



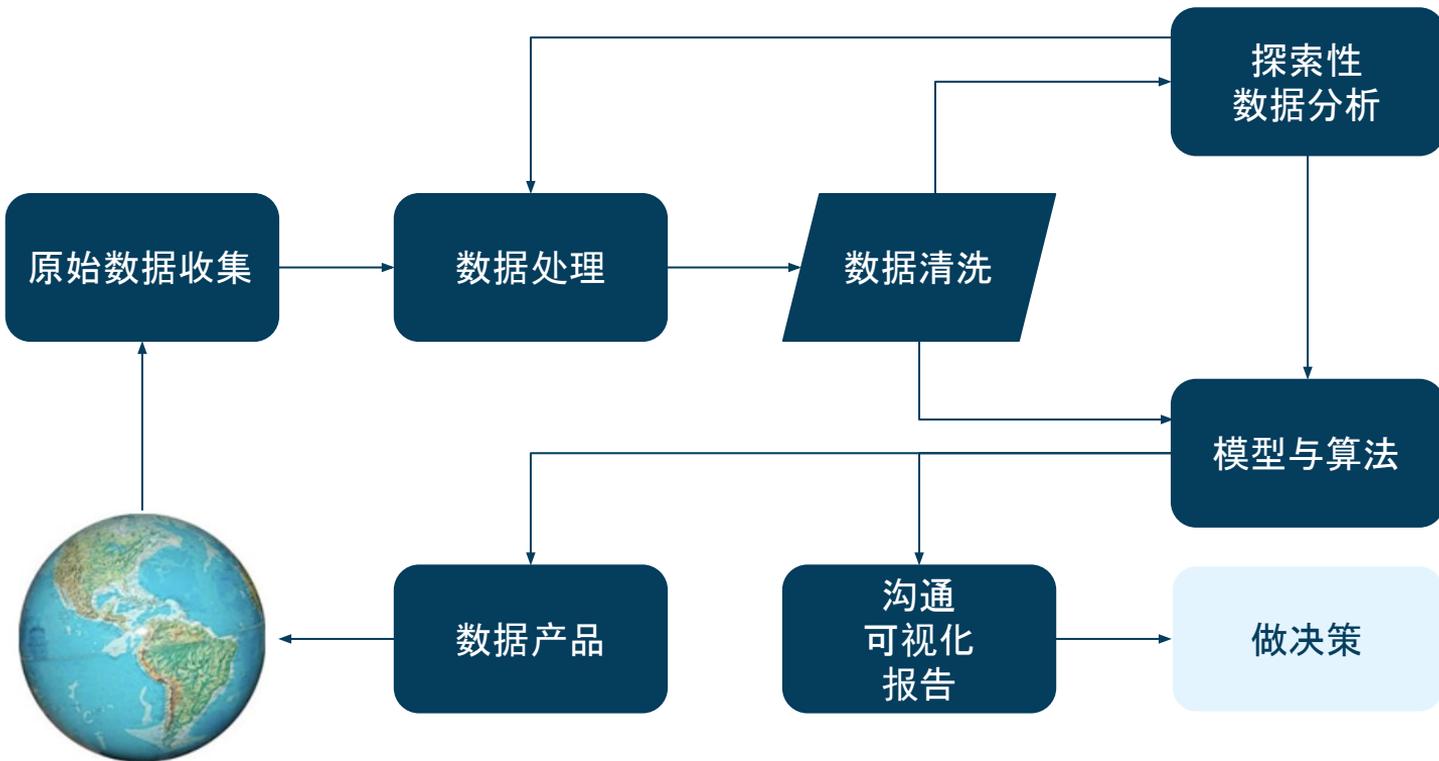
图与数据分析

林选磊

June 23, 2021

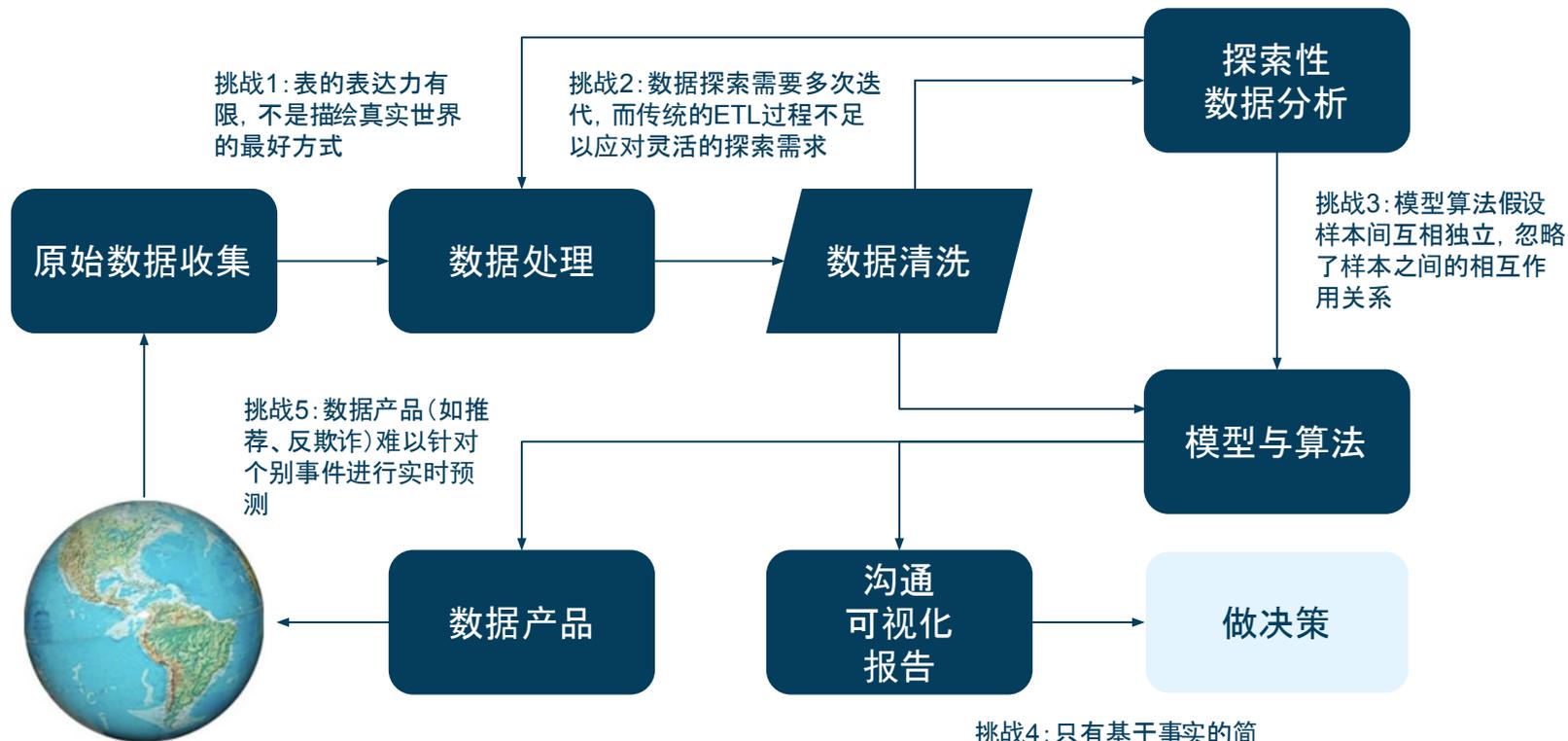


数据分析流程



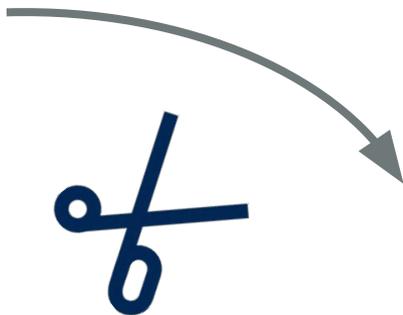
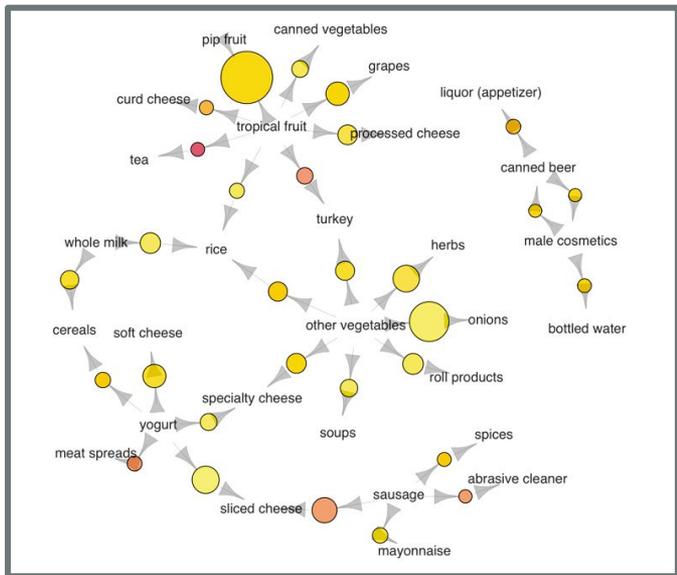
图片来源: https://en.wikipedia.org/wiki/Data_analysis

基于表思维的数据分析



挑战1: 表的表达力有限, 不是描绘真实世界的最好方式

真实世界

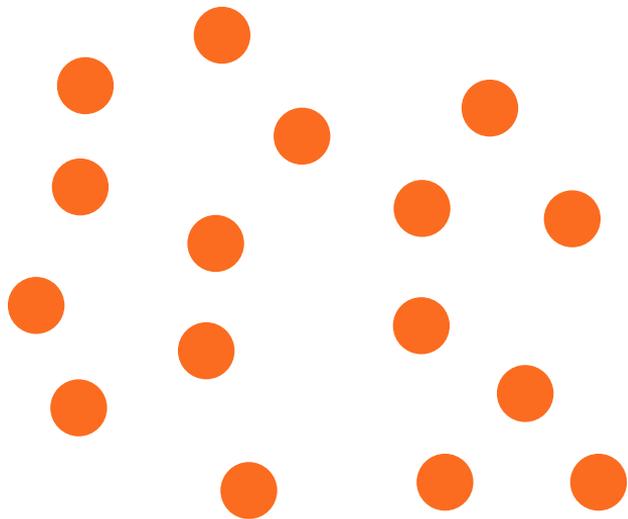


表

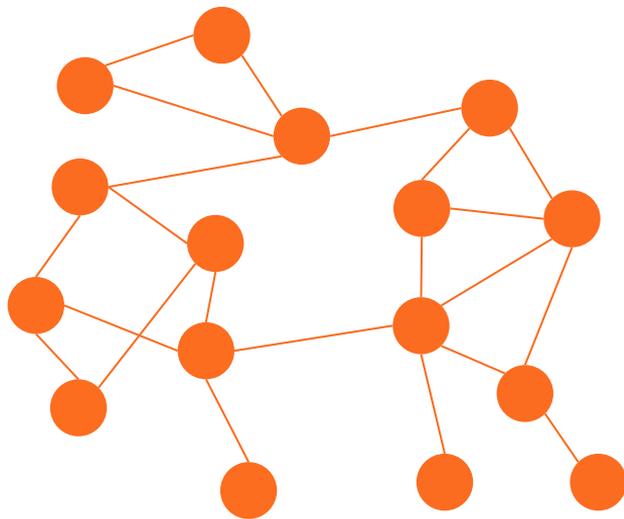
Transaction 1	   
Transaction 2	  
Transaction 3	 
Transaction 4	 
Transaction 5	   
Transaction 6	  
Transaction 7	 
Transaction 8	 

使用表来表达真实世界容易丢失事物间的关联信息

什么是图？

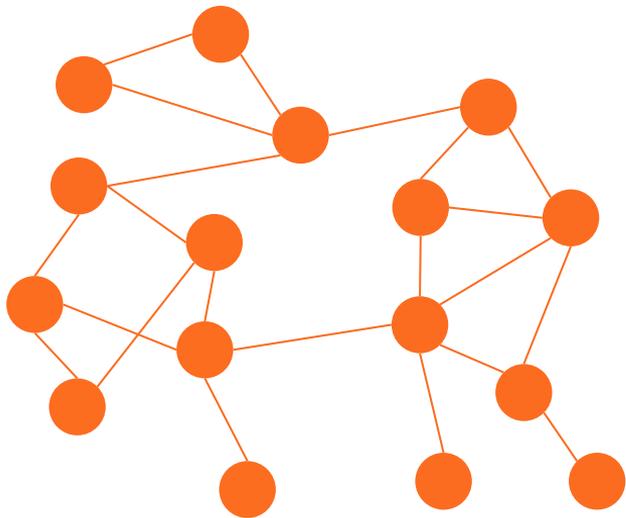


传统分析: 假设实体间互相独立

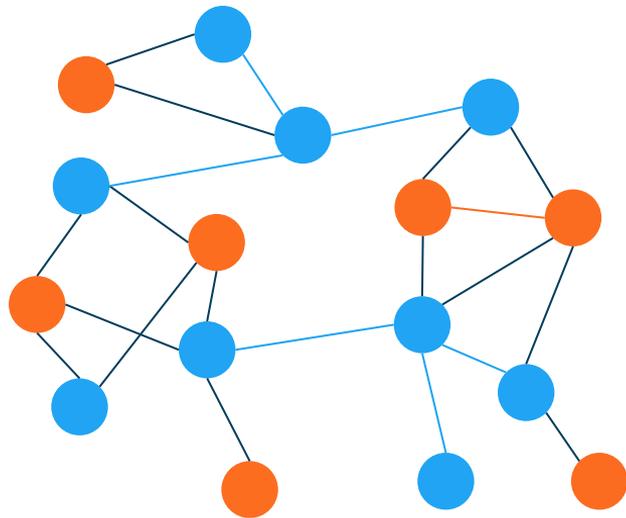


图分析: 探索实体之间的关系

同构图与异构图

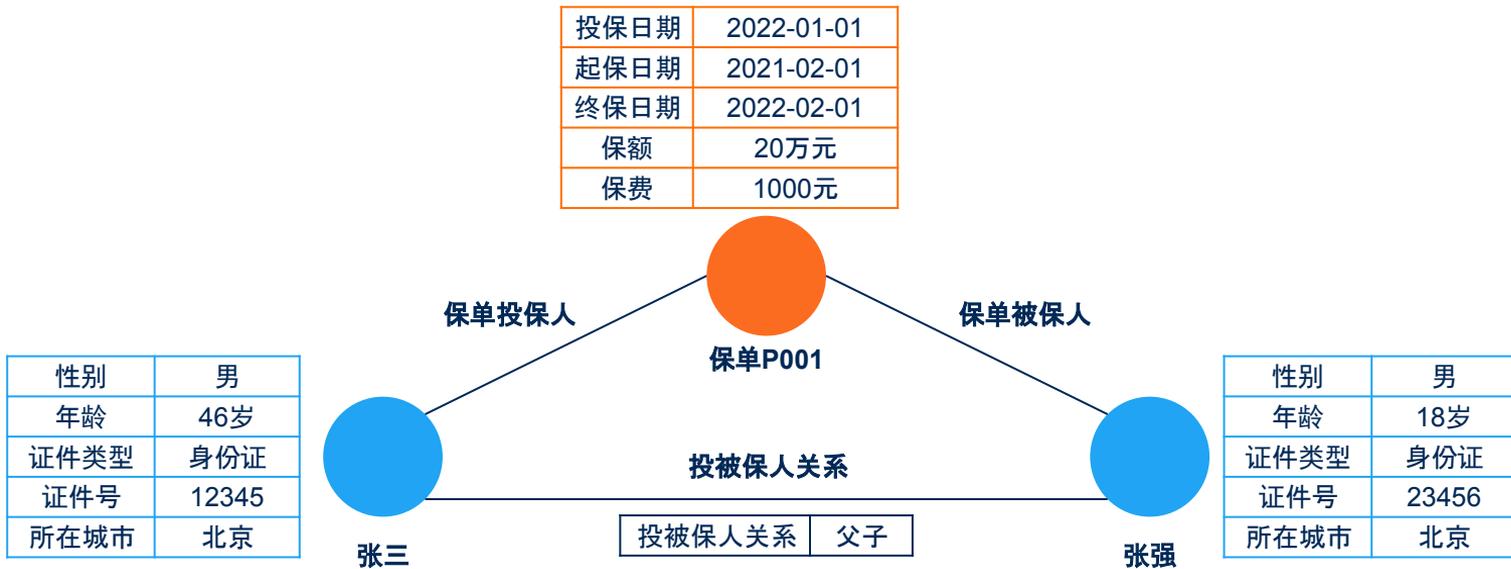


同构图: 顶点只有一种类型, 边也只有一种类型(如社交网络)

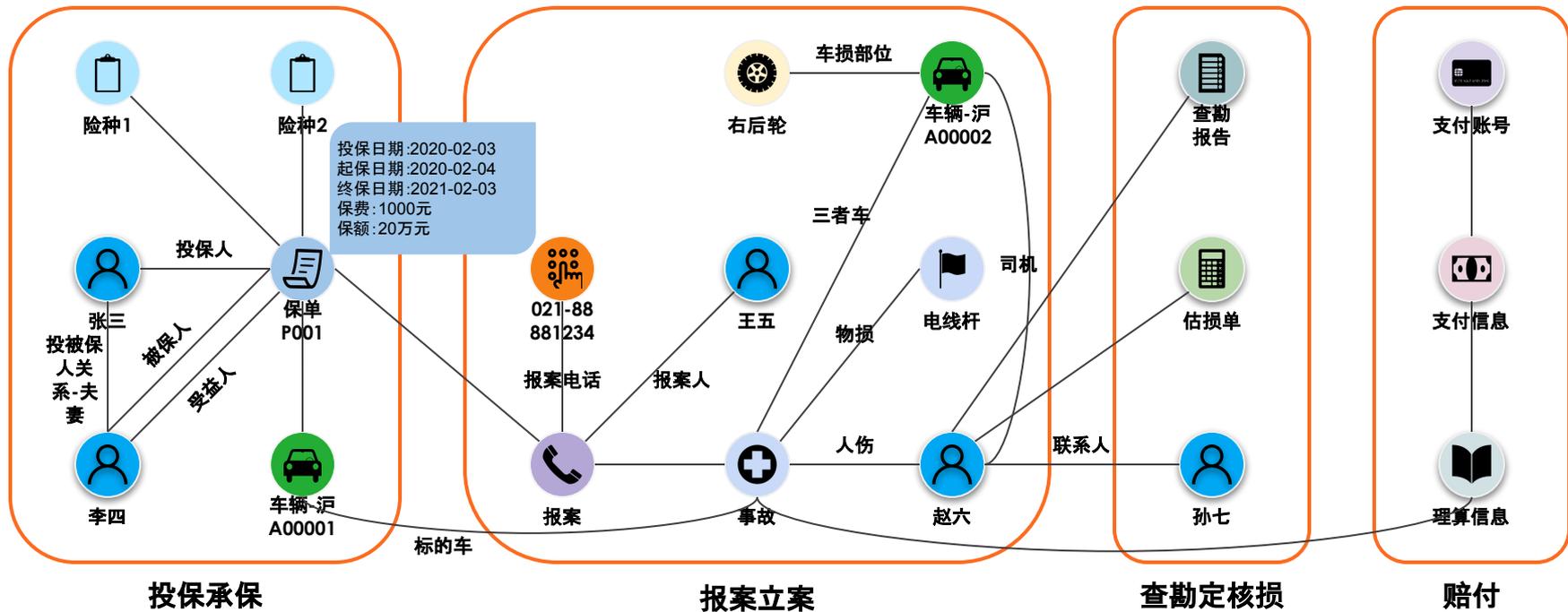


异构图: 顶点可能有多种类型, 边也有多种类型(如知识图谱)

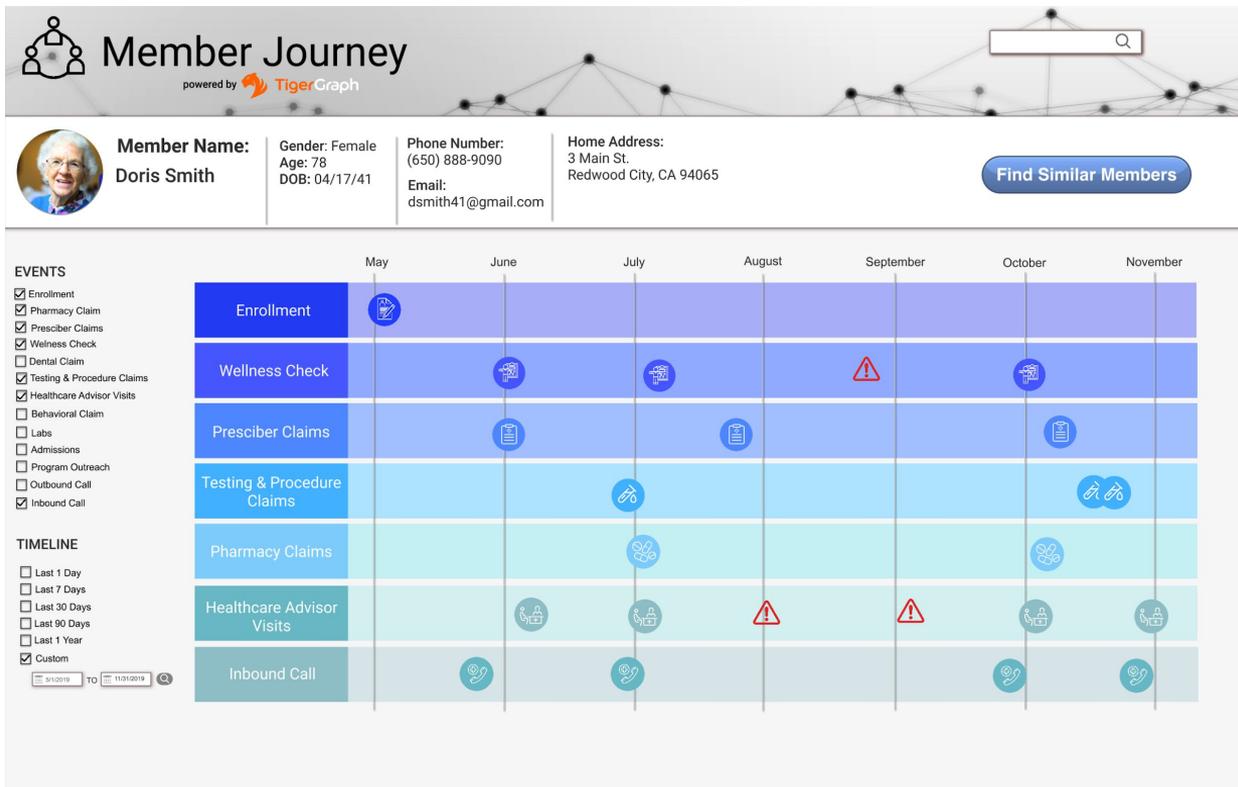
属性图



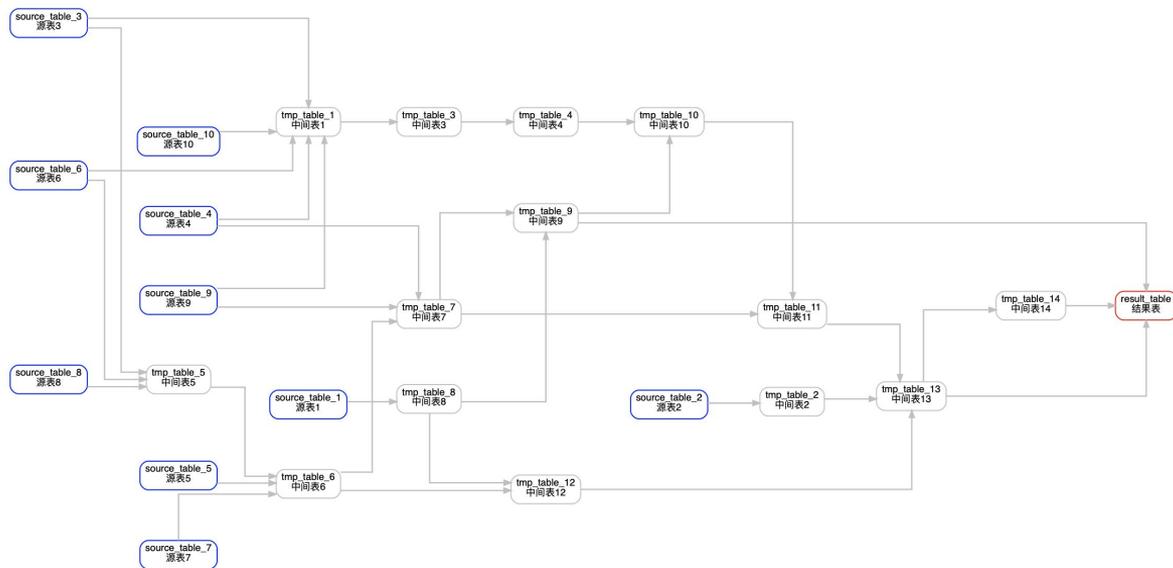
图的强大表达力



客户实时画像



挑战2: 传统的ETL过程不足以应对灵活的探索需求



传统的ETL从需求到落地消耗大量时间, 一点需求变更就要修改大量代码

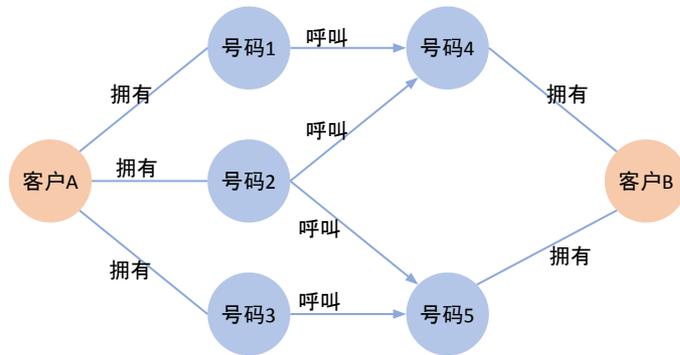
图 v.s. 表

客户	号码
客户A	号码1
客户A	号码2
客户A	号码3
客户B	号码4
客户B	号码5

表
关联

主叫客户	主叫号码	被叫号码	被叫客户
客户A	号码1	号码4	客户B
客户A	号码2	号码4	客户B
客户A	号码2	号码5	客户B
客户A	号码3	号码5	客户B

主叫号码	被叫号码
号码1	号码4
号码2	号码4
号码2	号码5
号码3	号码5



- 表关联非常消耗时间和资源
- 表忽略了业务关联关系
- 表作为AI的输入时, 有着样本间互相独立的错误假设, 只提供了样本的属性值

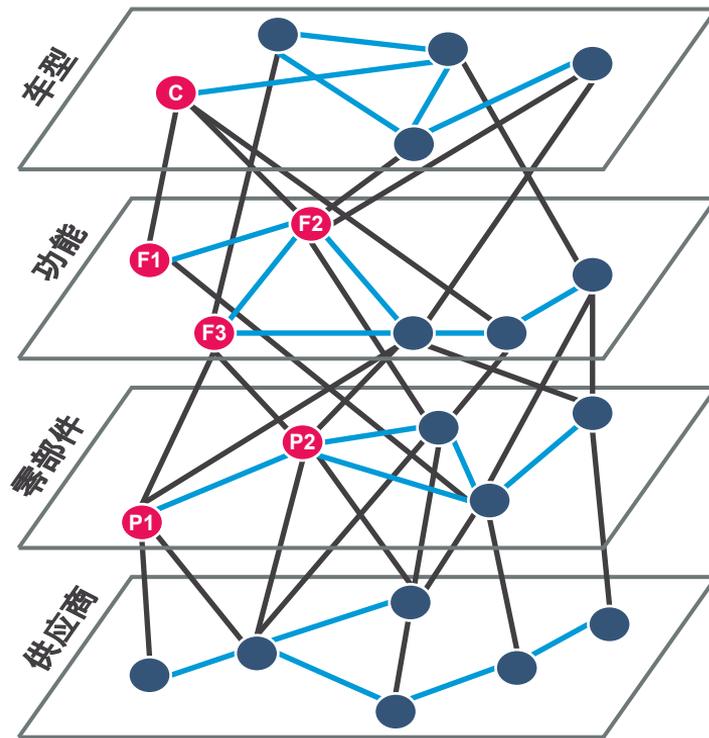
- 图因为事先已关联数据, 可进行实时查询
- 图有极强的业务表达力
- 图作为AI的输入时, 完善了信息, 为准确预测提供基础

用图加快汽车制造商的决策速度

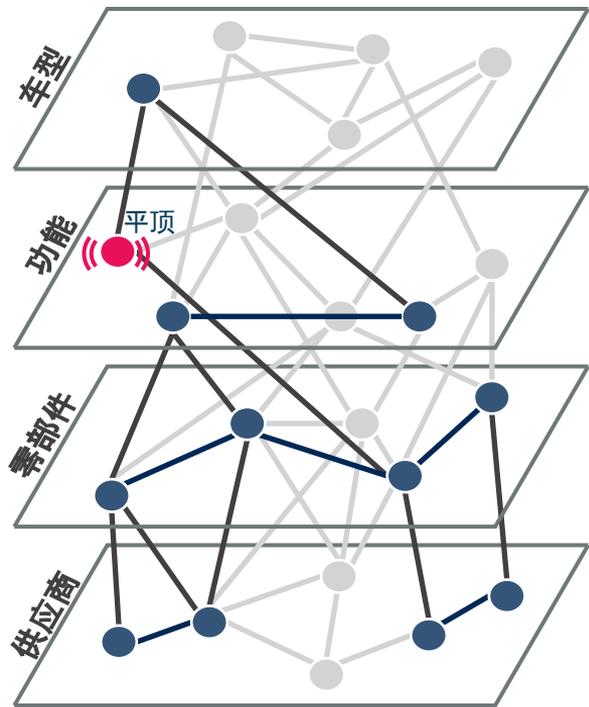
- 每辆汽车拥有大约有4500个零部件
- 有些系列的汽车共有超过3万个零部件可提供用户定制



“来自图的降维打击”



平顶版揽胜极光

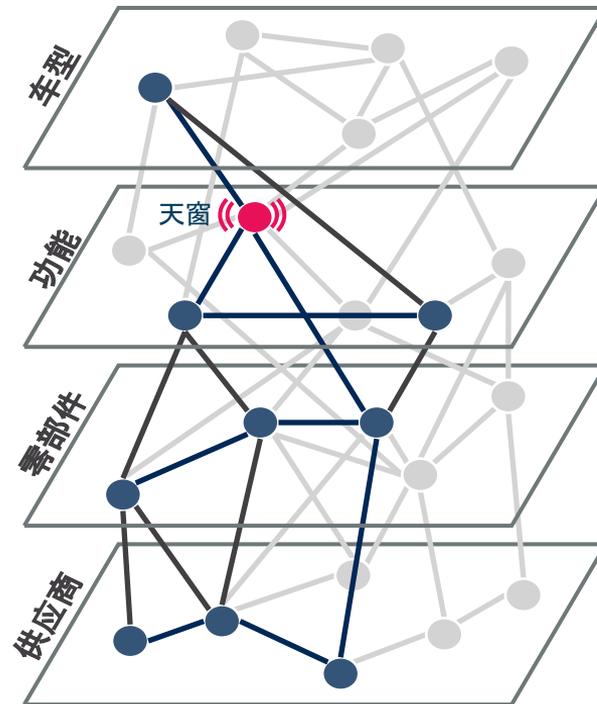


▲ 营收影响

▼ 成本影响



天窗版揽胜极光



45分钟

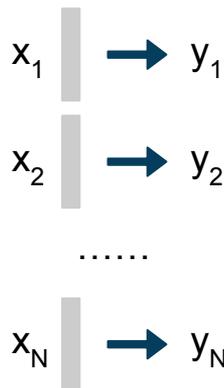


3周

挑战3: 模型和算法假设样本间互相独立

- 训练模型

- 逻辑回归
- SVM
- 神经网络
-



- 应用模型

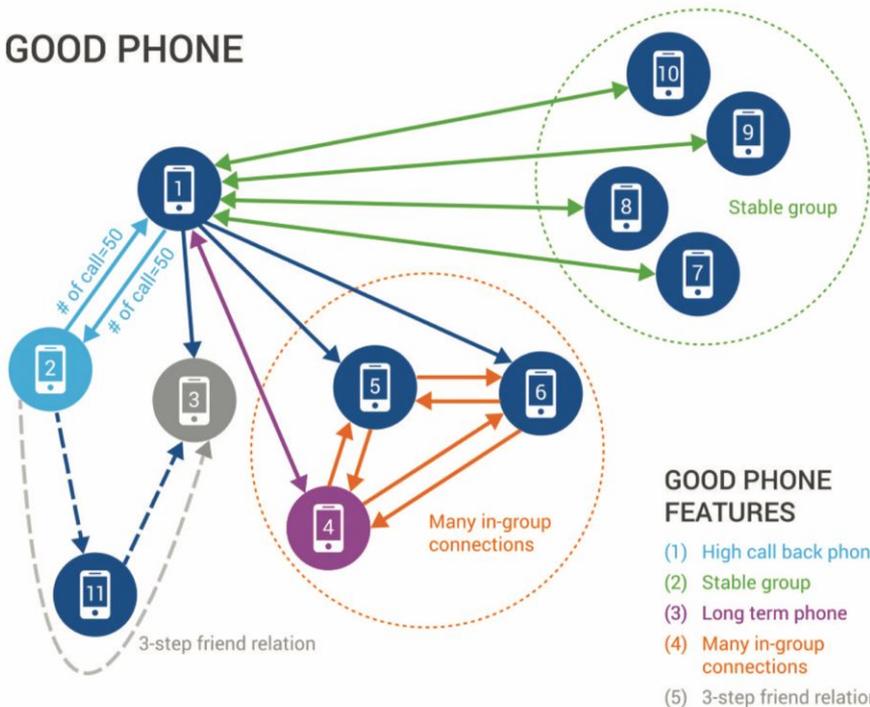
- 给定新数据, 获取其特征并使用训练好的模型进行预测



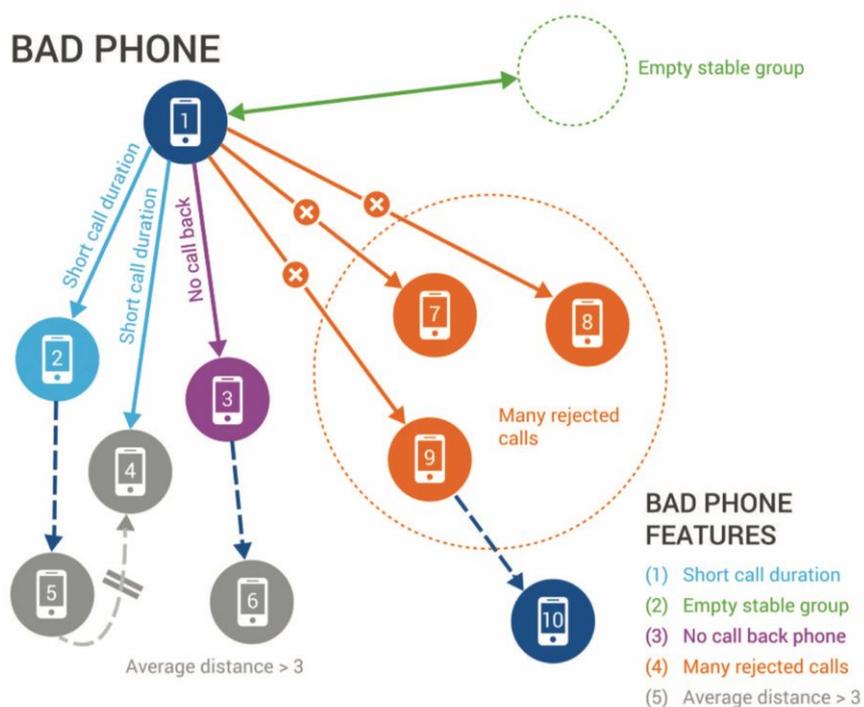
传统的机器学习方法忽略样本与样本之间的相互作用关系

图特征比常规特征更重要

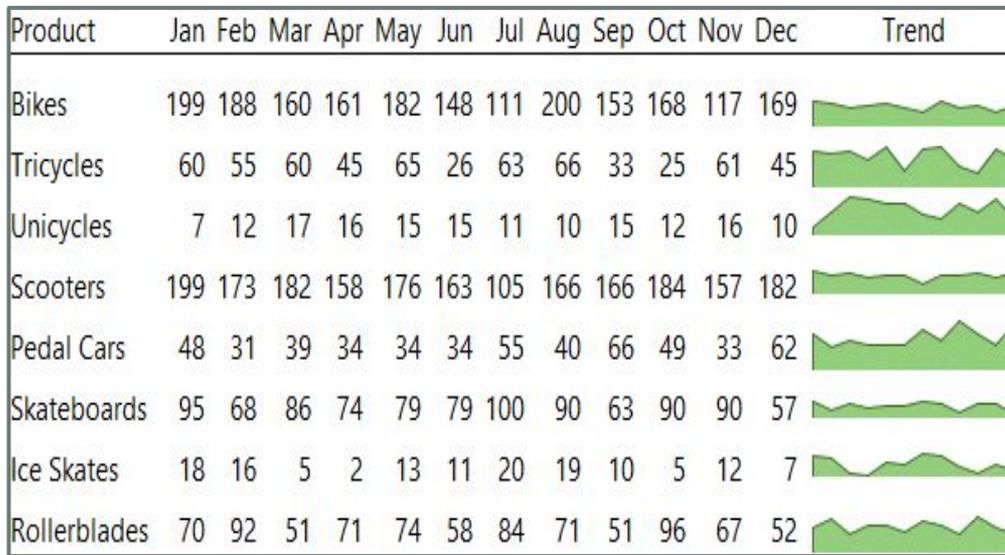
GOOD PHONE



BAD PHONE

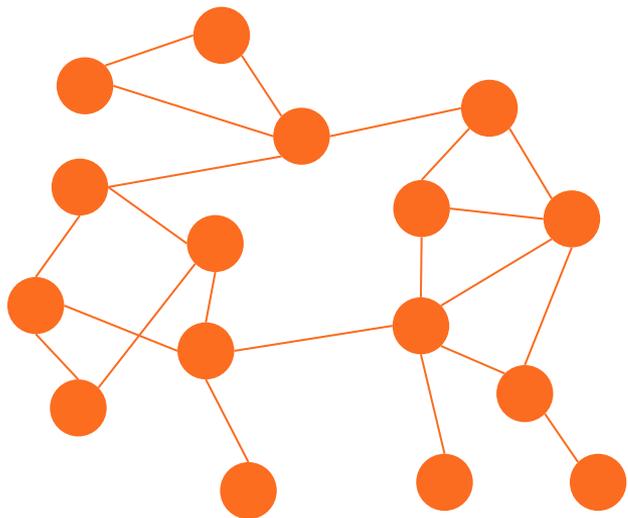


挑战4: 缺少解释事情发生原因的能力

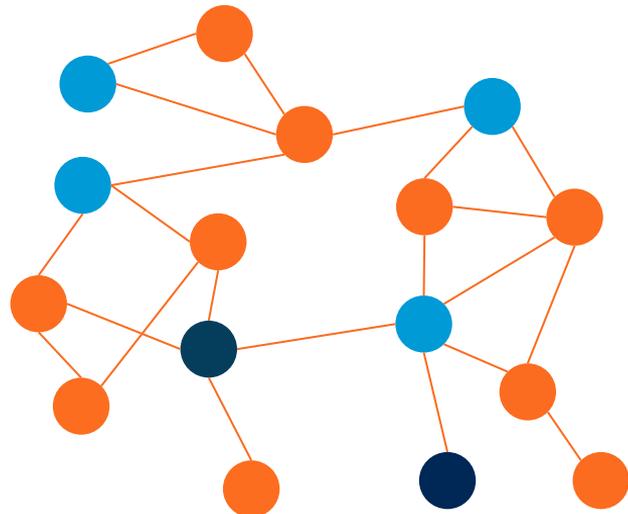


传统的BI分析基于事实进行简单统计, 却未能揭示事实背后丰富的因果链条

商户社区

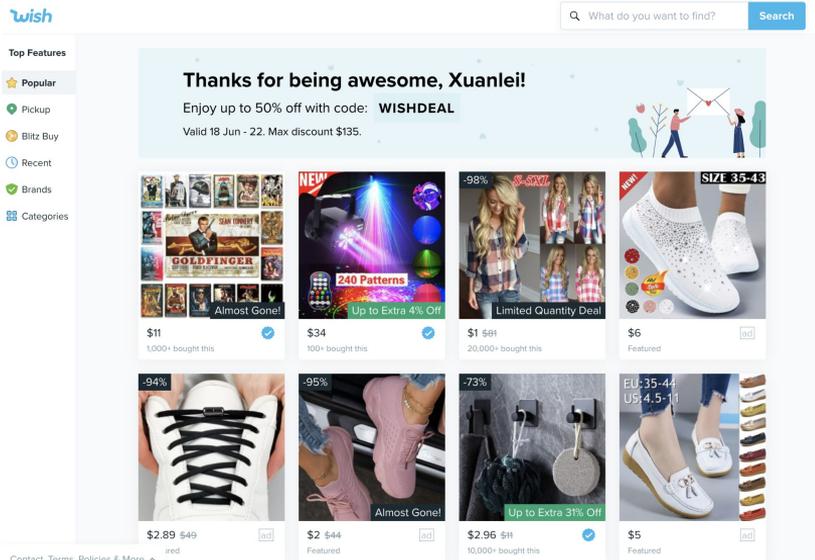


正常的商户应该都在同一商圈



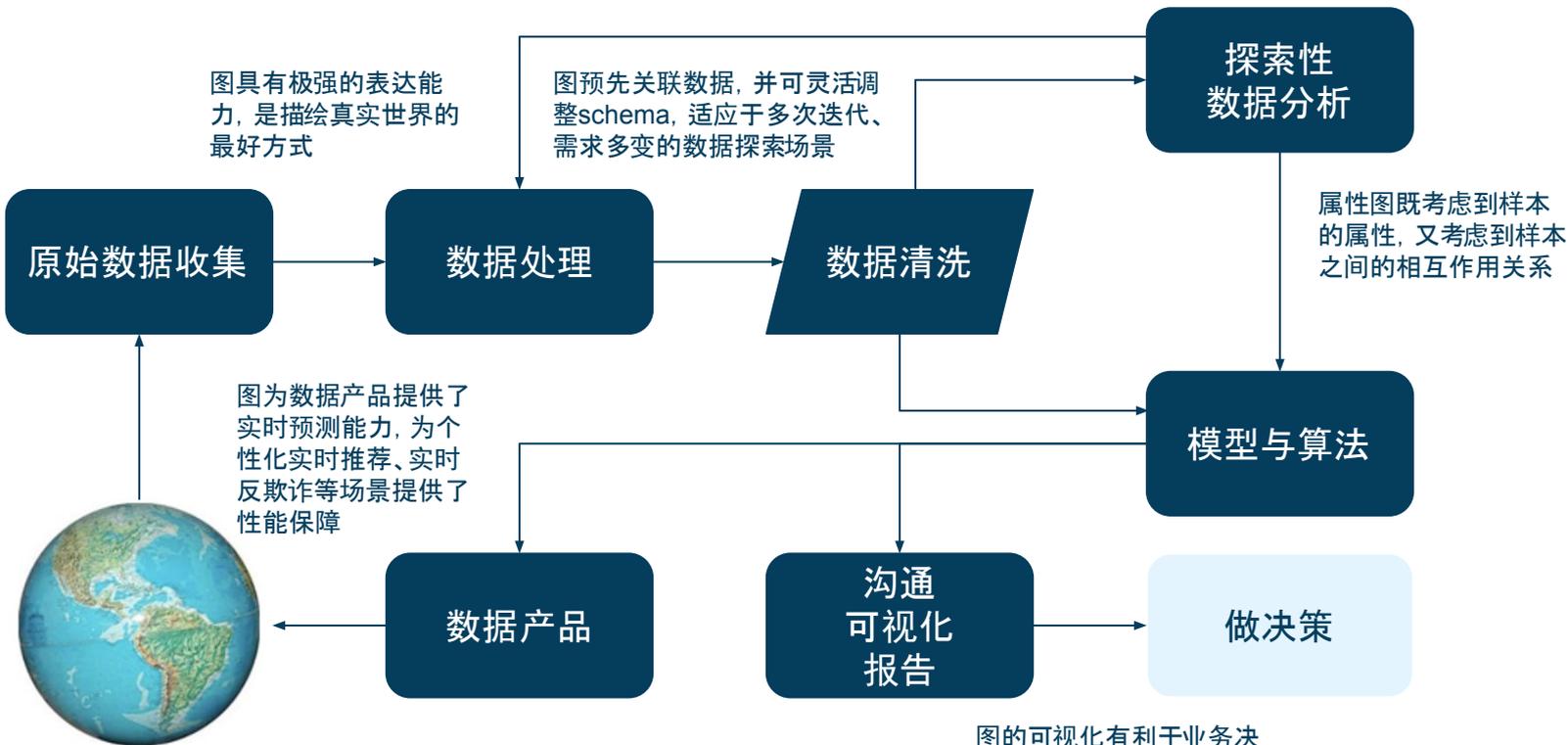
欺诈团伙的商户会分散在地理位置不同的商圈中

挑战5: 数据产品难以针对个别事件进行实时预测



推荐、反欺诈等数据产品都需要针对当前事件进行个性化的实时预测

基于图思维的数据分析



对TigerGraph感兴趣，想了解更多吗？

不管你是开发者，还是业务或管理人员，都可以从这里找到想要的内容。

我想试用TigerGraph:

- ➔ 下载免费企业版(容量<50G, 社区支持)
- ➔ 企业版试用(容量>50G, 原厂服务支持)
- ➔ Test Drive – 在线DEMO演示
- ➔ 免费开始云应用tgcloud.io
- ➔ 20多个入门套件及DEMO
- ➔ 预约现场交流
- ➔ 成为合作伙伴

我想了解更多的资料:

- ➔ 为什么要用图分析?
- ➔ Why TigerGraph?
- ➔ 成功案例
- ➔ 解决方案
- ➔ Gartner报告
- ➔ Graph + AI峰会
- ➔ 免费下载《图数据库购买者指南》

我是开发者，希望学习使用:

- ➔ 图课堂
- ➔ 开发者社区
- ➔ 技术文档
- ➔ 认证培训&报名
- ➔ 基准测试报告
- ➔ 免费下载《原生并行图》白皮书
- ➔ 免费下载O'Reilly抢先版

*关注微信公众号，输入“图思维”获得完整版演讲课件；

*以上链接请访问：TigerGraph.com.cn/link/



TigerGraph

© 2021. ALL RIGHTS RESERVED. | TIGERGRAPH.COM



扫/码/关/注