110.37243
9188ca5268c3971c3d964330e8409	投资类	93.6144
ccadine350934112291ac795fb4fb2	投资类	92.51511
f63fb6e1279110da386078213a7520e1	投资类	90.66344
7efabbc290f189052e0a780379e5879a	投资类	87.88099
c425089d5917ba4725aee919424cd	投资类	86.38917
21a655cc743d494c25eef4fc023124a9f8	投资类	86.34228
b6f8aa10e0771122aef1e9a208a	投资类	86.18618
9ec0e503394pc750mws00995v3467	投资类	85.44447

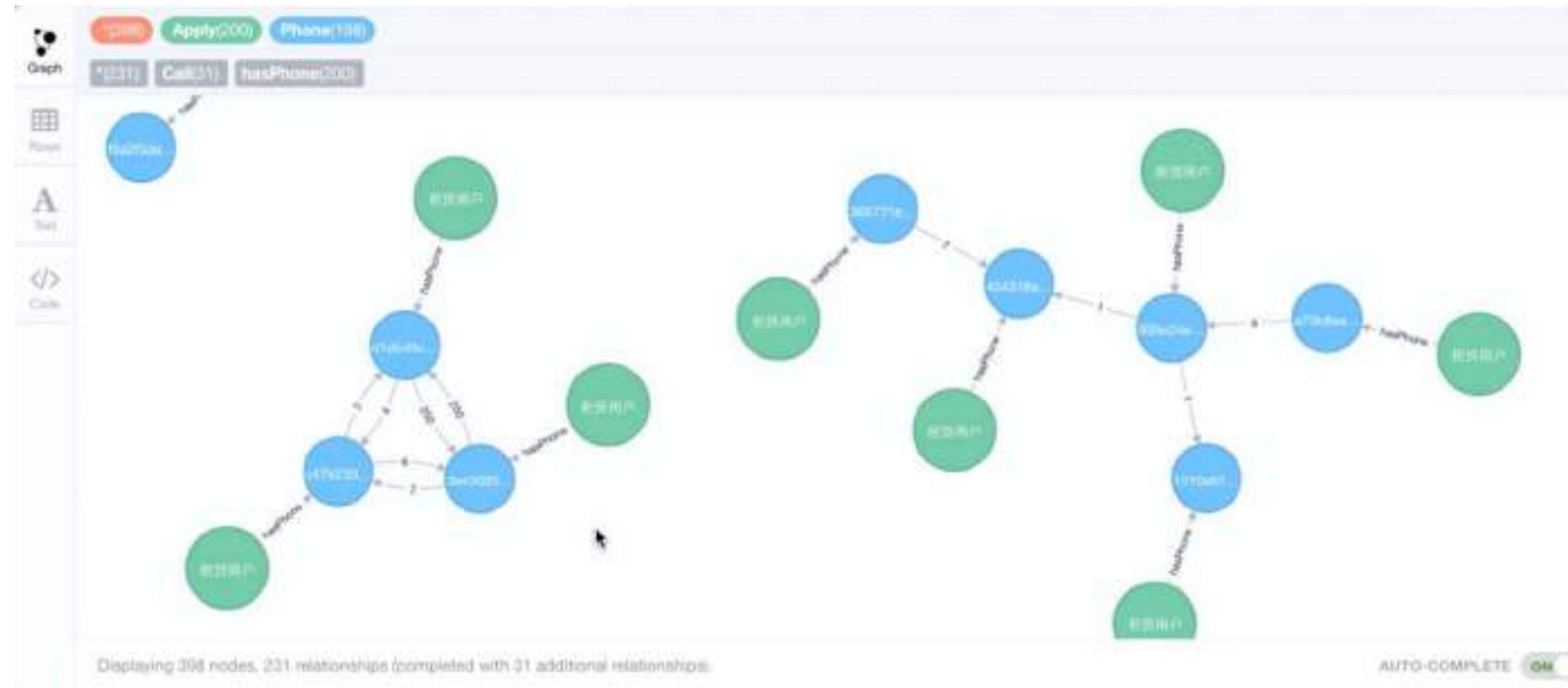
# 风险分析实践：人以群分(2)

- 借款用户通信社交网络与欺诈风险
  - 结论：PageRank高分段用户的坏账率是低分段用户的3.3倍



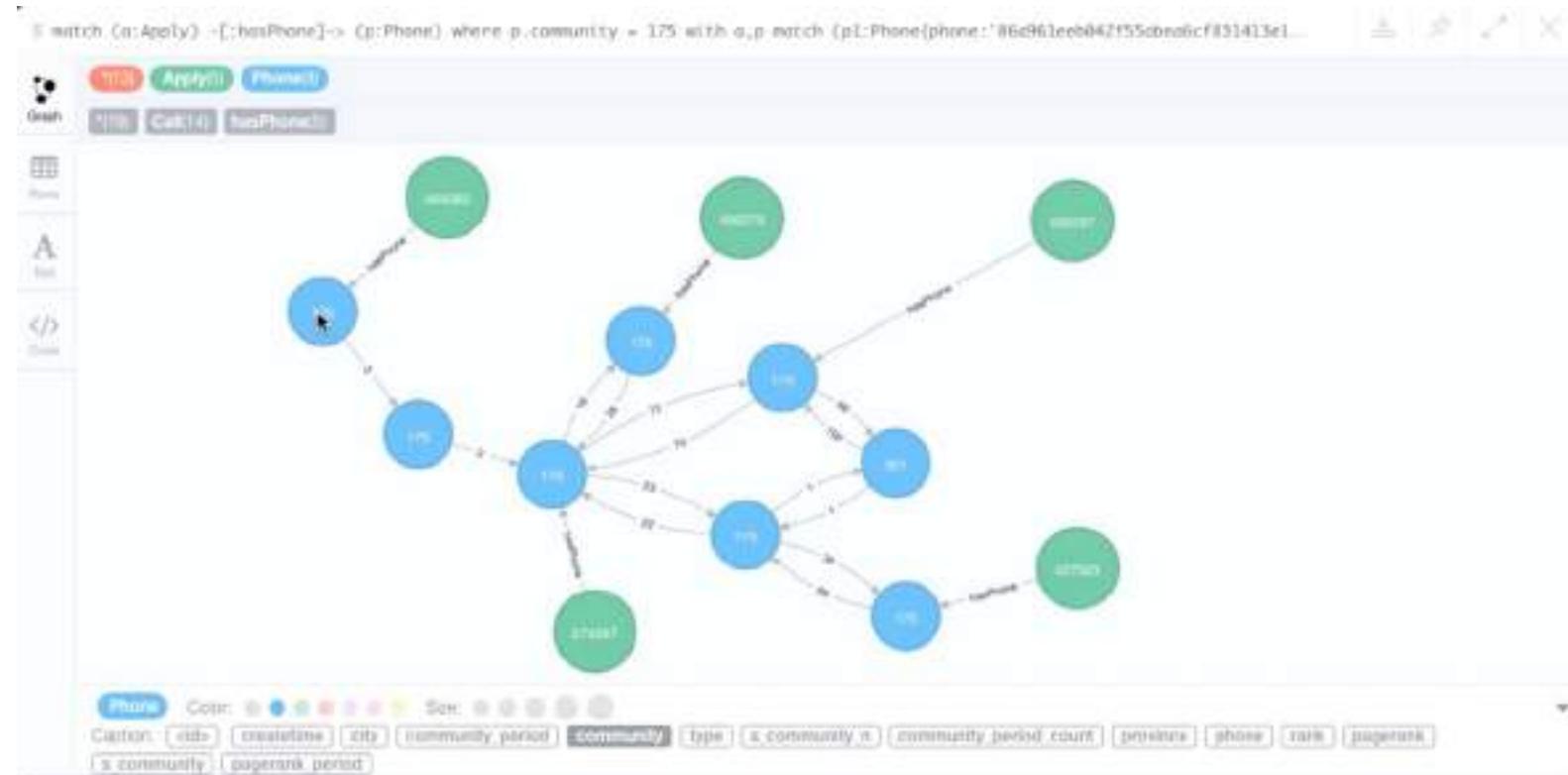
# 风险分析实践：识别组团欺诈风险

- 通过社区发现算法来实时评估每个用户的组团欺诈风险



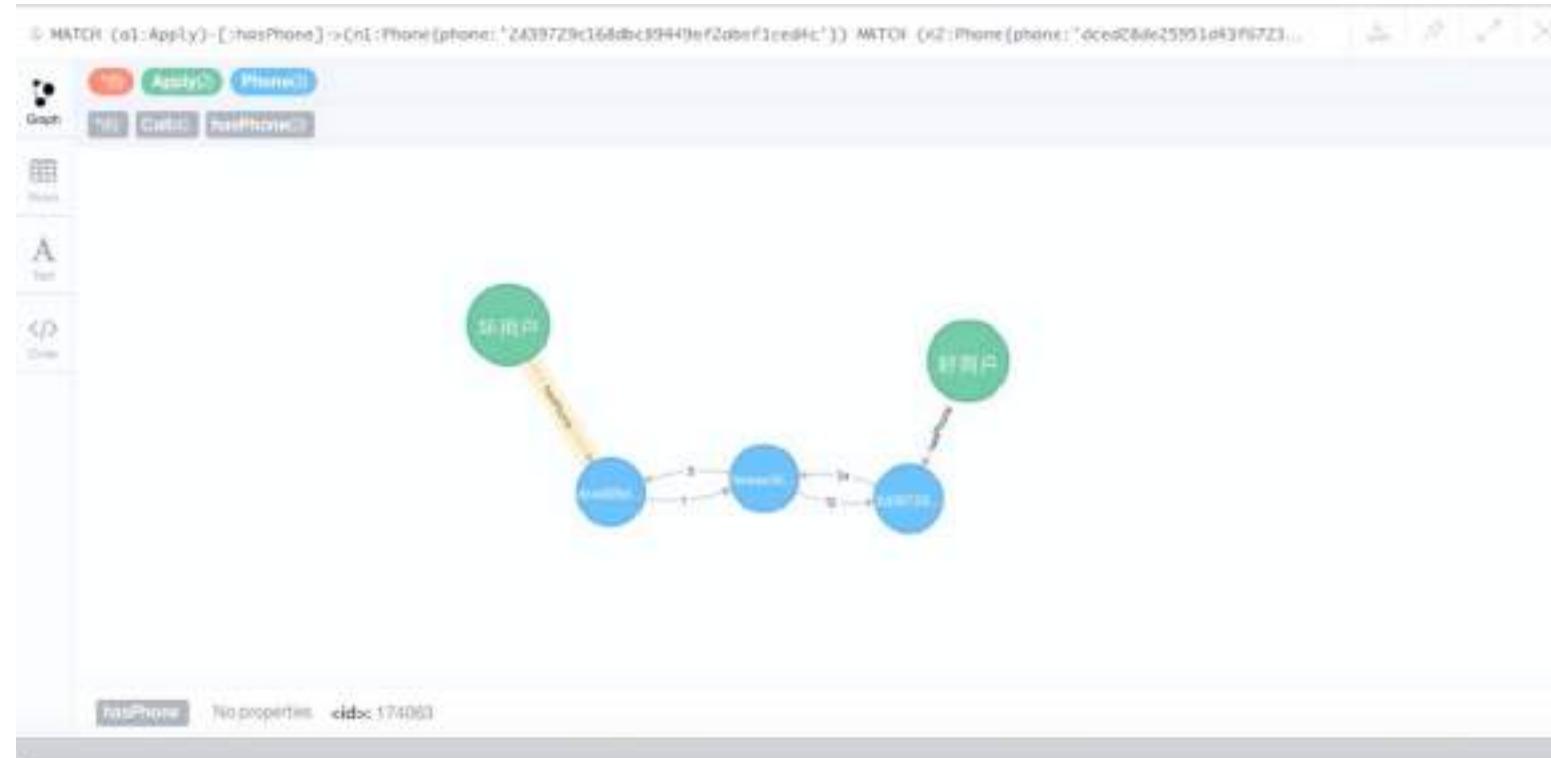
# 欺诈案例调查的挑战：失联用户找寻

- 通过社区发现算法挖掘失联用户的通信社交网络

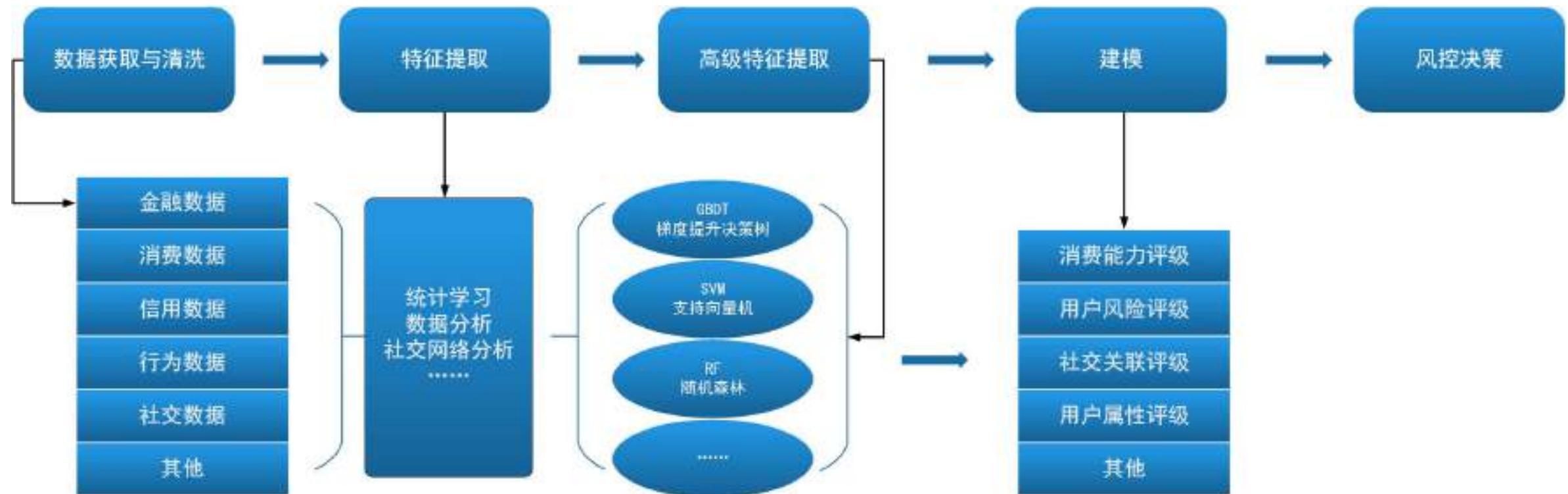


# 欺诈案例调查的挑战：失联用户找寻

- 在社区内使用最短路径算法来发现失联用户与一个正常还款用户的  
关系链

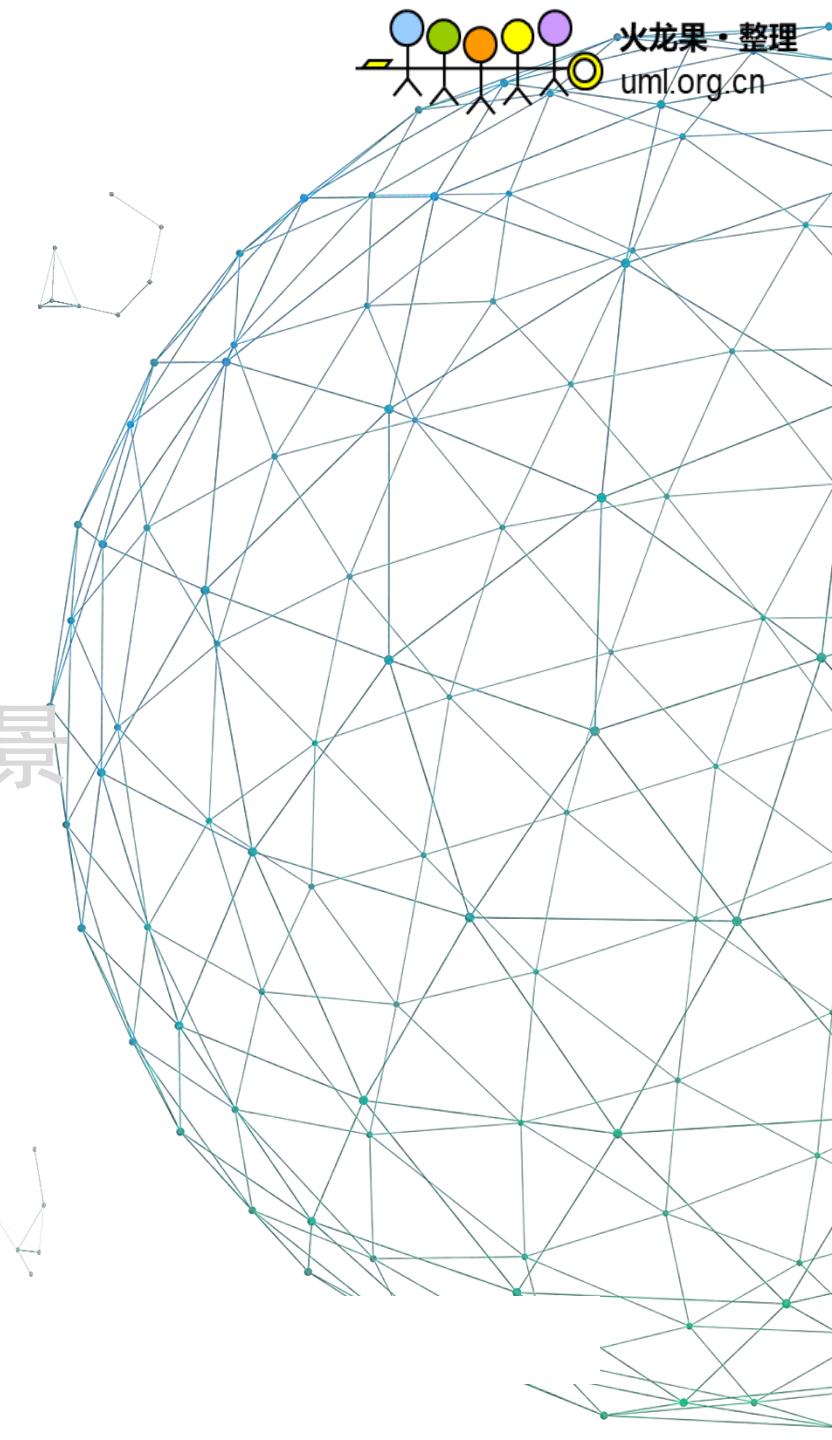


# 风控建模中的数据科学



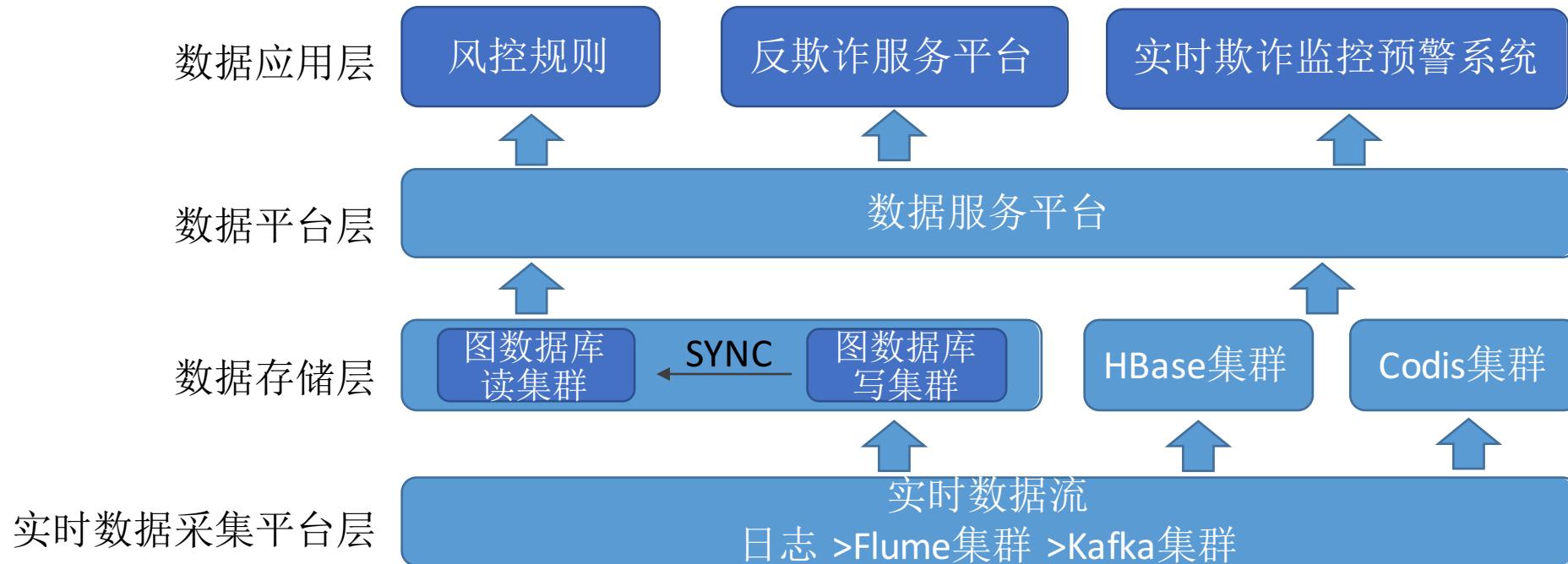
# 互联网金融风控中的数据科学

- ① 互联网金融服务面临的风险
- ② 知识图谱在金融风控中的应用场景
- ③ 大规模金融服务中的实时架构

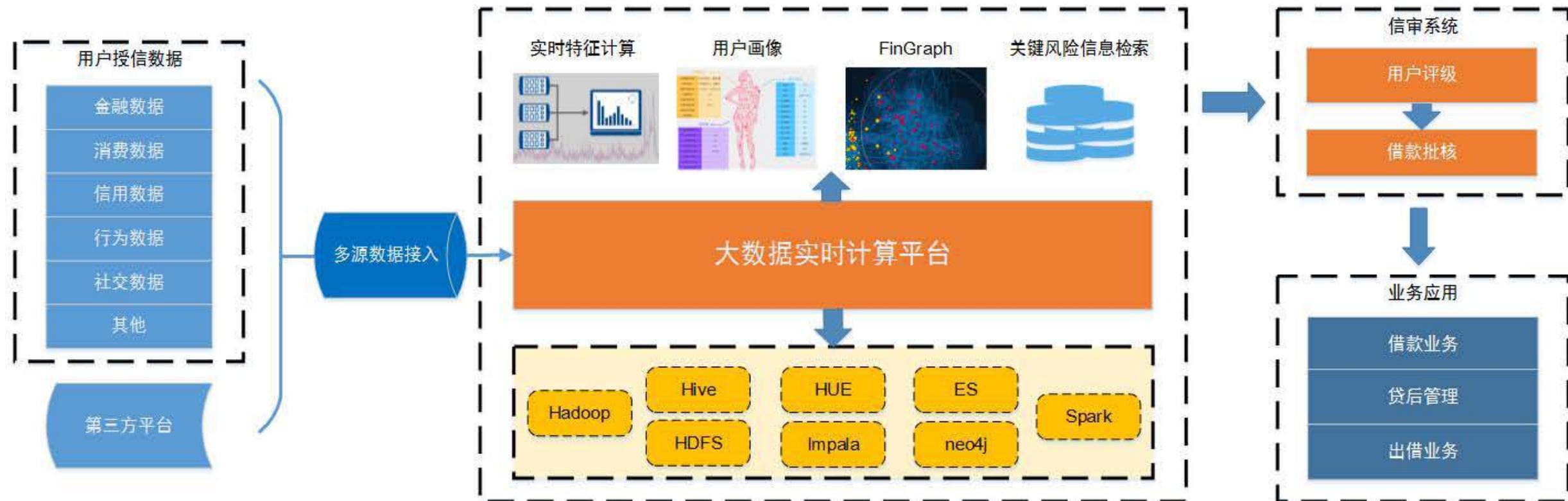


# FinGraph是实时风险监控的重要支撑

- 基于实时数据采集平台和图数据库，可实时捕捉风险特征，控制风险
  - 数据在采集、传输、存储时达到99.999%的可靠



# FinGraph是线上风控系统中的关键一环



# 总结：数据科学在互联网金融风控中发扬

互联网金融  
风控工作  
是一种机器  
学习过程

实时风险控制  
在线上P2P借款与理财  
咨询服务行业中，是全  
自动信用评估的关键

图挖掘技术  
可以把风控工  
作，从局部考量  
提升到全局考量

数据科学  
已逐渐渗透到  
离线建模流程  
中的多个环节

## FinGraph 知识图谱平台

- 结合了上亿节点和十亿关联关系提供全面风控能力
- 用户社交网络关系特征可以分辨出3倍欺诈风险的人群
- 用户的组团欺诈风险可以通过社区发现算法实时评估
- 失联用户也可以通过图挖掘算法来找寻

FinGraph  
已逐渐成为线  
上风控系统中  
的关键一环



# Thank You !

