

实验 6：智能营销实验指南

一、实验目的

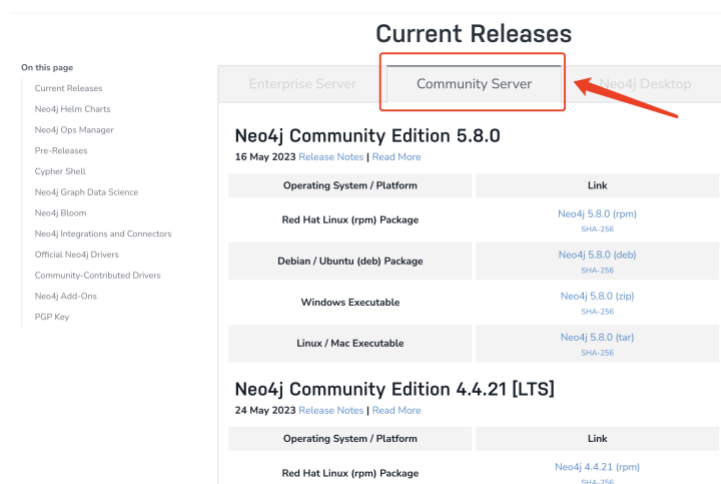
基于企业数据集，学习知识图谱构建以及知识图谱的可视化分析。

二、实验步骤

1. 环境安装与配置

a. 本地安装 Neo4j Community 版本

- i. 下载链接：<https://neo4j.com/download-center/#community>



- ii. 注意 1：Neo4j 5.x.x 需安装 java 17，Neo4j 4.x.x 需安装 java 11
- iii. 注意 2：从各种包的适配性角度考虑，不推荐安装最新版 Neo4j 5.x.x
- iv. 打开 <http://localhost:7474/>，若能成功则表示安装成功。打开后需要修改密码。

b. 安装相关 python 包：pandas、numpy、py2neo，建议使用 pip 安装

- i. 注意 py2neo 版本应该与 Neo4j 版本相匹配，否则无法运行
- ii. 若安装 Neo4j 4.x.x，推荐安装 py2neo 2021.2.3 版本

2. 基于 py2neo 在 Neo4j 数据库上构建企业知识图谱

参考代码：<https://github.com/jm199504/Financial-Knowledge-Graphs>

a. 数据读取：基于 pandas 读取所有的数据文件

- b. **连接 neo4j:** 注意由于 py2neo 版本差异, 一些版本对应的代码不同。以下是 py2neo 2021.2.3 版本的实例代码。

```
uri = "bolt://localhost:7687"
auth = ("neo4j", "password")
graph = Graph(profile="http://localhost:7474/", auth=auth)
```

- c. **创建实体:** 基于 py2neo.Node 依次创建企业、行业、股东、概念实体, 并且补充相关节点属性, 请避免实体重复创建问题。
- d. **创建关系:** 基于 py2neo.Relationship 依次创建企业-企业、企业-行业、企业-股东、企业-概念关系, 并且补充相关关系属性。
- i. 基于 py2neo.NodeMatcher 查找 node
- ii. 基于企业的股票之间的皮尔逊相关系数来创建企业-企业关系。皮尔逊相关系数的计算代码如下所示。

```
import numpy as np

def compute_corrcoef(stock_return: np.ndarray):
    corrcoef = np.corrcoef(stock_return)
    corrcoef[np.isnan(corrcoef)] = 0 # nan位置赋0
    return corrcoef
```

3. 基于 Crypher 语言, 在 <http://localhost:7474/> 上以平安银行为例完成以下可视化查询。鼓励进行更多其他的有意义的可视化查询。

参考代码: <https://github.com/jm199504/Financial-Knowledge-Graphs>

- a. 查看平安银行的持股股东
- b. 查看与平安银行的相关系数绝对值>0.8 的企业实体
- c. 查看所有与平安银行共处一个行业的企业

三、提交内容

1. 实验报告: 企业图谱构建和可视化分析报告。
2. 实验代码: 企业图谱构建代码。

注意: 请在学在浙大平台上提交, 请勿上传数据集。