# 审计图谱设计及开发实施方案

# 项目概述

## 项目背景

华夏银行EAST5.0数据体系内，存在大规模客户信息、交易记录、担保信息等结构化数据，这些数据之间存在复杂的关联关系。为更好地支持数据价值挖掘和业务决策，我们计划构建以社交网络为主旨的知识图谱应用，通过可视化分析平台实现图数据的交互式分析。

知识图谱可以打通数据孤岛,支持跨域知识推理。可视化分析则可以直观呈现复杂网络,辅助决策。本项目旨在把两者优势结合,以知识图谱为基础,通过可视化分析提升运营效能。

## 项目目标

本项目的总体目标是基于EAST5.0客户相关数据，构建知识图谱可视化平台，通过直观的交互式分析提升数据应用效率，为业务决策提供智能支持。

具体包含以下细分目标：

* 构建社交网络应用知识图谱模型
* 支持EAST5.0源数据全量和增量同步至图引擎
* 实现基础的网络分析算法，如中心性分析、社区发现等
* 开发交互式可视模块，如网络概览、节点查询、逐级扩展等

## 项目范围

本项目的范围包括：

* 构建核心知识图谱
  + 收集并整合EAST5.0内部客户数据
  + 数据清洗、映射与融合
  + 构建实体-关系图谱模型
* 社交网络分析知识图谱可视化平台开发
  + 图数据查询接口
  + 交互式可视化组件
  + 图和表格数据下载与重载

## 里程碑阶段划分

本项目采用分阶段递进的方式推进，主要里程碑包括：

* 第一阶段(4周：2023-06-05~~2023-07-04)
  + 组建项目团队
  + 评估技术方案，确定平台结构
  + 确定业务需求
  + 搭建项目基础开发框架及知识图谱模型
* 第二阶段(7周：2023-07-05~~2023-08-30)
  + 实现数据清洗、映射及融合，结合调度组件实现增量同步
  + 六大模块核心接口开发
  + 可视化模块和基础功能开发
  + 测试并迭代优化性能
* 第三阶段(2周：2023-08-31~~2023-09-13)
  + 编写测试案例，交付测试中心提测
  + 修复Bug及上线部署

## 团队和资源

本项目核心团队组成：

* 项目经理：1名
* 需求分析师：1名
* 前端开发工程师：2名
* 后端开发工程师：1名

资源：

* 应用服务器：8核16G 4台
* MySQL服务器(主从架构)：4核16G 3台

# 系统整体架构

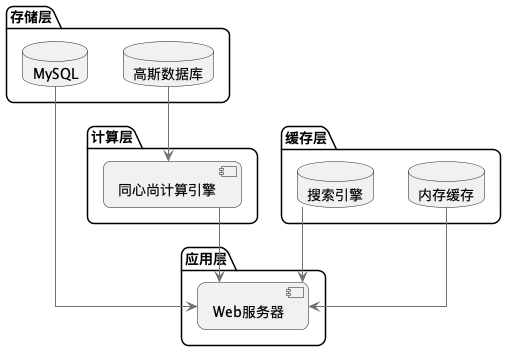


图 1 逻辑架构图

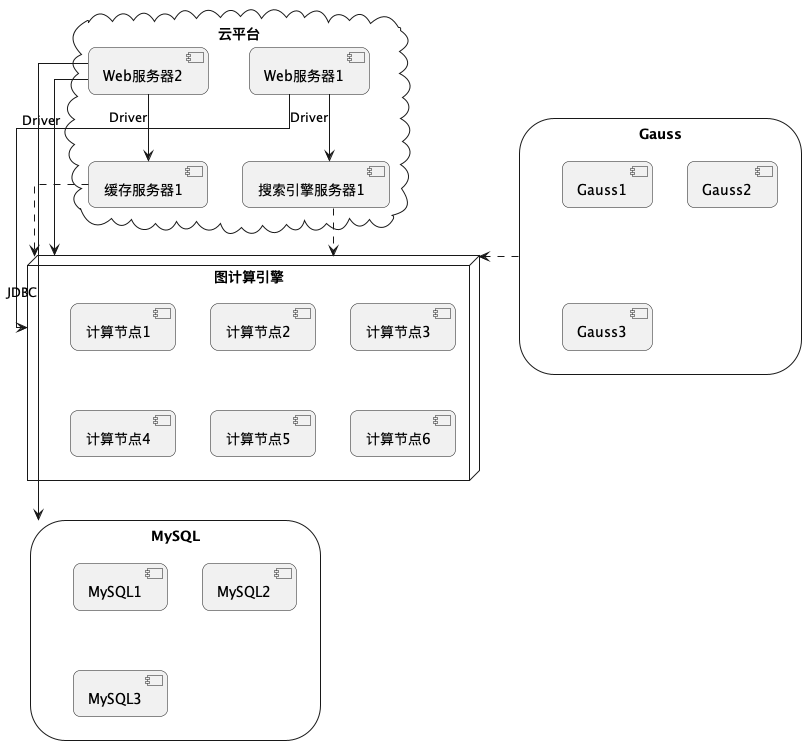


图 2 物理架构图

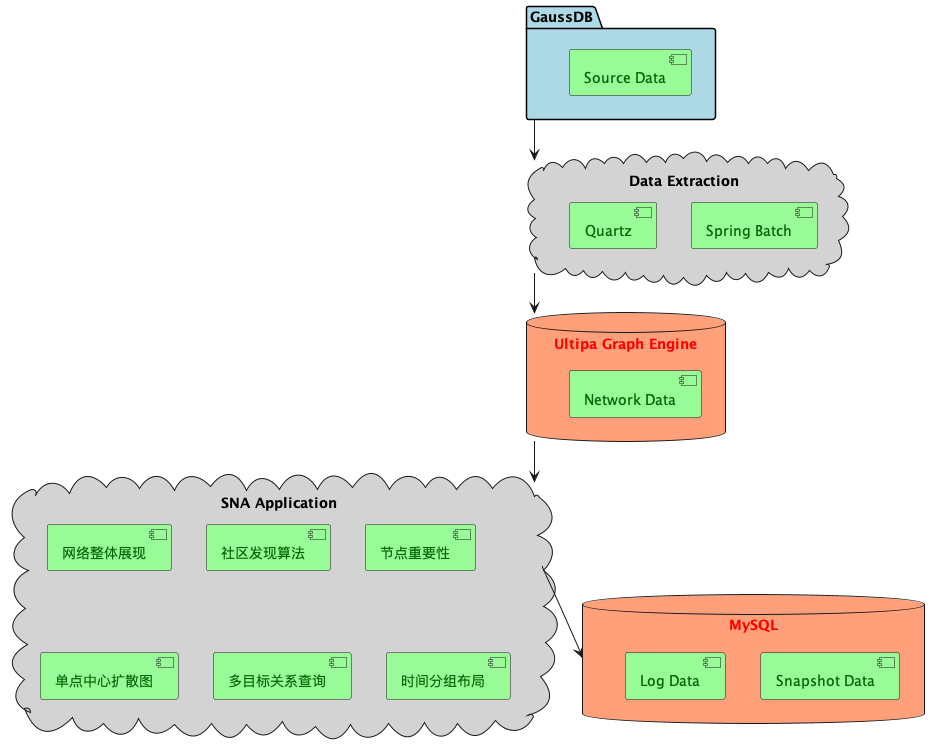


图 3 数据流程架构图

# 业务功能模块

## 数据同步

### 构建社交网络分析图模型

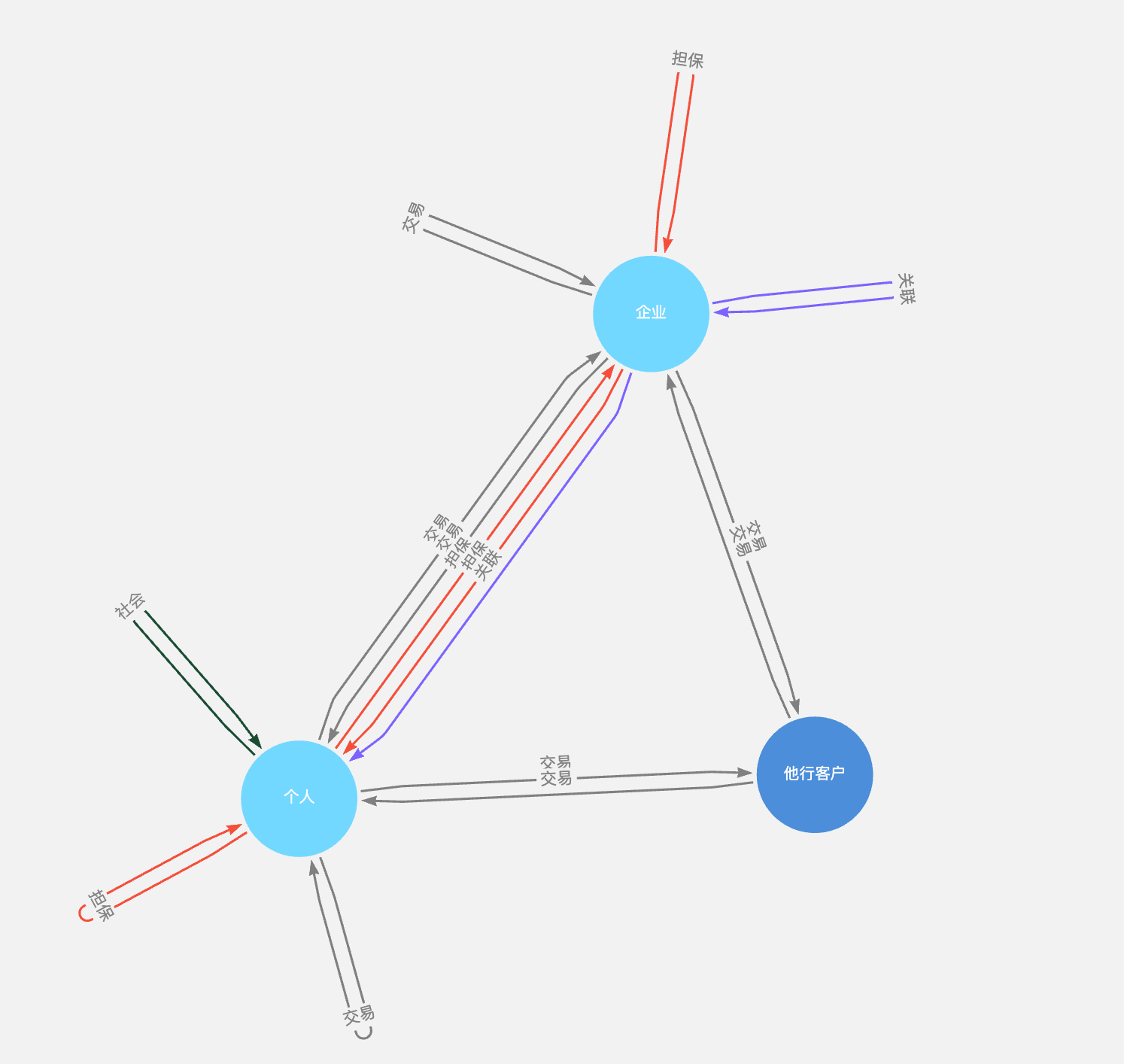
定义点

* 个人客户
* 企业客户
* 他行客户

定义边

* 社交关系
* 关联关系
* 交易关系
* 担保关系

模型



### 高斯数据全量同步

同步策略：

1. 先同步三个点，对应3张表；
2. 再同步四条边，顺序不分先后，每条边涉及多张表，SQL查询使用关联查询，一条语句分页查询；
3. 每500条查询一次，批量写入图引擎；
4. 查询源数据时只有业务限制条件

涉及技术：

1. spring batch：批量读写；

### 高斯数据增量同步

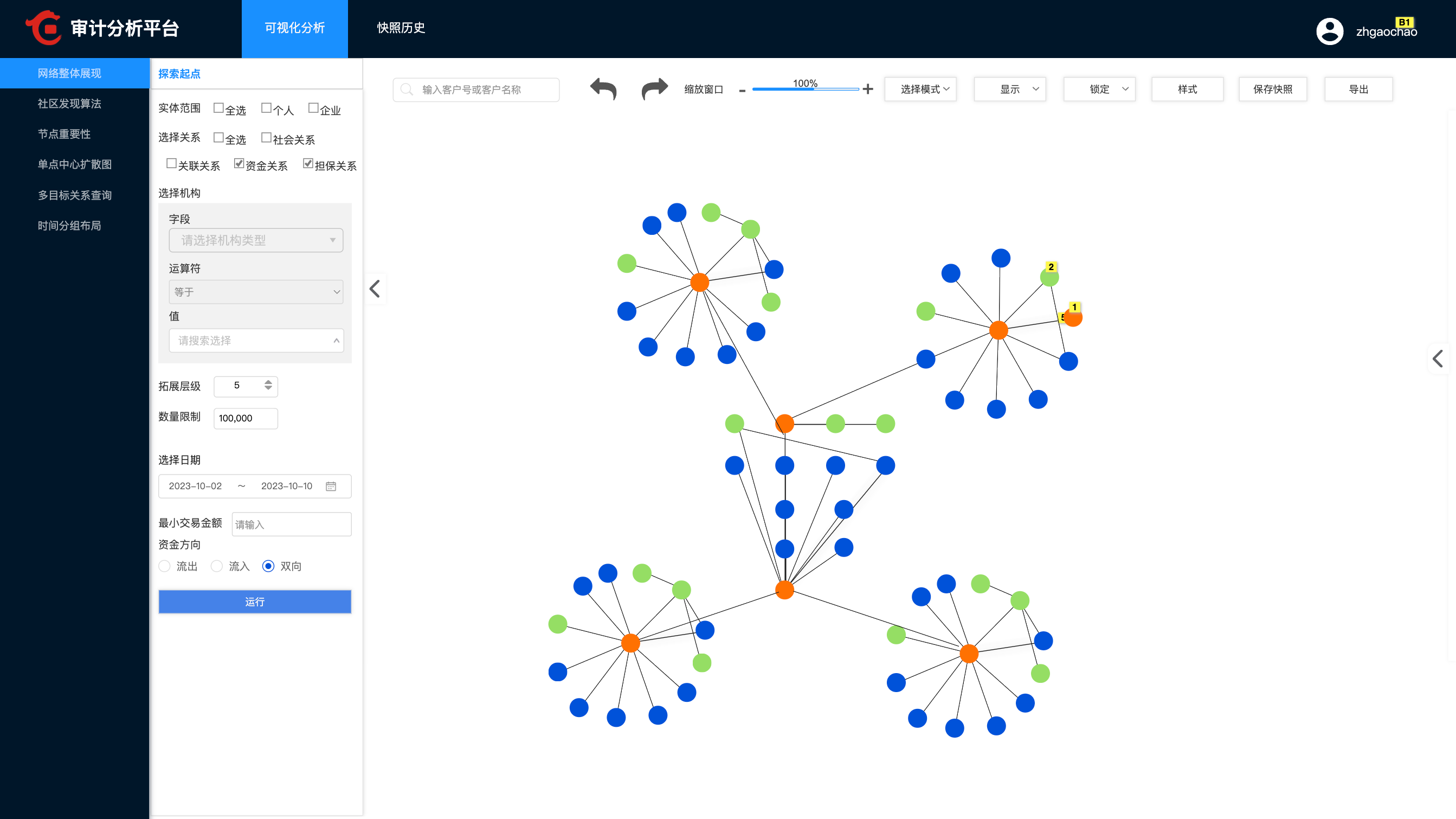
同步策略：

1. 依然是先同步三个点表，再同步四条边的数据；
2. 查询源数据时除了业务限制条件，还要根据已同步数据的插入日期，来选取之后日期的源数据；

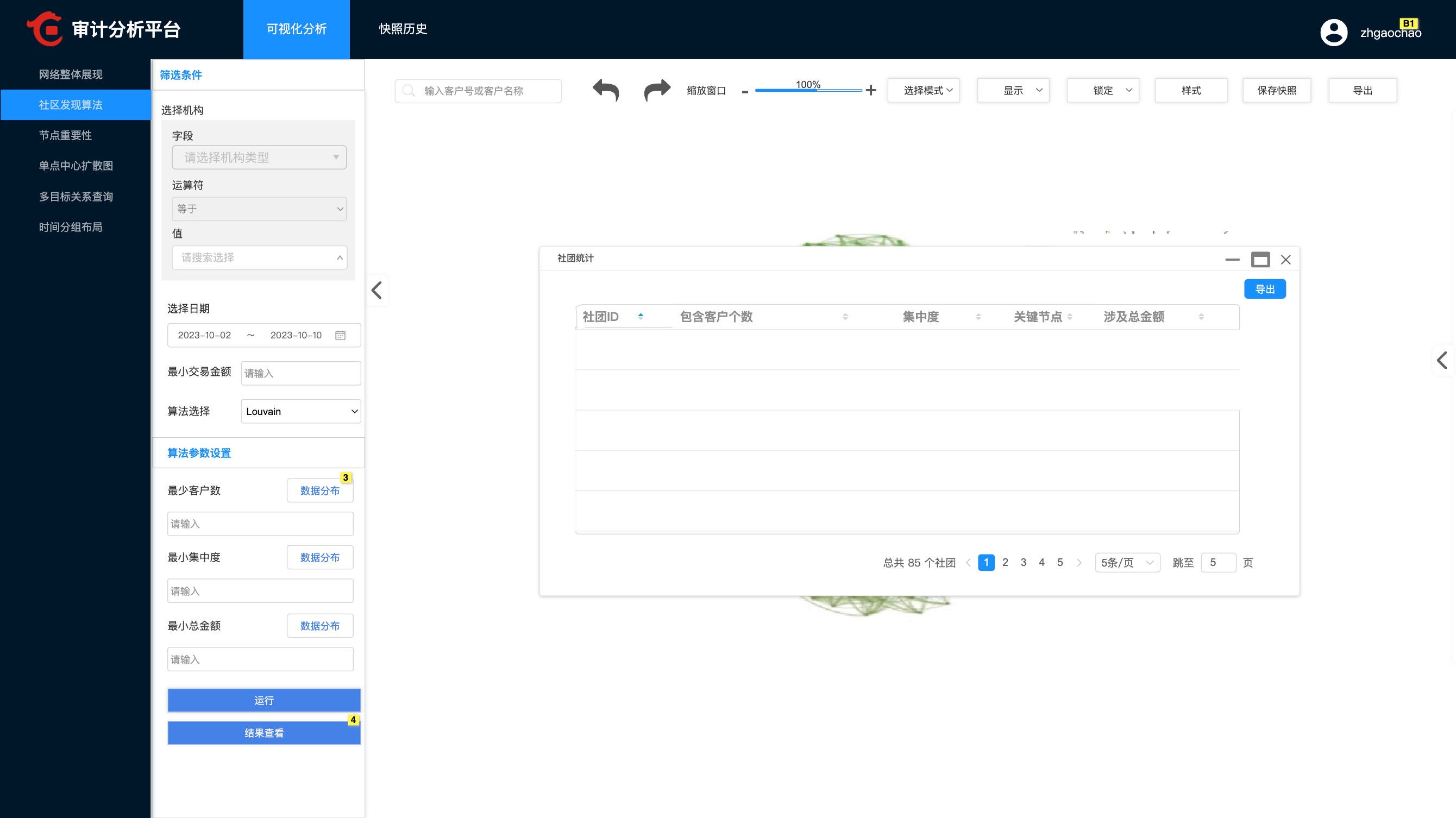
涉及技术：

1. spring batch：批量读写
2. Quartz：调度器

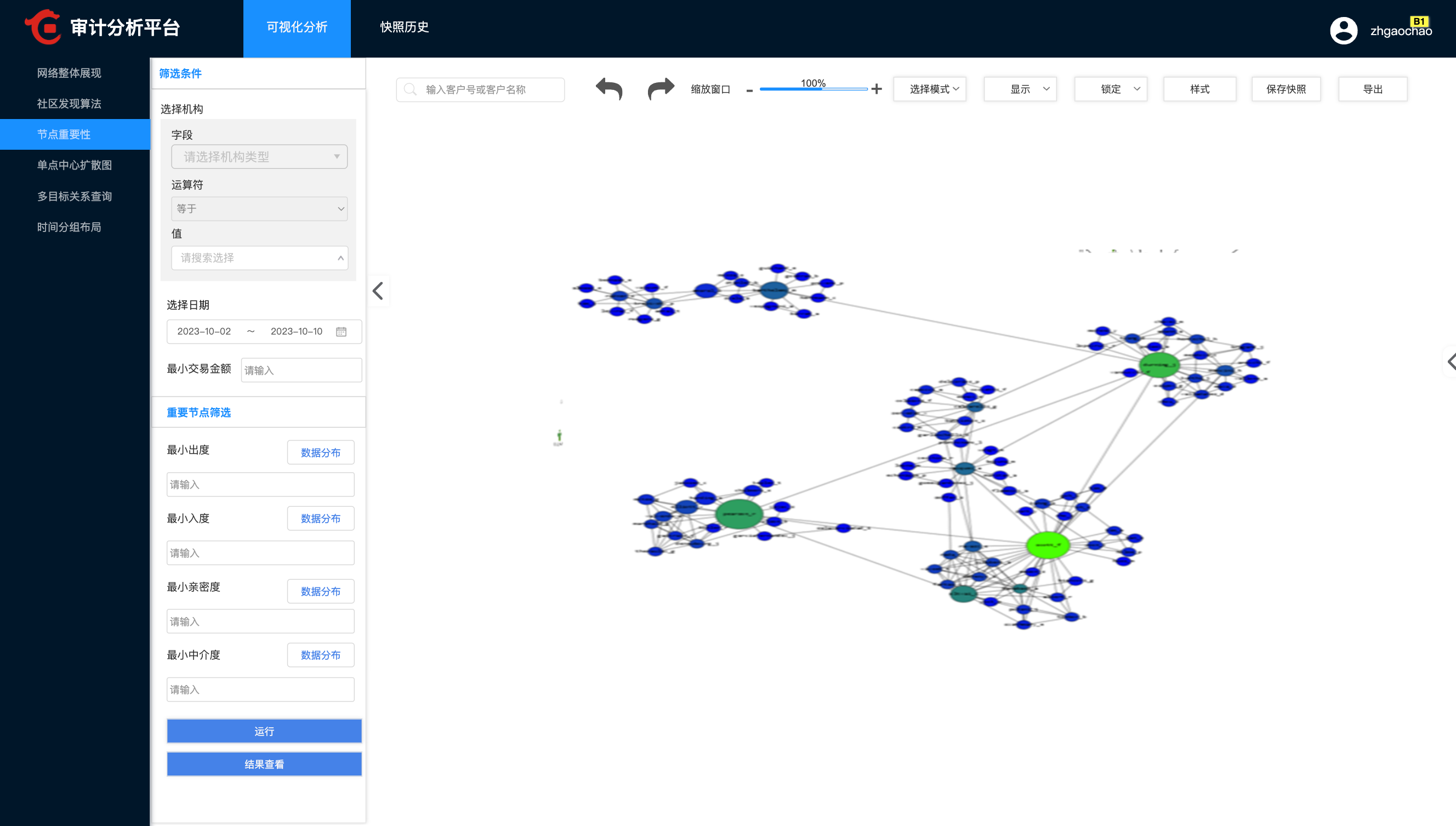
## 网络整体展现



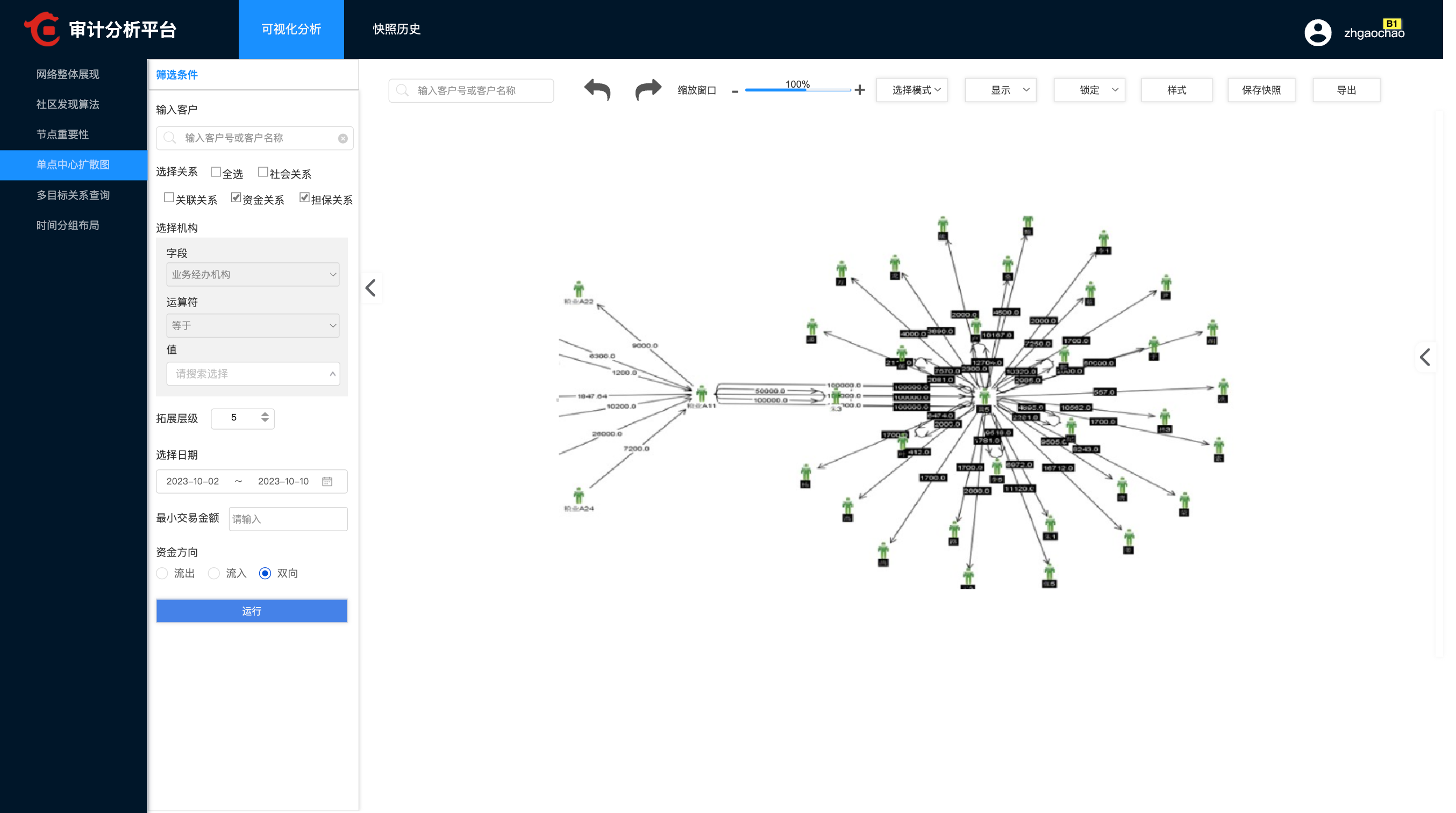
## 社区发现算法



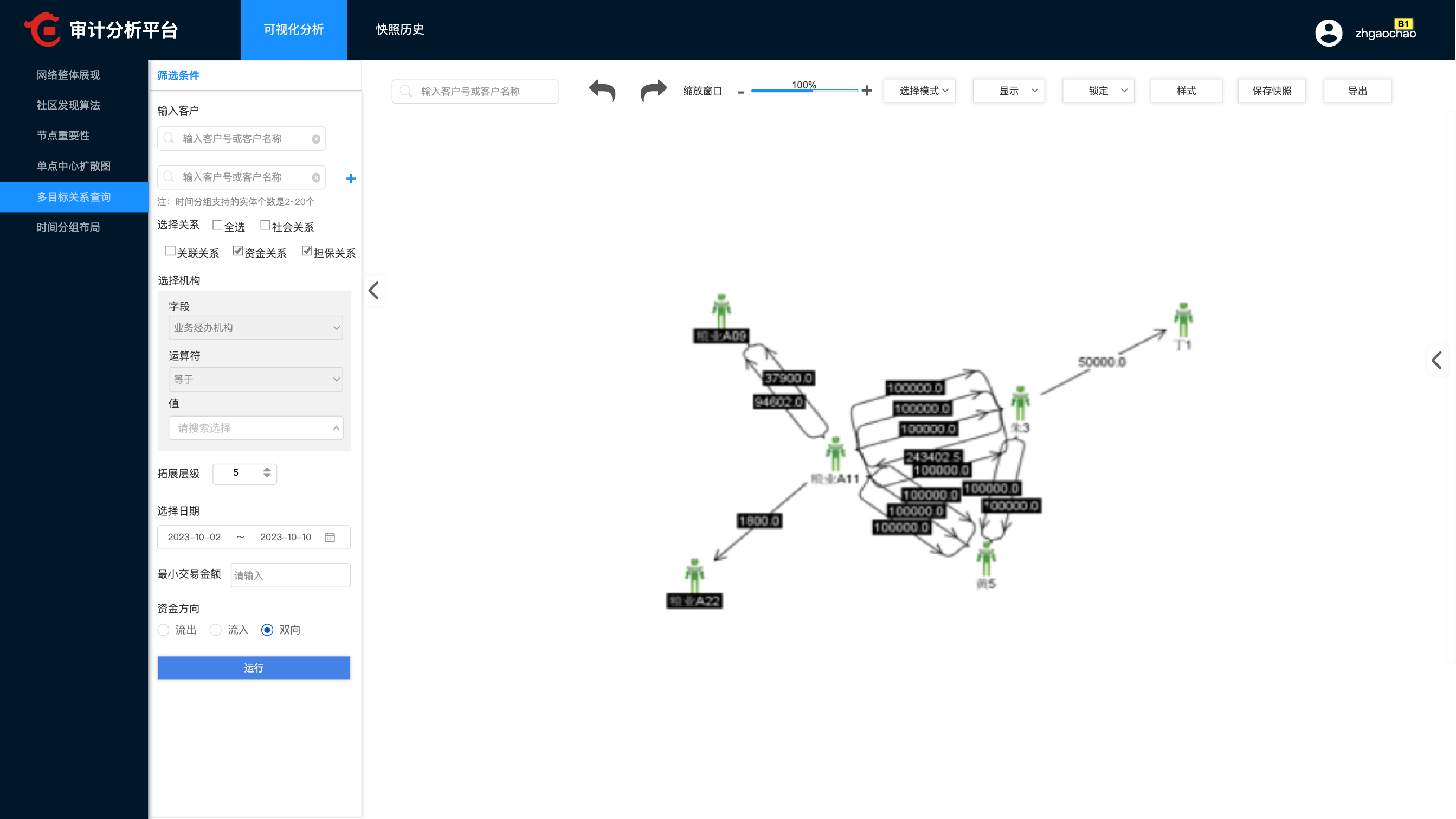
## 节点重要性



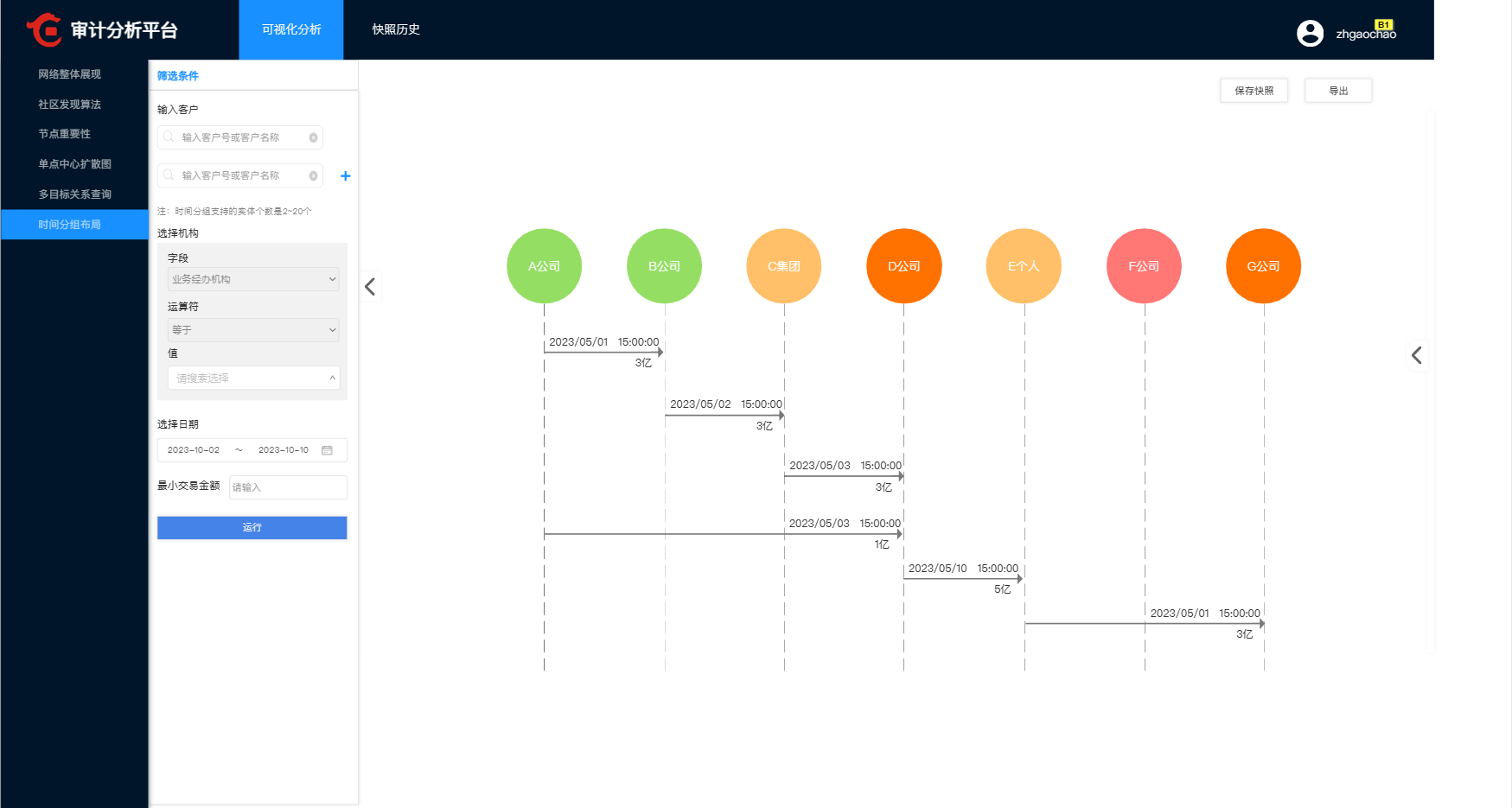
## 单点中心扩散图



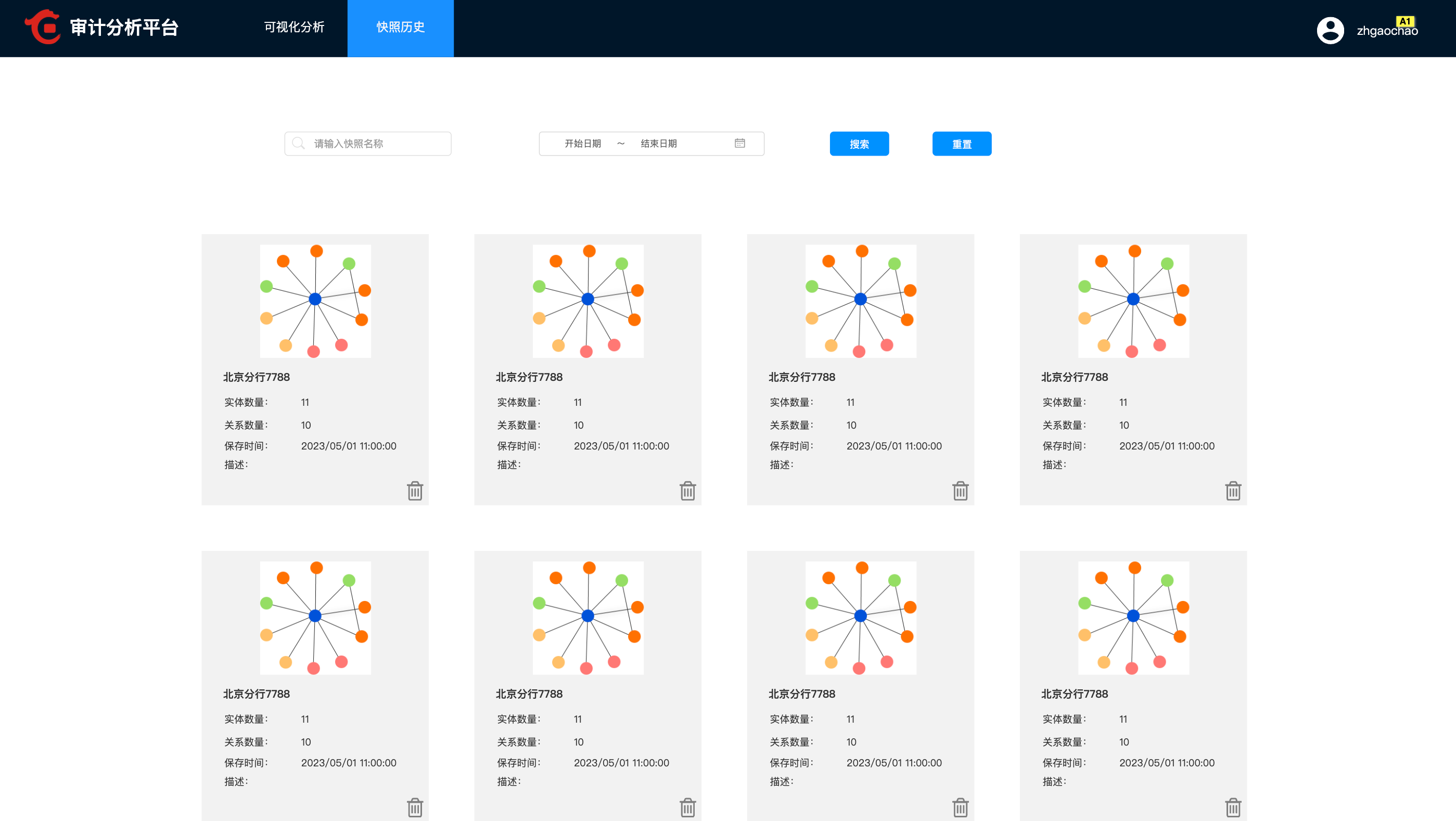
## 多目标关系查询



## 时间分组布局



## 快照历史



# 系统接口设计

## 网络整体展现

* 模型管理接口
* 网络概览数据查询接口
* 拓展接口

## 社区发现算法

* 社区列表查询接口
* 社区客户数统计接口
* 社区集中度统计接口
* 社区总金额统计接口
* 社区网络查询接口

## 节点重要性

* 节点出度统计接口
* 节点入度统计接口
* 节点亲密度统计接口
* 节点中介度统计接口
* 节点网络查询接口

## 单点中心扩散图

* 客户模糊查询接口
* 单点扩散查询接口

## 多目标关系查询

* 多目标路径查询接口

## 时间分组布局

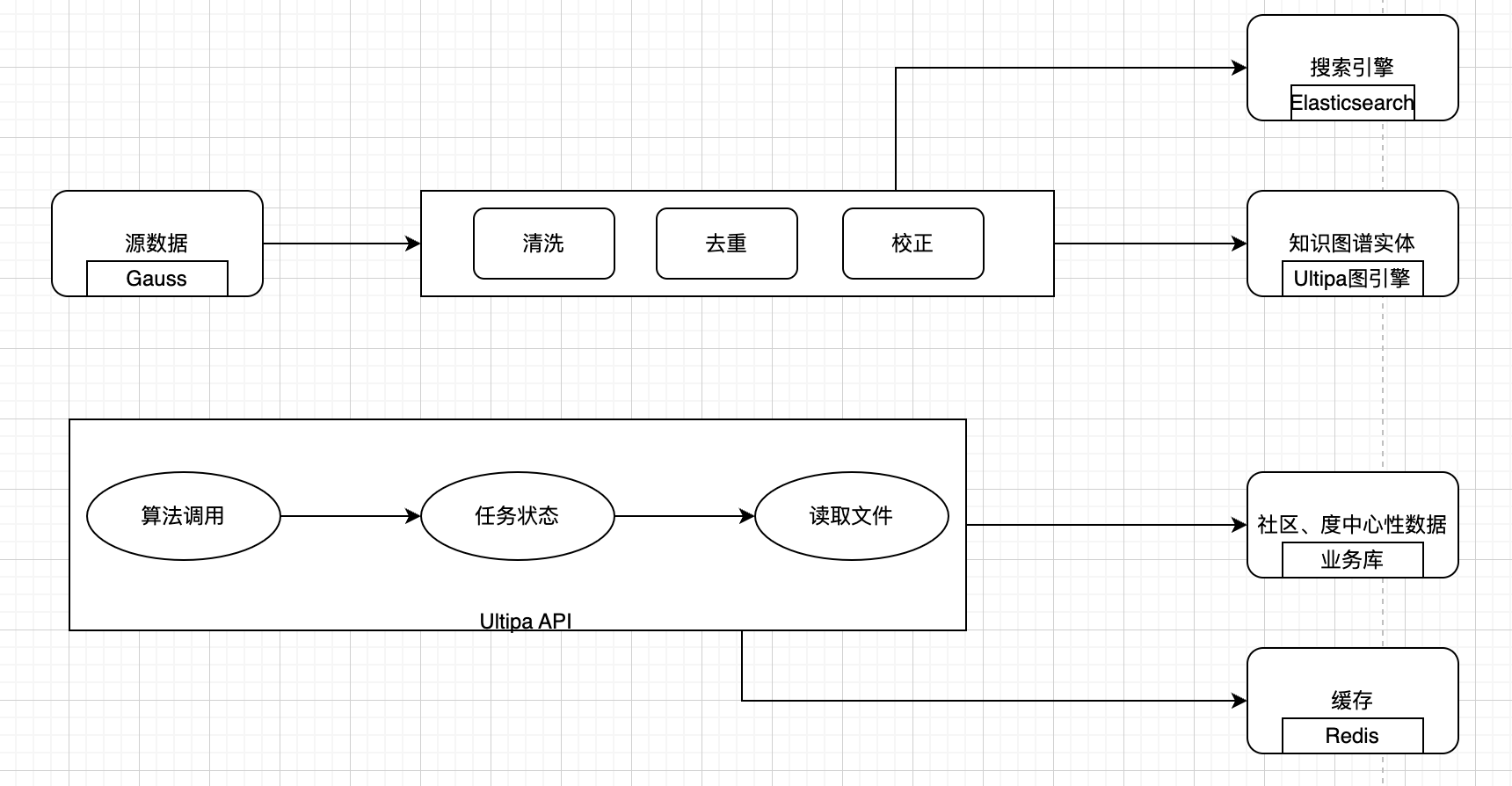
* 同多目标路径查询接口

## 快照历史

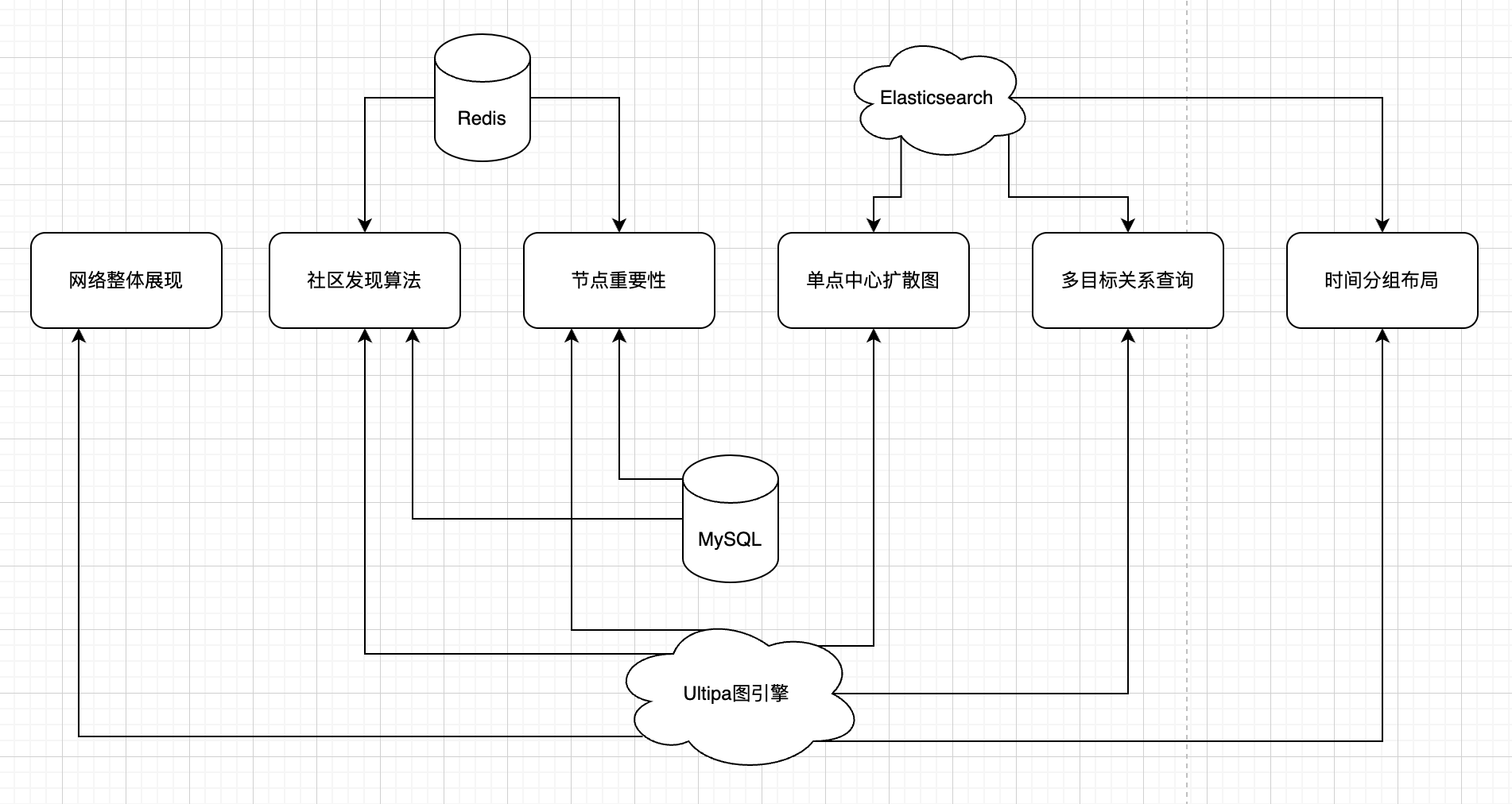
* 快照保存接口
* 快照删除接口
* 快照重放接口

# 数据流程描述

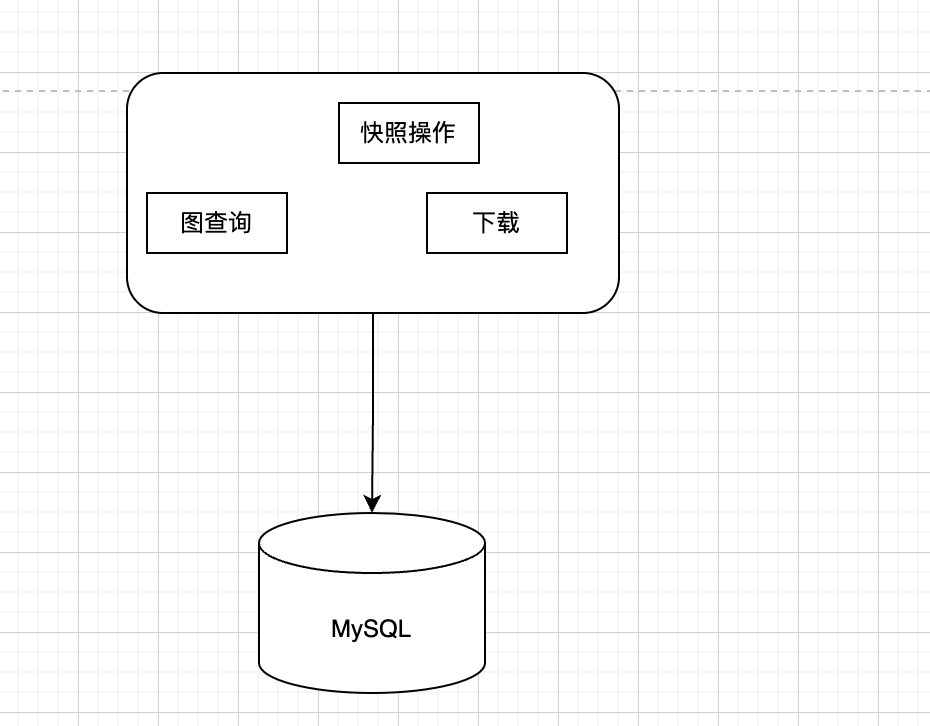
## 数据同步流



## 知识应用流



## 用户操作流



# 非功能性需求

* 系统性能：接口响应延迟<200ms，可视化交互<100ms
* 系统可用性：服务可用性>=99.95%
* 系统容量：知识图谱实体数量>3000万，边数量>6000万
* 系统可扩展性：存储和服务支持水平扩展
* 用户体验：可视化友好、操作便捷
* 开发效率：组件化和模块化设计，提高开发效率