

# 路线三 V2.0

《数据集成》



# 目录

- 路线三 综述
- 路线三 作业2：任务；产物
- 路线三 作业3：任务；产物
- 数据源 参考

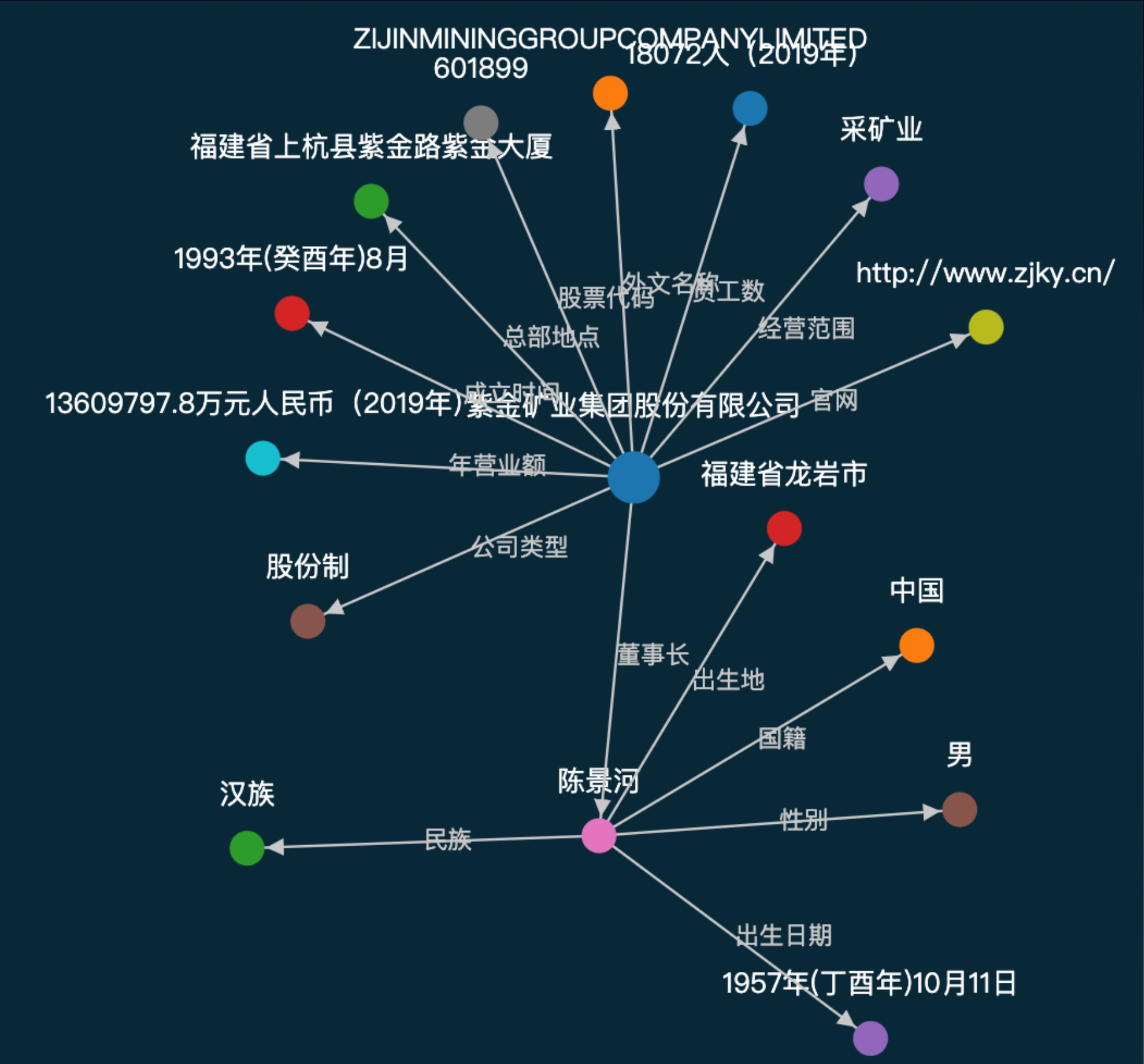
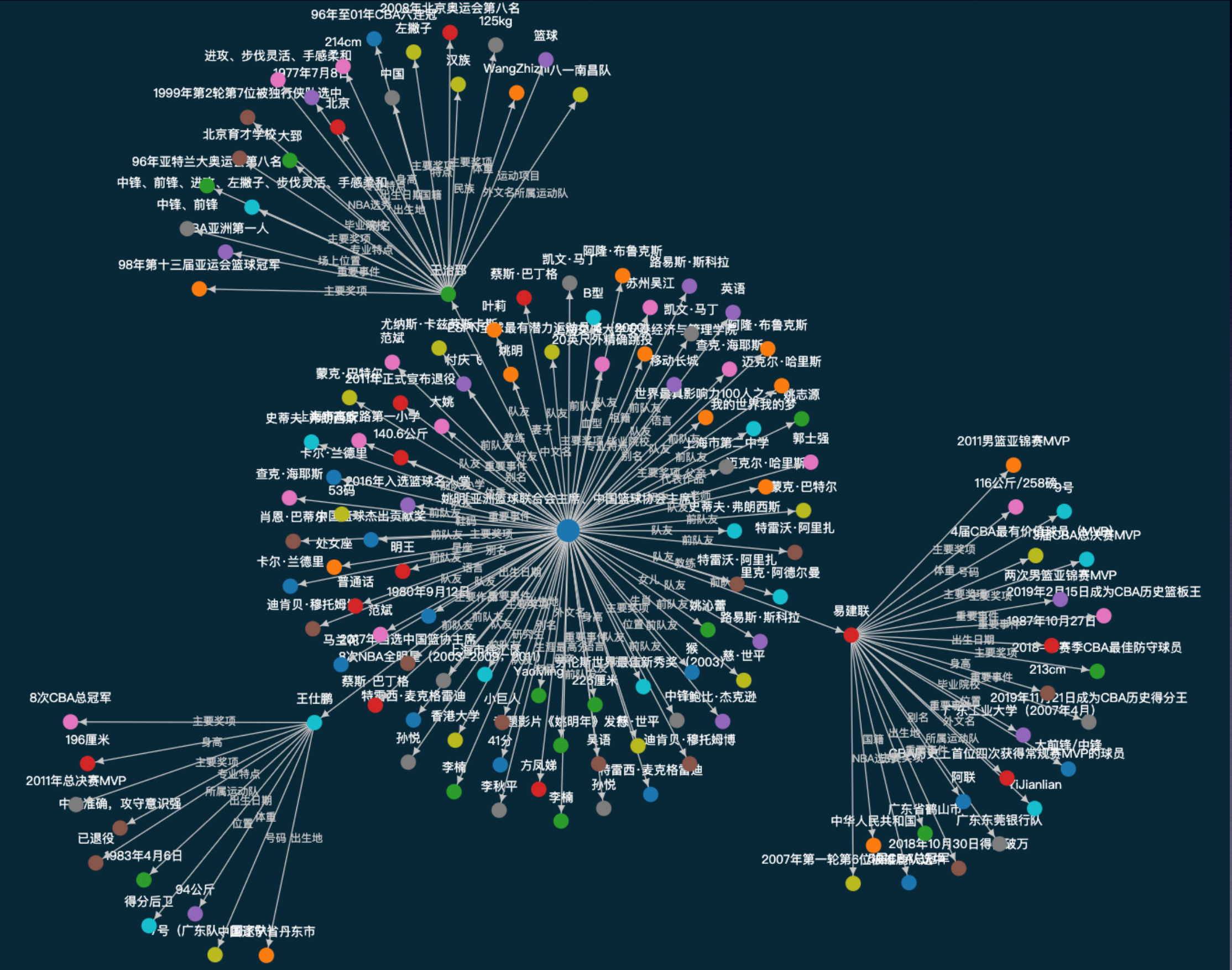


# 路线三 综述

- 运用数据集成相关技术，获取、整理 证券行业相关数据
- 分析数据，得出 实用结论
  - 如 cc 公司 或 ii 行业 股票价格 波动特征、走势 .....
- 集成更多数据，设计、构建 知识图谱，探索 投资结论
  - 如 c1 公司 与 c2 公司关联过多，出于分散风险考虑，不建议全部投资 .....



# 知识图谱大概模样





# 路线三 作业2

- 任务
  - 数据获取、数据整理（各组自行开展，数据结构、数据库类型不作要求）
- 产物
  - 小组信息.doc
    - 组号，组长姓名学号手机，组员姓名学号
  - 项目文档.pdf
    - 组员分工，项目 git 地址
    - 数据源，获取方式，关键代码（如增量爬取等）
  - 数据库设计文档.pdf
    - 数据库名称，数据表字段、含义、约束信息，数据库访问方式



# 路线三 作业3

- 任务（三选一）（得分会综合考虑任务难度与完成度）
  - 1) 分析数据、得出**结论**（小组内数据）
  - 2) 小范围集成，分析数据、得出**结论**（与个别组合作，交换数据）
  - 3) 大范围集成，构建知识图谱，探索**结论**（与多个组合作，交换数据）
- 产物
  - 小组信息.doc
    - 组号，组长姓名学号手机，组员姓名学号
  - 项目文档.pdf
    - 任务序号，组员分工，项目 git 地址
    - 针对任务2、3：数据源，集成过程与方法，结果数据库地址与描述（参考作业2数据库设计文档的要求）
  - 数据分析报告.pdf
    - 分析思路，相关算法，所得**结论**



# 数据源 参考

- <http://pycs.greedyai.com/>
- <https://tushare.pro/>
- .....