

The logo for DTCC (Database Technology Conference China) features the letters 'DTCC' in a bold, orange, sans-serif font. A thick orange horizontal line is positioned below the letters, curving slightly upwards at the ends. The background of the entire image is light gray with various technical and data-related graphics, including a large circular network diagram on the left, a DNA double helix in the center, and a city skyline silhouette at the bottom left.

第十一届中国数据库技术大会

DATABASE TECHNOLOGY CONFERENCE CHINA 2020

架构革新 高效可控

IT168.com logo featuring a stylized red and blue 'IT' icon followed by the text '168.com'.

ChinaUnix.net logo featuring a blue circular icon with a white 'U' and the text 'ChinaUnix.net'.

ITPUB logo featuring the text 'ITPUB' in a bold, black, sans-serif font with a red underline.



北京国际会议中心 |



2020/12/21-12/23

HugeGraph图数据库实践2.0

典型图算法的应用

百度 李章梅





目录

1. 图数据库概念
2. HugeGraph典型图算法实践
 - 1) 知识问答
 - 2) 协同推荐
 - 3) 关系挖掘
 - 4) 循环担保
 - 5) 群控作弊
 - 6) 团伙欺诈
3. HugeGraph优势

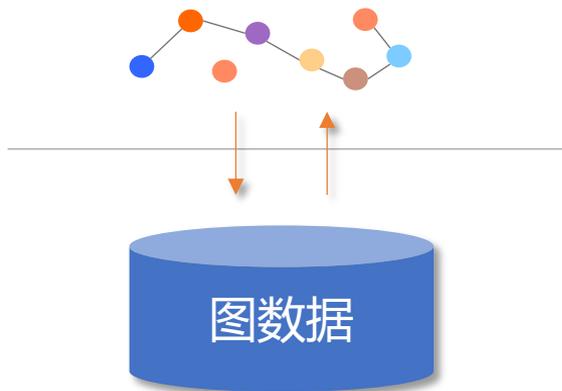


目录

1. 图数据库概念
2. HugeGraph典型图算法实践
 - 1) 知识问答
 - 2) 协同推荐
 - 3) 关系挖掘
 - 4) 循环担保
 - 5) 群控作弊
 - 6) 团伙欺诈
3. HugeGraph优势

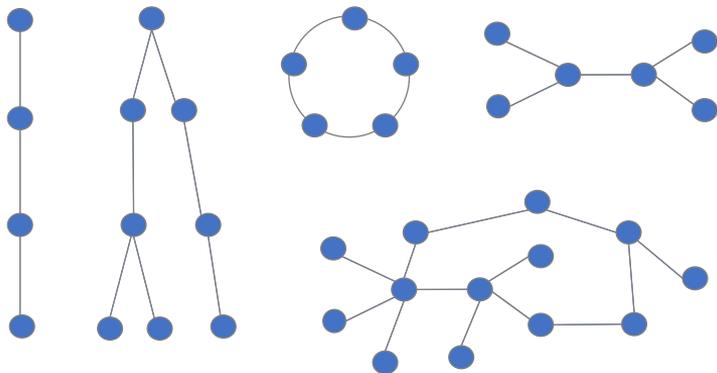
图数据库是什么？

- 存储、查询/分析图数据
- 灵活、高效处理图数据



图数据是什么样的数据？

- 由实体和关系组成的线、树、
- 环、叉、网等各种形状的数据



HugeGraph是什么？

百度自研图数据库

国内首家开源图数据库

HugeGraph

Apache 2 License

Apache Tinkerpop Enabled

大规模分布式图存储与图计算

HugeGraph应用分类

OLTP，即时查询，以图遍历为主



图查询

图计算

HugeGraph

HugeGraph应用分类

OLAP，离线分析，全图多轮迭代计算



图查询

图计算

HugeGraph



目录

1. 图数据库概念
2. HugeGraph典型图算法实践
 - 1) 知识问答
 - 2) 协同推荐
 - 3) 关系挖掘
 - 4) 循环担保
 - 5) 群控作弊
 - 6) 团伙欺诈
3. HugeGraph优势



架构革新 ◎ 高效可控
第十一届中国数据库技术大会
DATABASE TECHNOLOGY CONFERENCE CHINA 2020

目录

1. 图数据库概念

2. HugeGraph典型图算法实践

- 1) 知识问答
- 2) 协同推荐
- 3) 关系挖掘
- 4) 循环担保
- 5) 群控作弊
- 6) 团伙欺诈

OLTP

3. HugeGraph优势



目录

1. 图数据库概念

2. HugeGraph典型图算法实践

1) 知识问答 2) 协同推荐 3) 关系挖掘

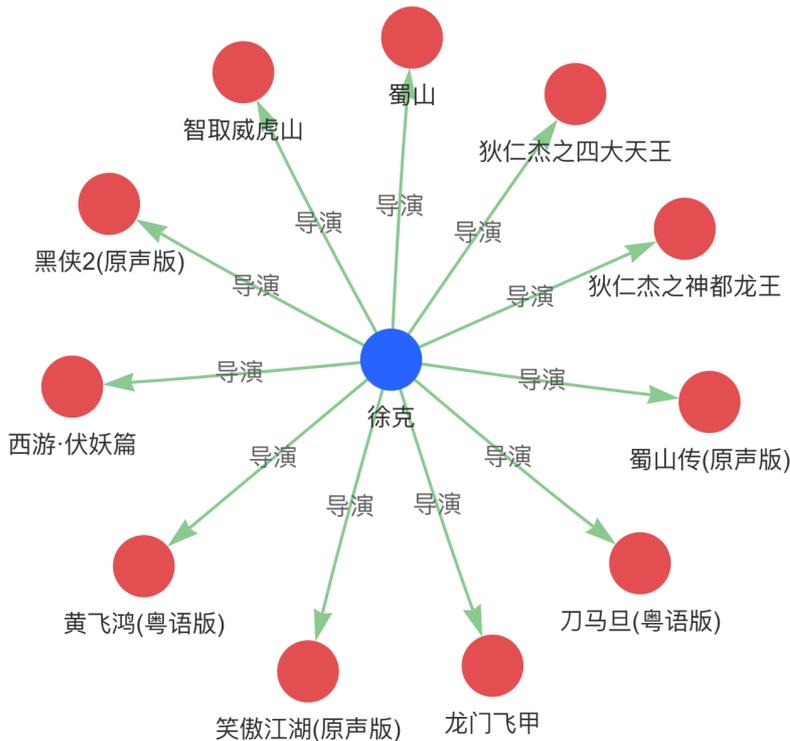
4) 循环担保 5) 群控作弊 6) 团伙欺诈 OLAP

3. HugeGraph优势

1、知识问答

- 徐克导演的电影
- 梁家辉演出的电影
- 徐克和梁家辉合作的电影
- 与徐克合作的武侠片导演

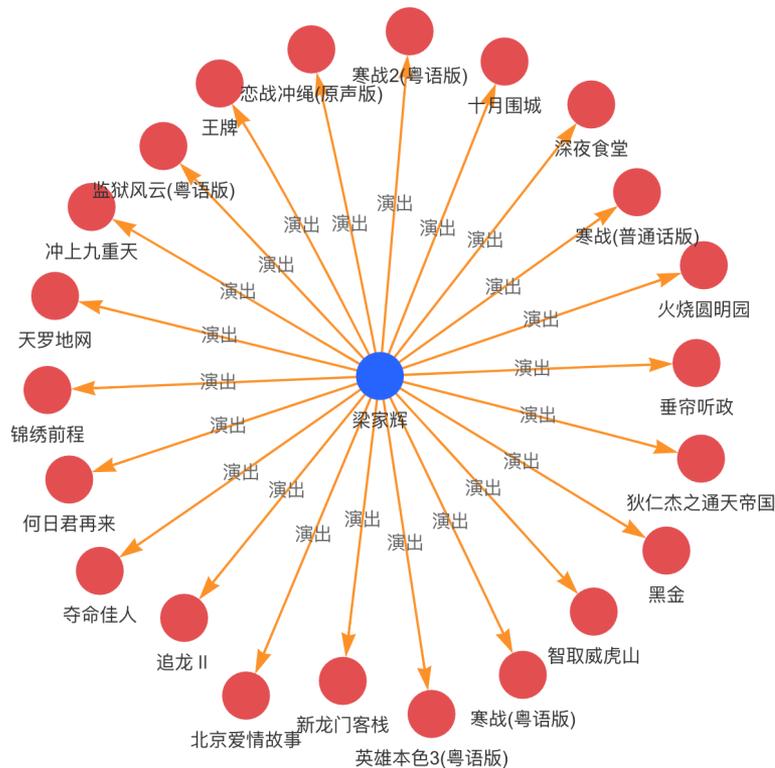
`g.V('徐克').out('导演').path()`



1、知识问答

- 徐克导演的电影
- 梁家辉演出的电影
- 徐克和梁家辉合作的电影
- 与徐克合作的武侠片导演

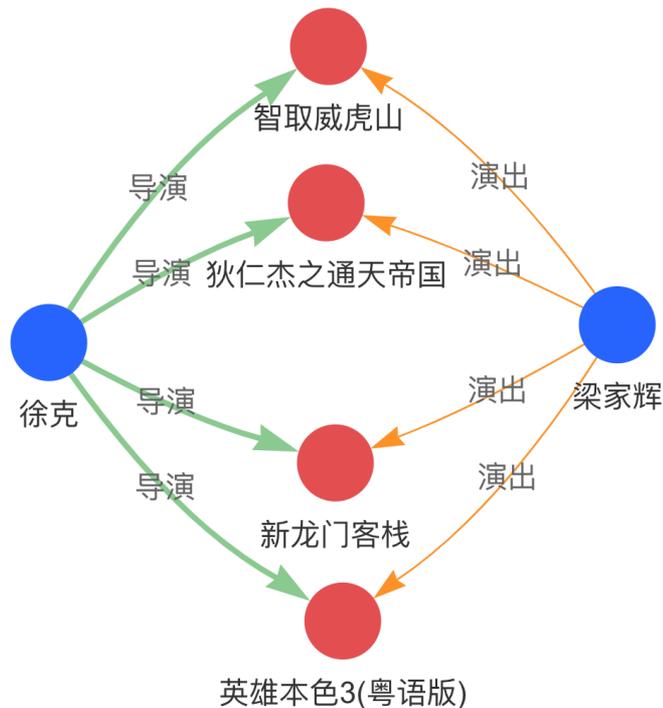
g.V('梁家辉').out('演出').path()



1、知识问答

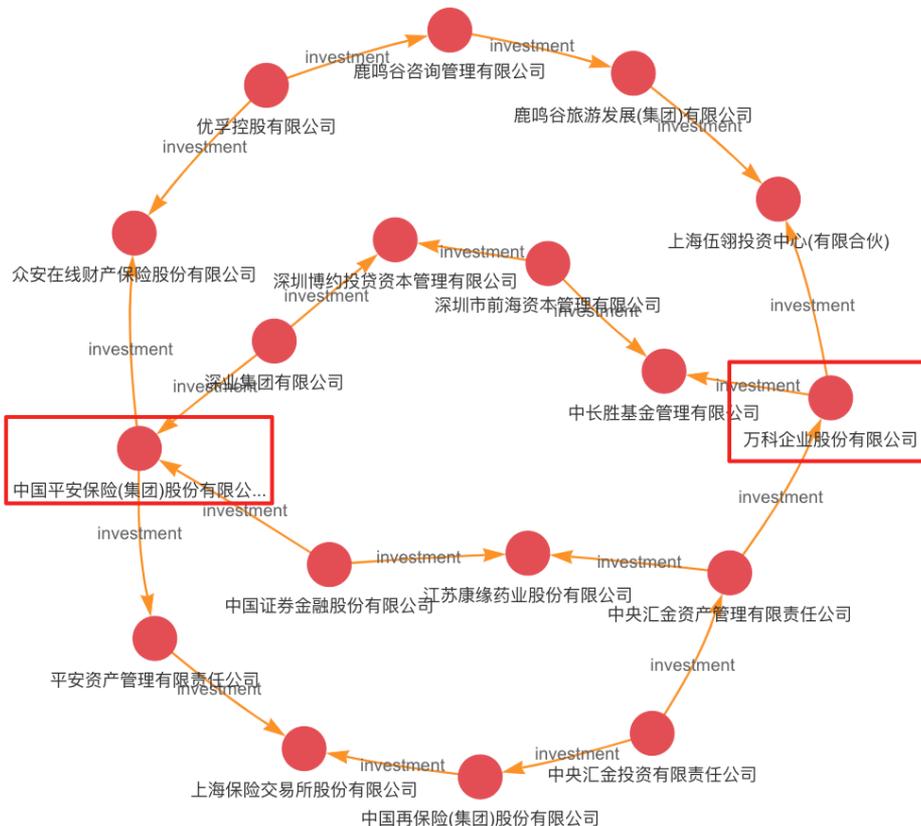
- 徐克导演的电影
- 梁家辉演出的电影
- 徐克和梁家辉合作的电影
- 与徐克合作的武侠片导演

```
g.V('徐克').match(
  as('d').out().as('m'),
  as('m').in().as('a'))
.where(as('a').hasId('梁家辉'))
.select('m')
```



3、关系挖掘

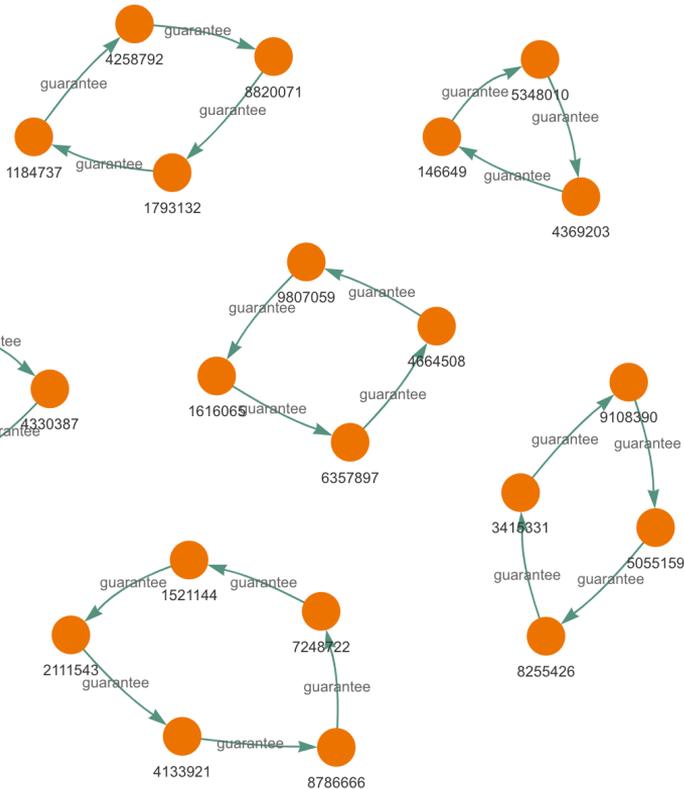
- 股权穿透
- 深度分析



最短路径、全最短路径

4、循环担保

- 借贷风控（贷款前循环担保检测）
- 企业责任担保（授信、连带责任担保）

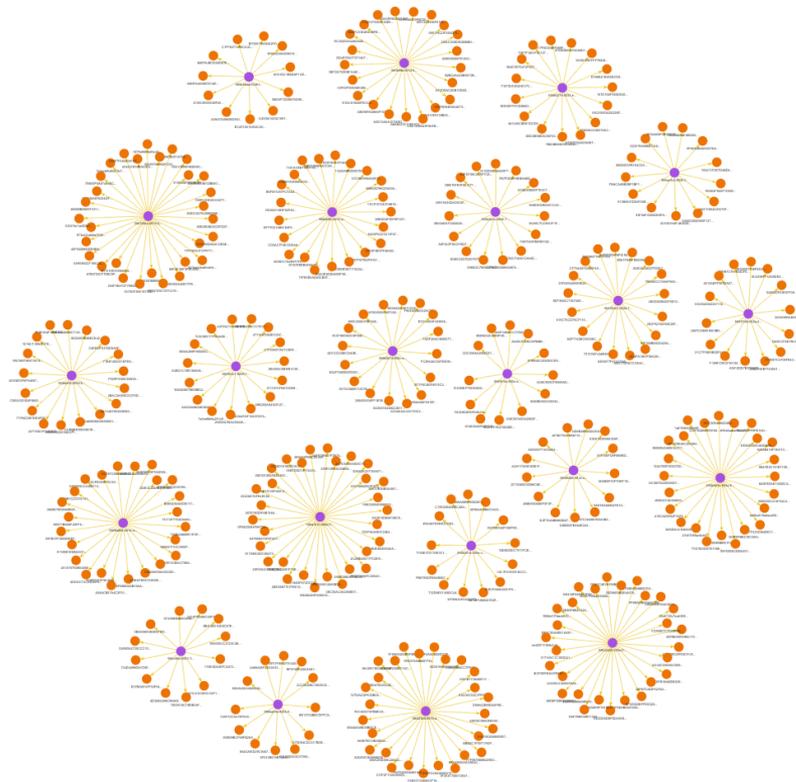


使用环路检测算法进行全图计算

5、群控作弊

- 控制大量账号 - 共享同一手机
- 群控大量账号 - M对一共享IP
- 群控大量账号 - M对N共享IP
- 群控大量账号 - M对M共享IP

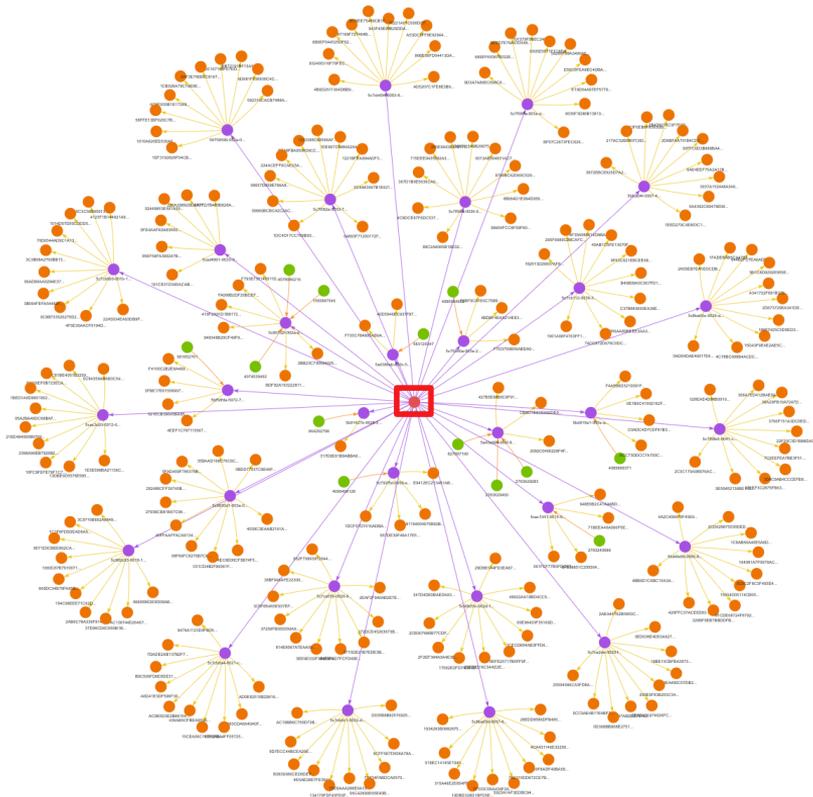
找出登录账号的度中心性大于10的手机



5、群控作弊

- 控制大量账号 - 共享同一手机
- 群控大量账号 - M对一共享IP
- 群控大量账号 - M对N共享IP
- 群控大量账号 - M对M共享IP

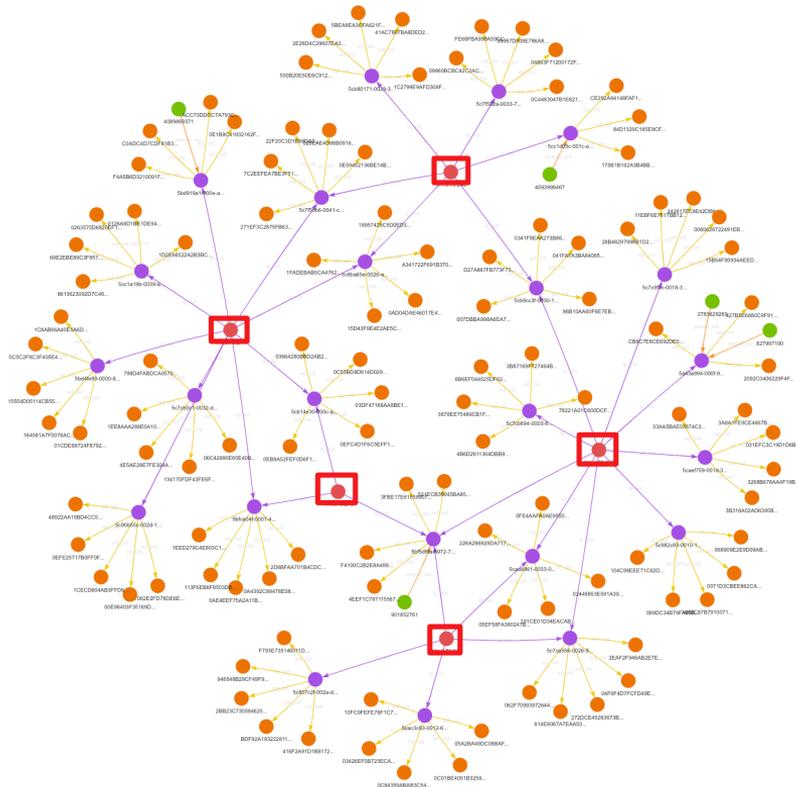
找出登录账号的2阶度中心性大于500的IP



5、群控作弊

- 控制大量账号 - 共享同一手机
- 群控大量账号 - M对一共享IP
- 群控大量账号 - M对N共享IP
- 群控大量账号 - M对M共享IP

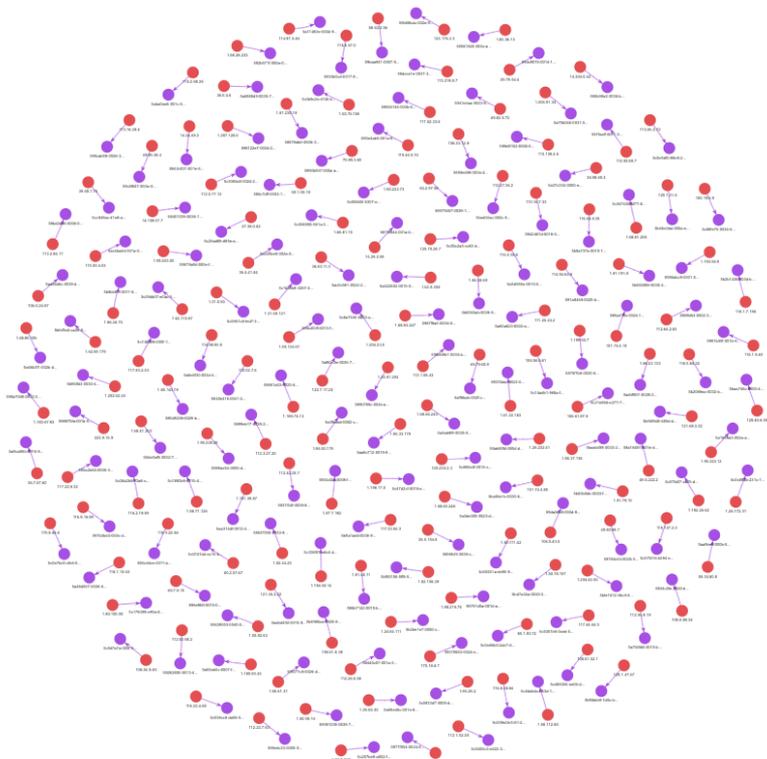
找出登录账号的2阶度中心性大于100的IP
且IP两两之间均存在二度关联



5、群控作弊

- 控制大量账号 - 共享同一手机
- 群控大量账号 - M对一共享IP
- 群控大量账号 - M对N共享IP
- 群控大量账号 - M对M共享IP

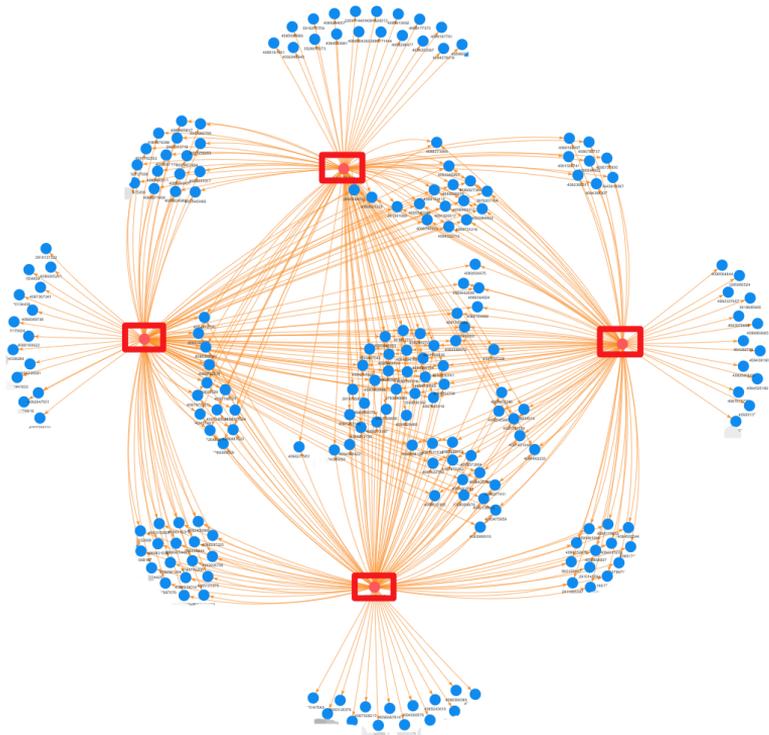
一个IP上仅登录一个账号，看起来正常



5、群控作弊

- 控制大量账号 - 共享同一手机
- 群控大量账号 - M对一共享IP
- 群控大量账号 - M对N共享IP
- 群控大量账号 - M对M共享IP

一个IP上仅登录一个账号，看起来正常，但是历史时间里，共享IP可能被大量用户使用过，可从其它维度找出IP之间的紧密关系，方法是使用梭型相似度算法

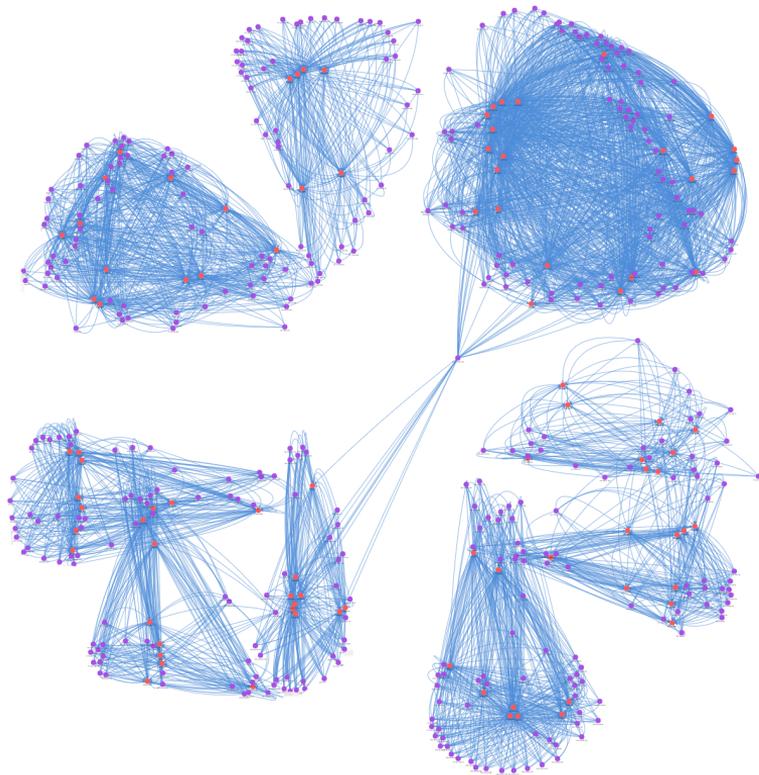


6、团伙欺诈

- 如何找出社区？
- 如何判断为风险社区？
- 如何找出社区核心成员？

方法：

- 连通子图
- LPA
- SCAN
- Louvain

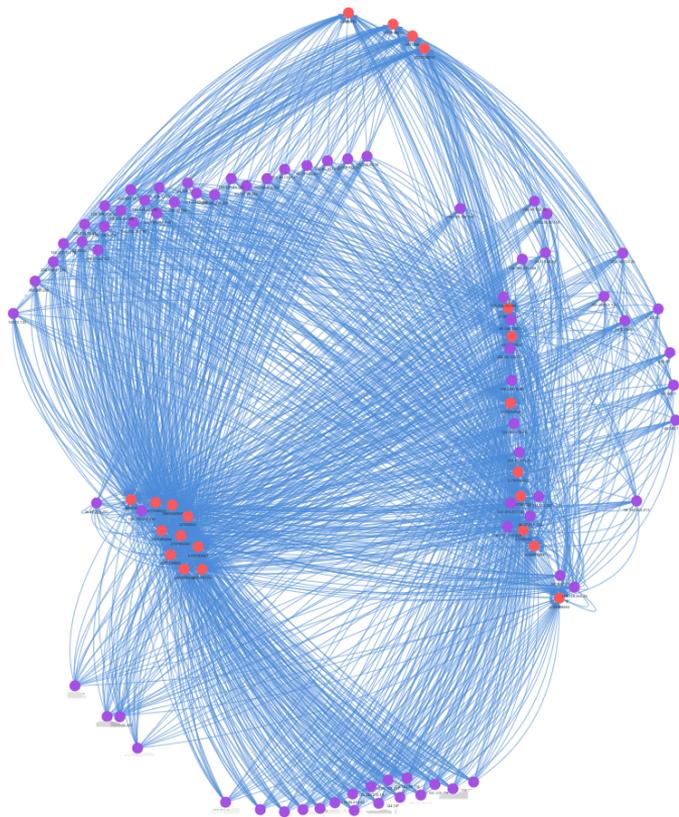


6、团伙欺诈

- 如何找出社区？
- 如何判断为风险社区？
- 如何找出社区核心成员？

方法：

- 孤立子图规模
- 社区成员平均特征值异常
- 关联关系密度极高

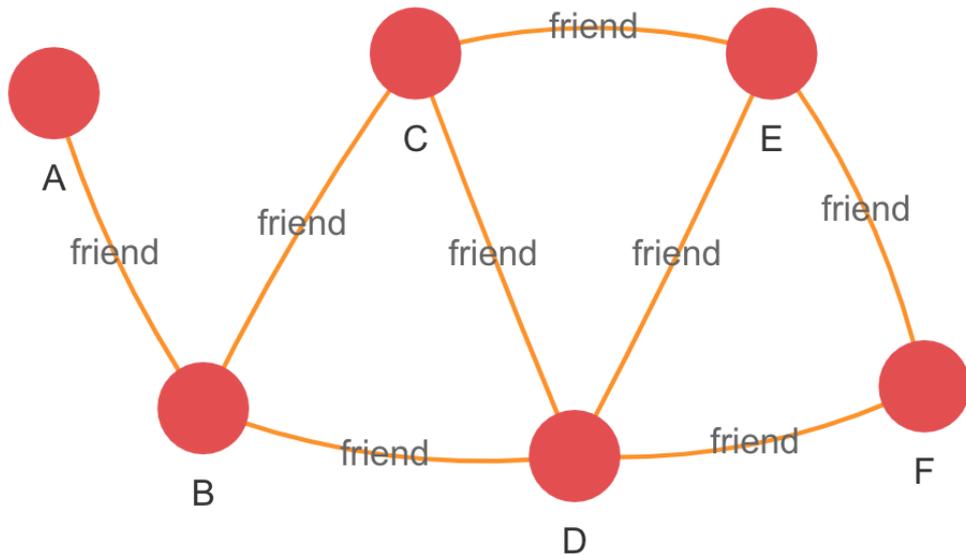


6、团伙欺诈

- 如何找出社区？
- 如何判断为风险社区？
- 如何找出社区核心成员？

方法：

- 紧密中心性高
- 中介中心性高
- PageRank值高

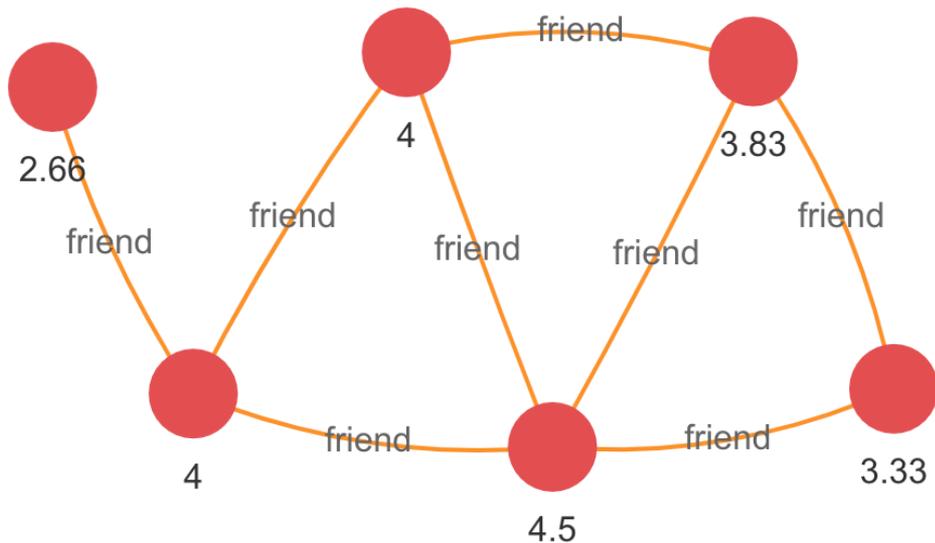


6、团伙欺诈

- 如何找出社区？
- 如何判断为风险社区？
- 如何找出社区核心成员？

方法：

- 紧密中心性高
- 中介中心性高
- PageRank值高

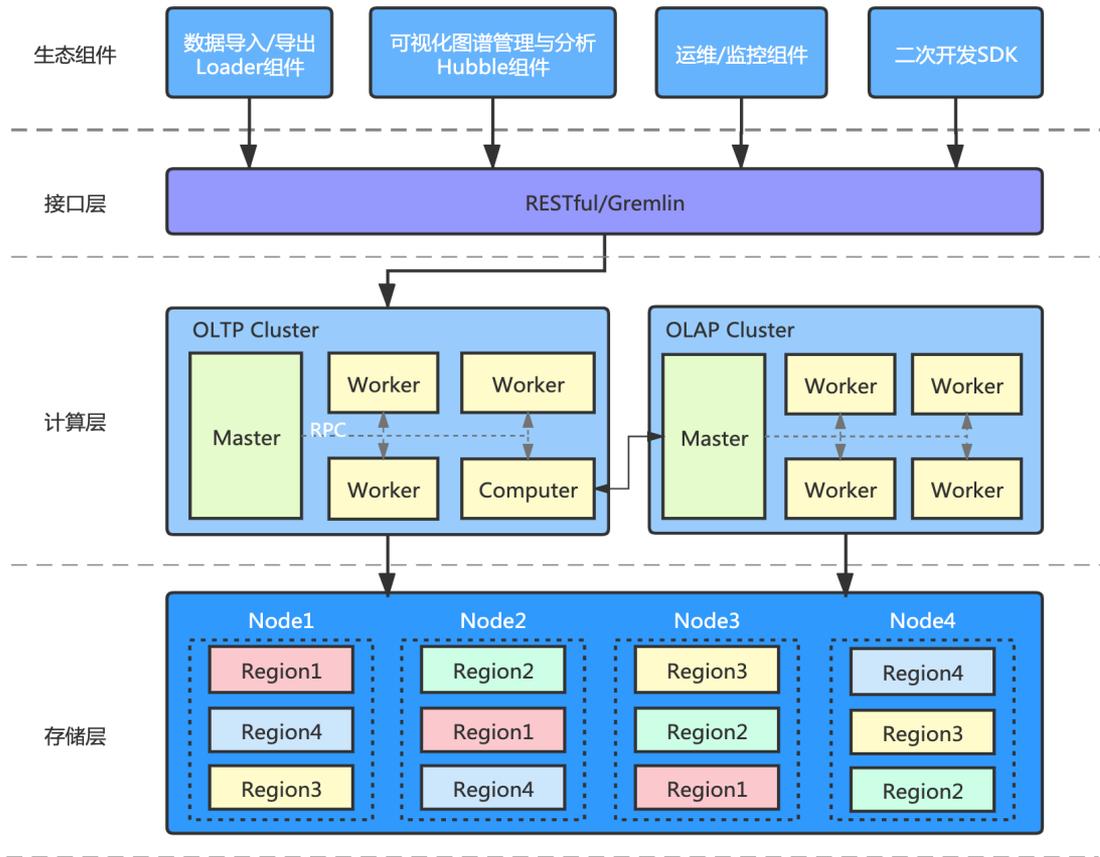




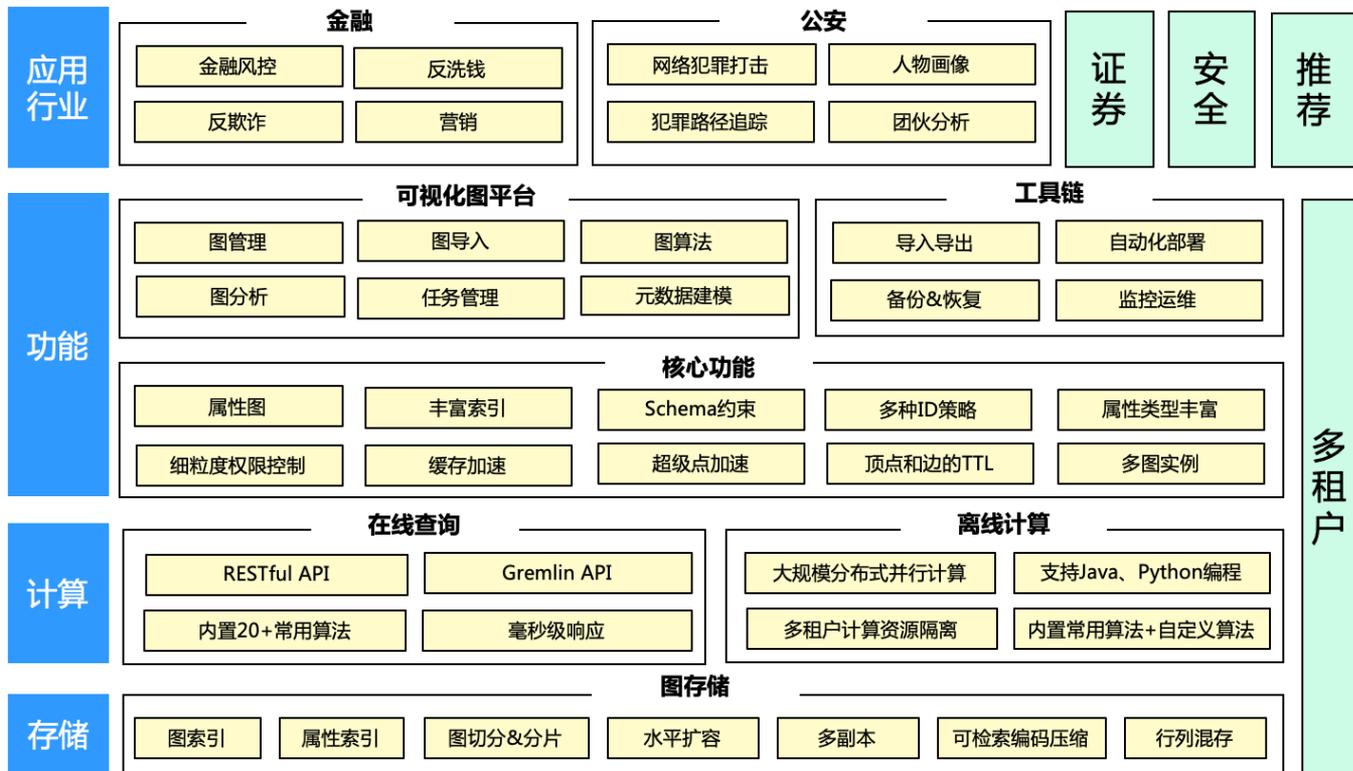
目录

1. 图数据库概念
2. HugeGraph典型图算法实践
 - 1、知识问答
 - 2、协同推荐
 - 3、关系挖掘
 - 4、循环担保
 - 5、群控作弊
 - 6、团伙欺诈
3. HugeGraph优势

- 支持千亿大规模图
- 支持实时在线写入，吞吐单机达30~50万每秒
- OLTP 实时在线图查询毫秒级延迟
- OLAP 全图离线分析在小时级耗时（10台普通机器集群）
- 水平扩展，支持分布式存储、横向扩容
- 高可用，支持数据多副本、备份恢复



- 代码开源、中文交流社区
- 工具链完善
- 通用图查询语言 Gremlin
- 图算法丰富，内置 OLTP 20+ & OLAP 10+ 算法
- 功能丰富



THANKS

