



七牛云
QINIU

IT大咖说
知识分享平台

点融网
Dianrong.com

大数据技术在点融网业务 的应用介绍

刘利

点融数据团队



目录

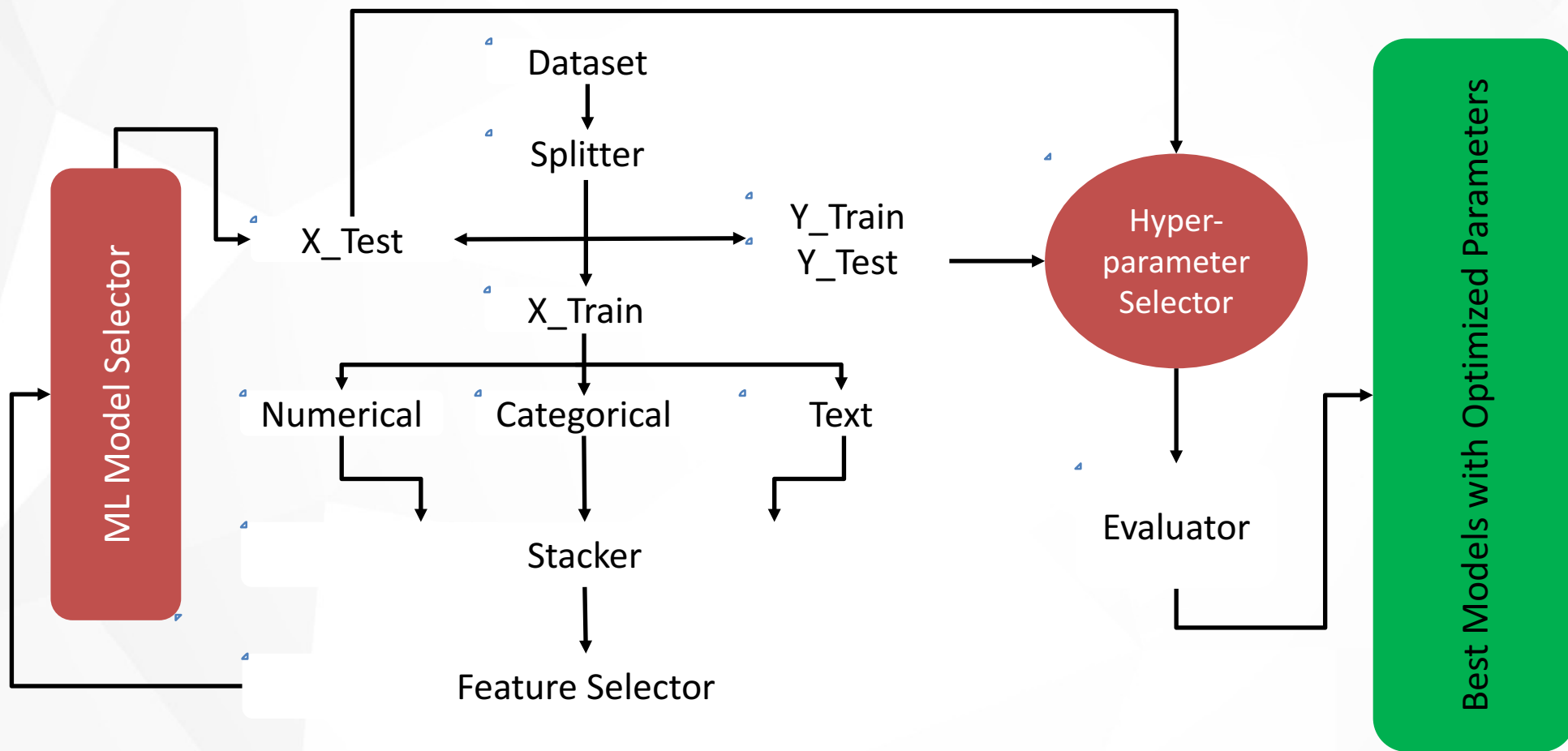
- 点融机器学习平台
- 风控业务案例分析
- 如何提升模型性能

点融简介



- 6年前成立于上海，拥有 Lending Club 的技术平台
- 获得Tiger Global, Standard Chartered Private Equity 等多家知名风投基金
- 在国内有28个办事处，超过2600名员工
- 发起了中国首个区块链平台
- 目前，总用户投资金额已经超过290亿RMB，对用户的利息回报已经超过10亿RMB

机器学习一般流程



已有的解决方案

- 收费
- 数据安全
- 数据可视化
- 分布式
- 模型结果部署

点融机器学习平台



spark-submit

Spark
Master
JVM

Spark
Worker
JVM

Spark
Worker
JVM

Spark
Worker
JVM

Dianrong ML Cluster

Spark
Executor
JVM

Dianrong
ML

Spark
Executor
JVM

Dianrong
ML

Spark
Executor
JVM

Dianrong
ML

Dianrong App
jar file

支持读取hdfs数据



variable_importance_demo

importFiles

Import Files

Search: Enter a file or directory path and press the Enter key

Selected Files: (No files selected)

Actions:

```
importFiles [ "hdfs://master-1f1.c.oss-cn-gz.com:8020/user/oper/german.csv" ]
```

1 / 1 files imported.

Files

Actions:

数据可视化

Data Visualization: german.hex

PAIR PLOTS

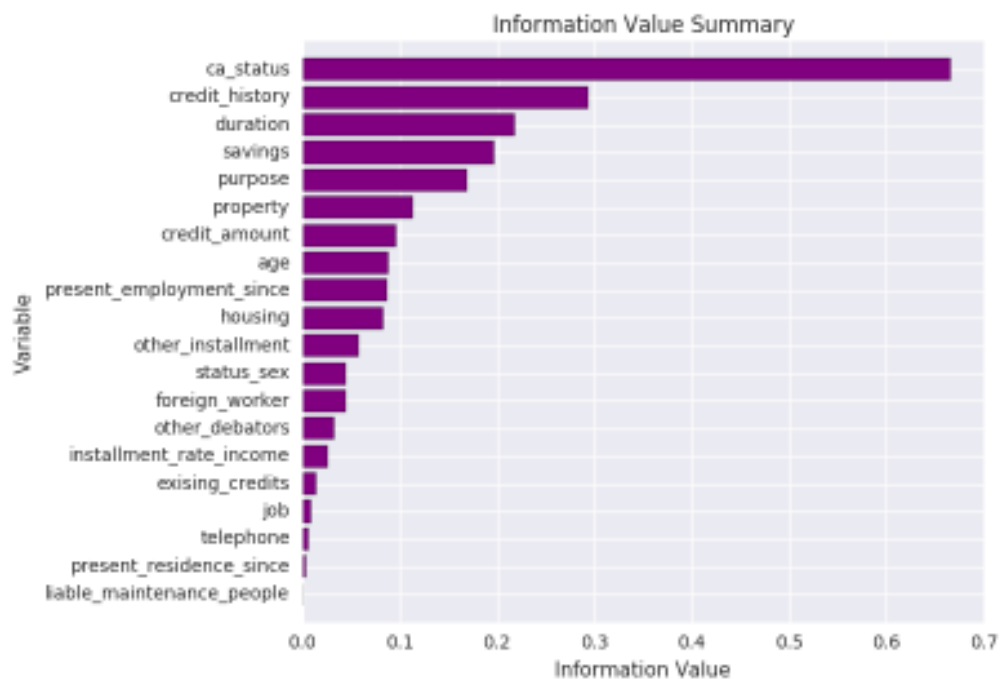


特征重要性排序

```
CS informationValue "german.hex", {"label":"gb","index":20}
```

Feature Importance: german.hex

INFORMATION VALUE RANKING



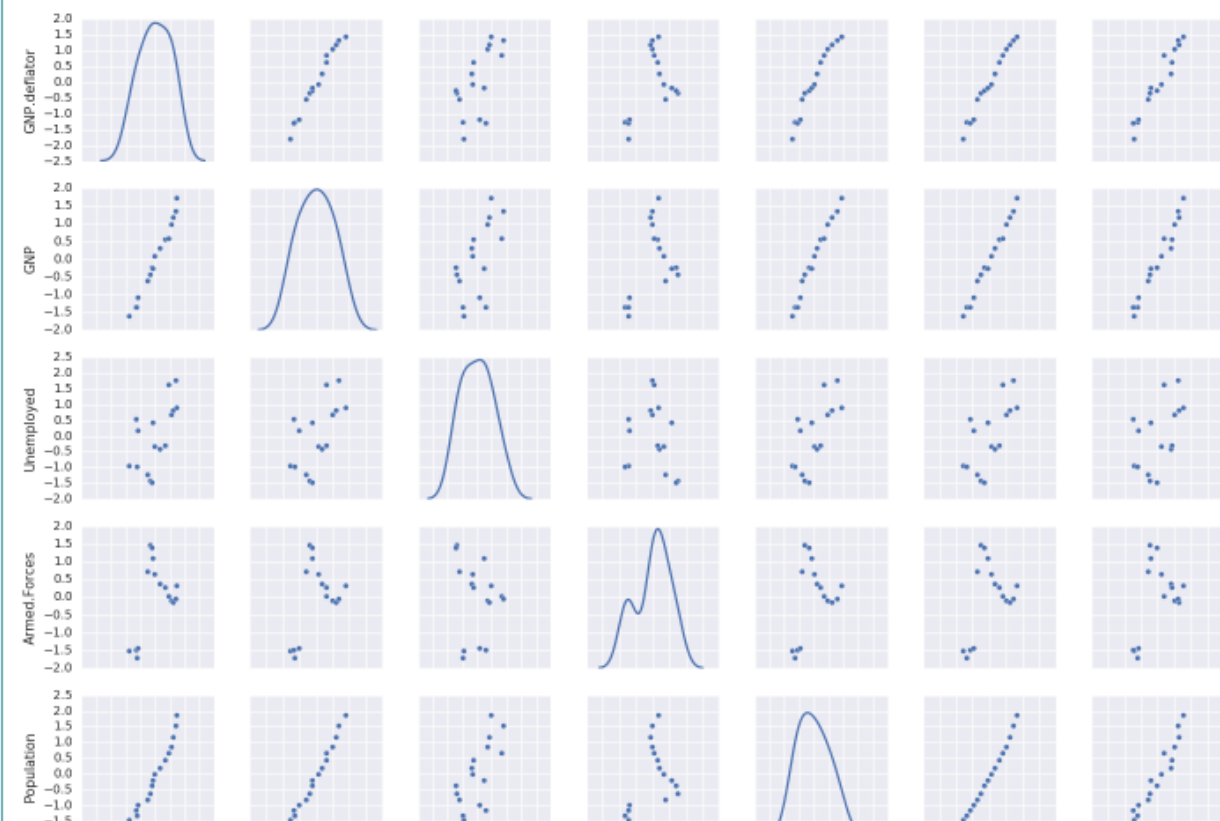
共线性分析

collinearity_analysis_demo



Collinearity Analysis: longley.hex

COLLINEARITY



模型库

```
CS buildModel "deeplearning"
```

Build a Model

Select an algorithm:

- Deep Learning
- Deep Learning
- Distributed Random Forest
- Gradient Boosting Machine
- Generalized Linear Modeling
- Generalized Low Rank Modelin
- K-means

PARAMETERS

validation_frame (Choose...)

nfolds 0

response_column (Choose...)

3d041-4cad-4f2b-9284-25d70a0b0f39

模型部署

Publish Model

Model:

Overwrite:

Actions:

```
CS publishModel "deeplearning-266c6485-581e-48a2-99ed-c890a9fd6fd8", overwrite: true
```

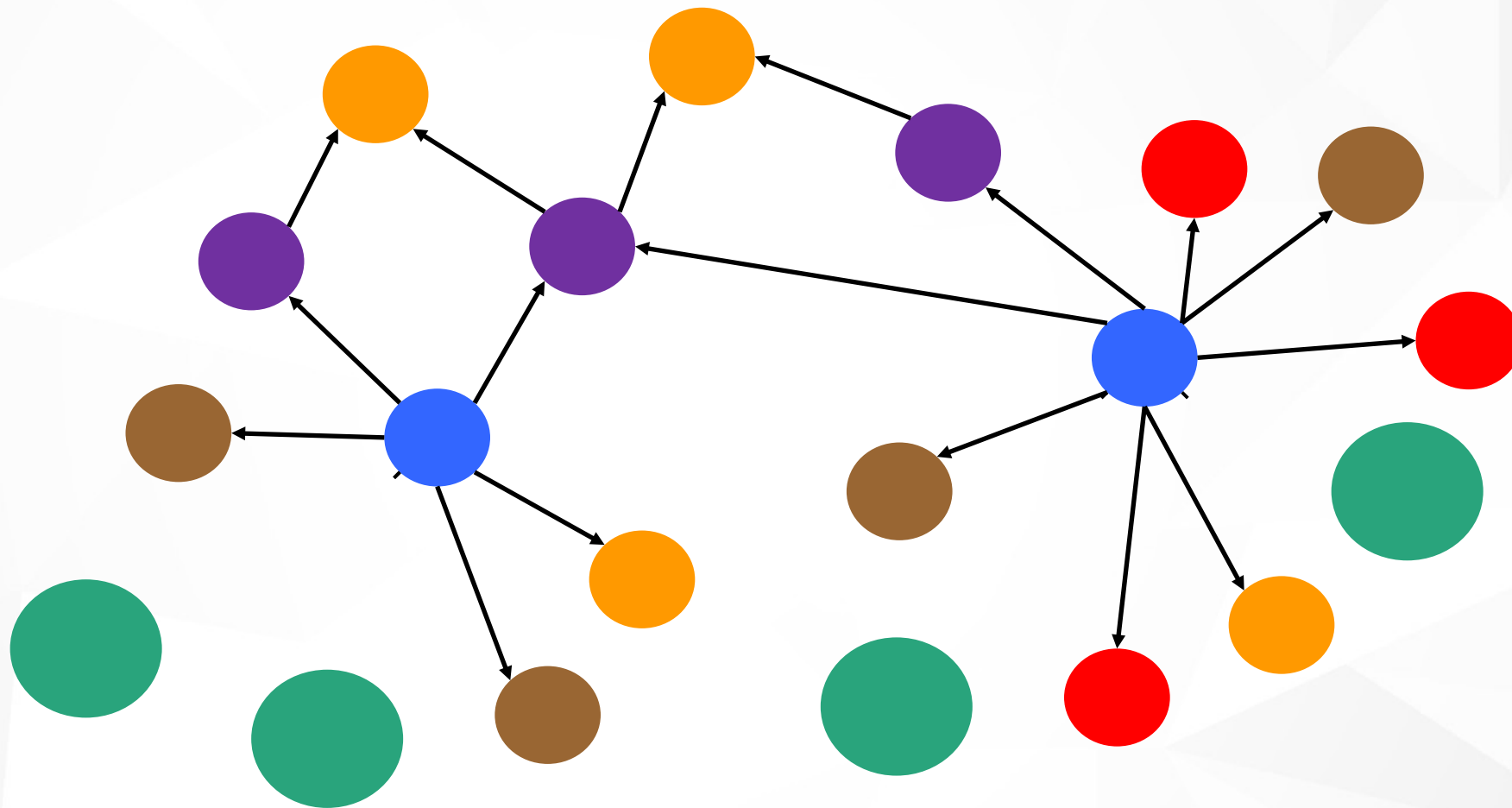
Model Published

The specified model was published successfully. You can get the model score for your data by:

<http://0.1x.cdn.com/score/deeplearning-266c6485-581e-48a2-99ed-c890a9fd6fd8>

风控业务案例分析



- Bankcard
- Company
- Person
- Email
- Loan
- Phone

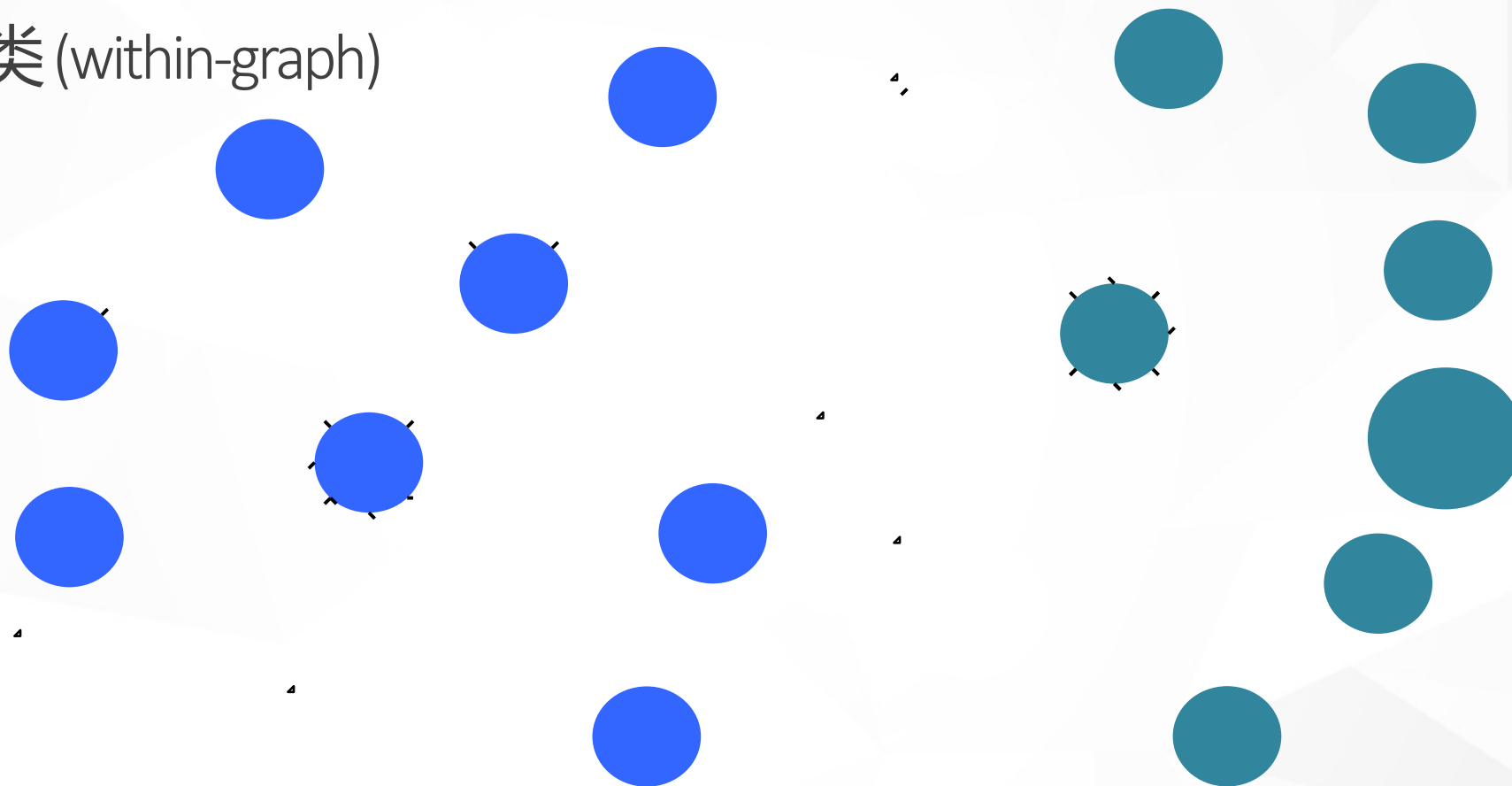


Graph Mining 在风控领域的应用

基于点的分类 (within-graph)

例: SNA

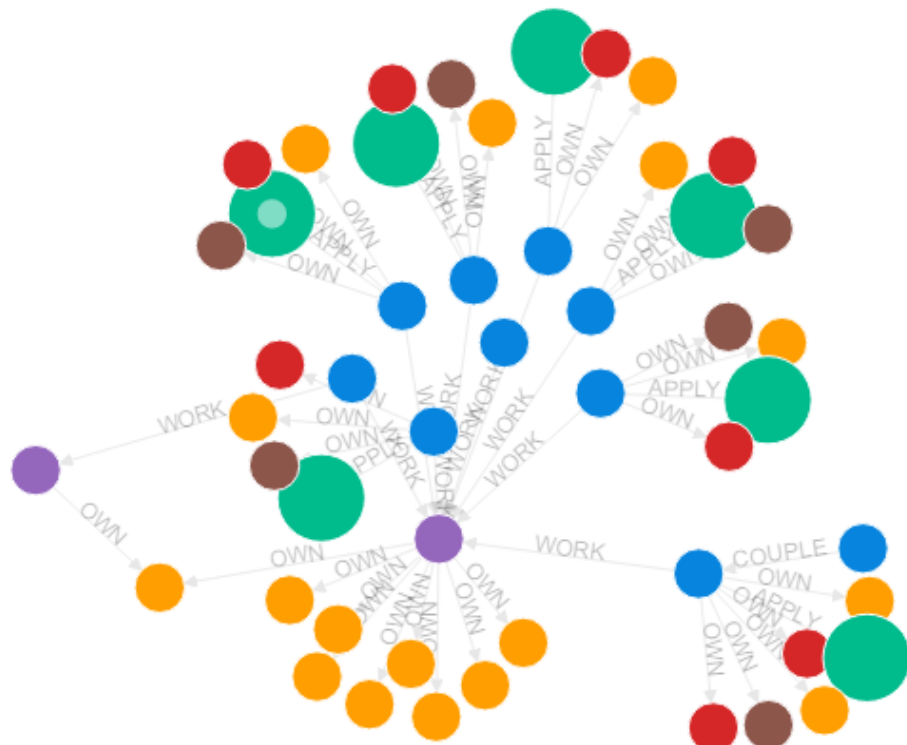
 GOOD
 BAD



Graph Mining 在风控领域的应用

基于图的分类 (between-graph)

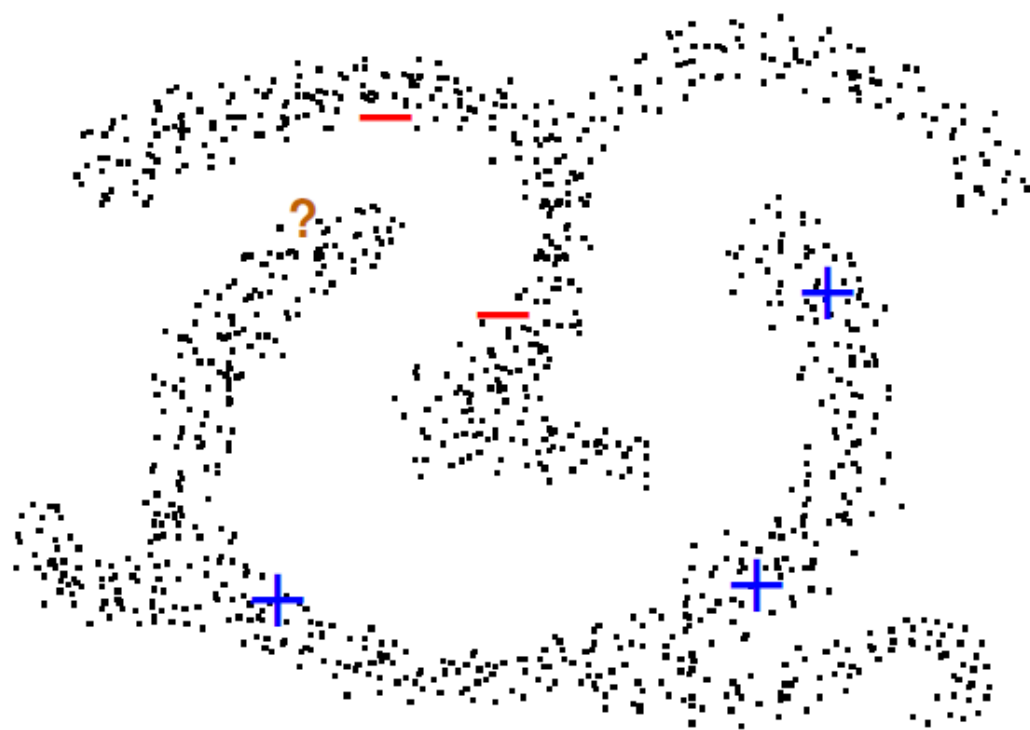
例: X度人脉



Graph Mining 在风控领域的应用

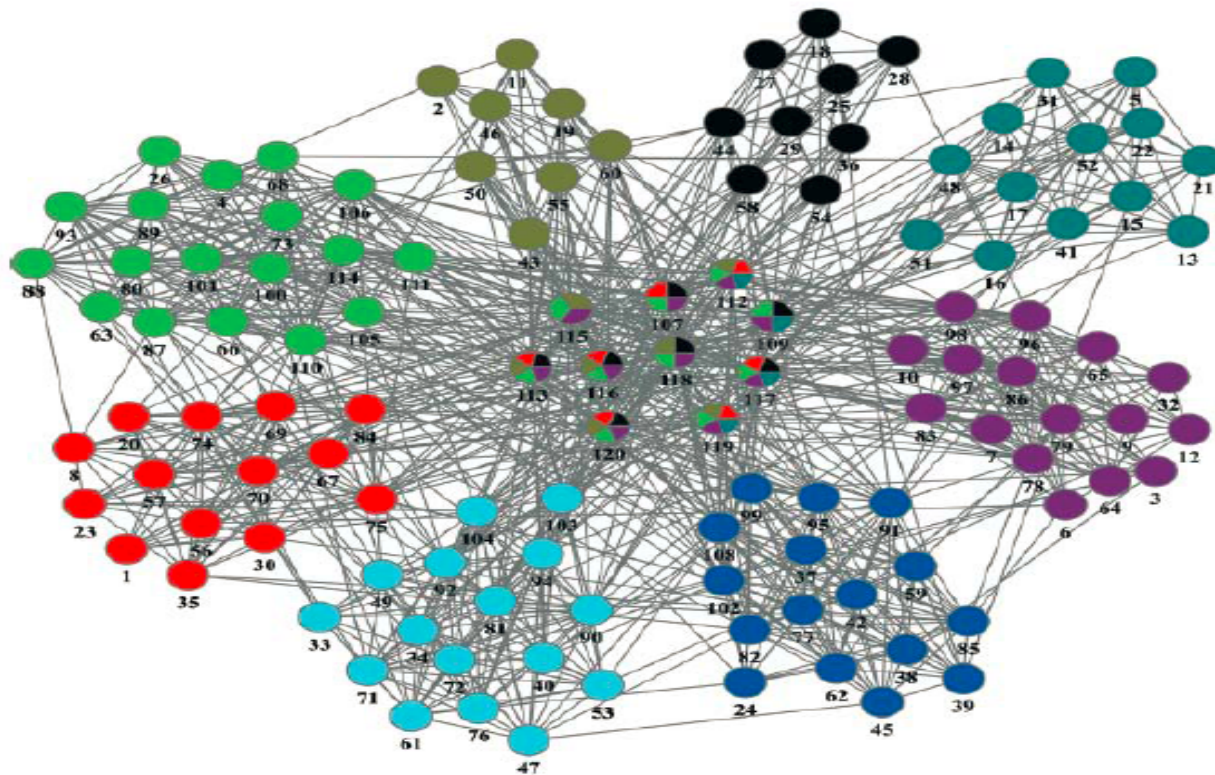
半监督学习

- 1) Smoothness 平滑假设
- 2) Cluster 聚类假设
- 3) Manifold 流形假设



Graph Mining 在风控领域的应用

社区发现



如何提升模型性能

- 基于数据
- 借助算法
- 用算法调参
- 借助模型融合



Q&A

