由中国人民大学信息资源管理学院冯惠玲教授、刘越男教授、严承希博士和哲学院王小伟副教授主讲的《数字人文导论》是中国人民大学本科生通识核心课程之一，面向全校本科生开设。课程内容包括数字人文的基本概念、历史源流、发展脉络、基本方法和技术、前沿议题等。本系列推文是该课程学生的学习成果展示。

一

团队成员

李沅达 财政金融学院

王可心 财政金融学院

王雨飞 明德书院

二

项目背景

《红楼梦》是中国古典四大名著之一，属于章回体长篇小说，共120回。小说以贾、史、王、薛四大家族的兴衰为背景，展现中国古代社会世态百相，揭露了中国古代封建社会的各种矛盾和封建制度的黑暗腐朽，被誉为“封建社会的百科全书”。一直以来，对《红楼梦》的各种研究数不胜数，而《红楼梦》中人物的社会关系也是一个值得研究的问题，“人是各种社会关系的总和”，对不同人物的社会地位以及人与人之间的社会关系的研究对进一步研究书中所描述的社会及各种矛盾有正向推动作用。据相关统计，《红楼梦》中出场过的大大小小的人物共有983人，主要人物129人，人物关系错综复杂，因此很难直接分析人物之间的社会关系，建立系统的框架，不利于进一步阅读和理解这一经典名著。

三

计划解决的问题

我们希望能通过对《红楼梦》中人物社会关系所进行的数据可视化处理——建立知识图谱，方便读者了解该人物与所有与之相关的人物之间的社会关系，更好地阅读和理解《红楼梦》。

四

项目实施过程

1.创建知识图谱

《红楼梦》人物社会关系这一数据集共包含188条数据，涉及书中主要人物128个。每一条数据的大致格式为：人物名+与之相关的另一个人物名+两者之间的关系（亲戚关系/隶属关系/朋友关系/恋人关系等）+该人物居住的府邸+与之相关的另一个人物住的府邸。该数据集中涉及的人物绝大部分都是贾家府中人物，也存在少量史家、王家、林家和其他家族的人物。其中与贾宝玉相关的数据共有31条，位列第一；其次是贾母，与之相关的数据共18条；第三是林黛玉，与之相关的数据共有13条。这三者是数据集中包含的所有人物之中出现次数最多的，也表明了三人在书中的重要地位，后续做出的知识图谱应该也相对复杂。

图数据库作为NoSQL的分支，有助于直观且清晰地呈现较为复杂的人物关系图谱。Neo4j是图数据库的主要代表。本组使用Neo4j对红楼梦人物社会关系数据集进行处理，建立知识图谱，主要采用了两种方法。

法一：运用Neo4j基础语法

在Neo4j中，可通过Cypher命令创建节点，本案例中，数据集内的每个人物均被作为节点处理。可通过CREATE和MATCH语句可建立不同人物节点以及不同人物、人地间的关系。如下图：

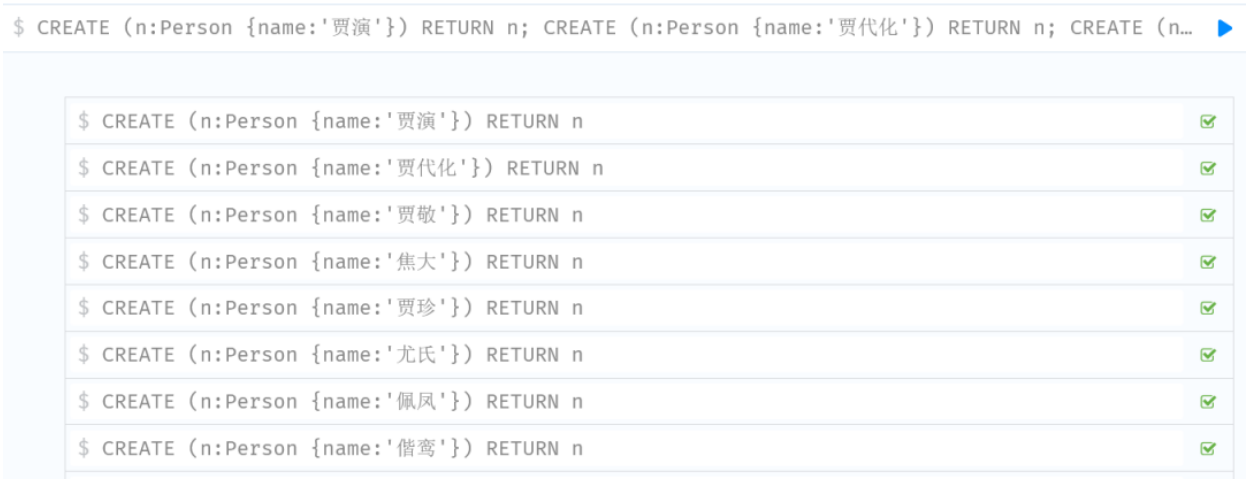

图 1 创建人物节点

图 2 创建地点节点

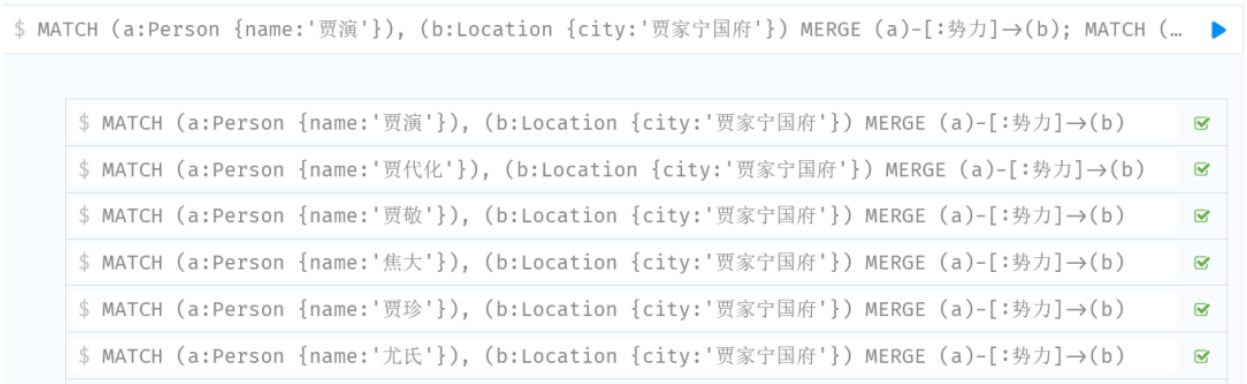

图 3 匹配人地关系

图 4 匹配人物关系

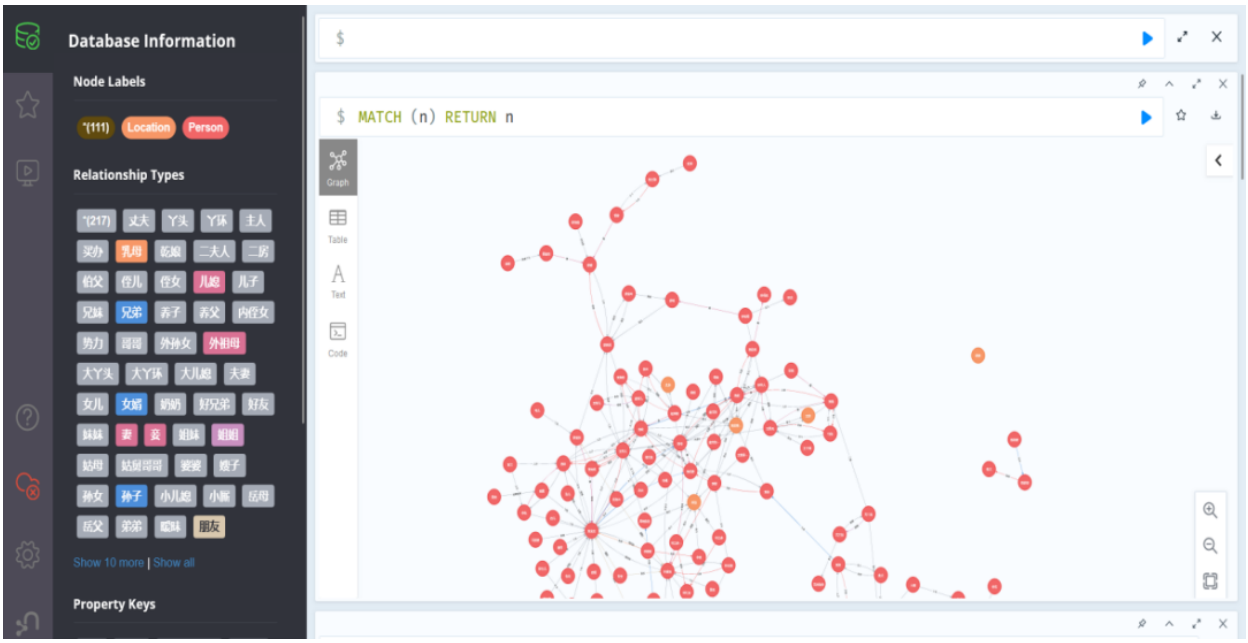

图 5 方法一 Neo4j知识图谱结果查询

法二：利用python创建知识图谱

（1）数据导入：将csv文件拷贝到Neo4j目录下的import文件

（2 ）创建知识图谱代码

```python
import csv
import py2neo
from py2neo import Graph, Node, Relationship, NodeMatcher
g = Graph('http://localhost:7474', user='neo4j', password='123456')

with open('D:/neo4j-community-3.5.31/import/relation.csv', 'r', encoding='utf-8') as f:
    reader = csv.reader(f)
    for item in reader:
        if reader.line_num==1:
            continue
        print("当前行数：", reader.line_num, "当前内容：", item)
        start_node=Node("Person", name=item[0])
        end_node=Node("Person", name=item[1])
        relation=Relationship(start_node, item[3], end_node)
        g.merge(start_node, "Person", "name")
        g.merge(end_node, "Person", "name")
        g.merge(relation, "Person", "name")
```

图 6 python pycharm中的知识图谱代码

（3 ）打开Neo4j网页并运行代码

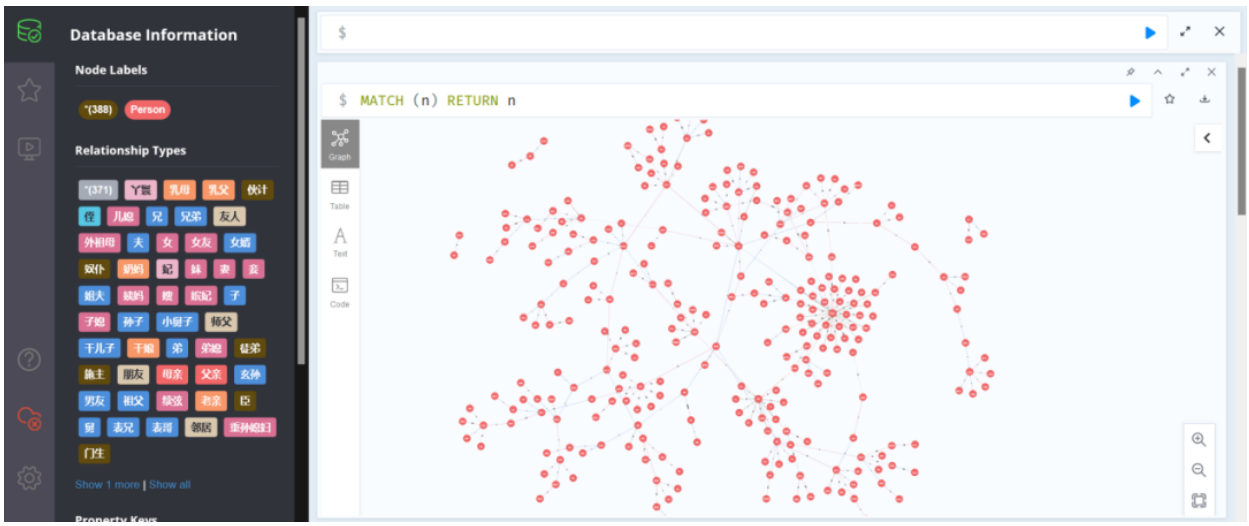

图 7 方法二Neo4j知识图谱结果查询

2.网页设计

在实现红楼梦社会关系知识图谱的建立后，设计专门网页以便于对其进行更好的可视化效果展示，并利用网页补充了相关主要人物的资料。

图 8 网页设计首页

考虑到核心人物及读者兴趣所在，我们特别地建立了贾宝玉、金陵十二钗（正钗、副钗、又副钗）的页面，在页面中展示了人物信息和相应的人物关系图谱。通过超链接，便可跳转至读者想要了解的人物所在页面。

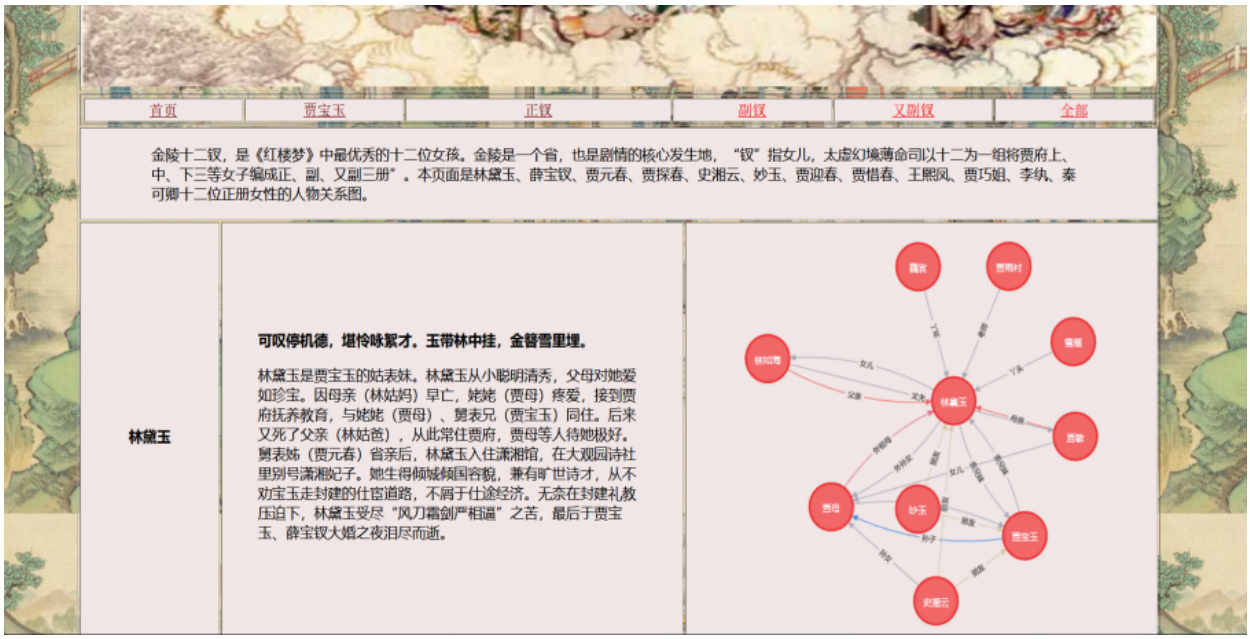

图 9 金陵十二钗正钗页面

五

项目成果

1.数据分析结果

在批量创设节点及建立关系后，可以得到数据集的整体人物图谱。《红楼梦》中的人物关系错综复杂却又紧密勾连，以贾家荣国府、宁国府为中心展开，其中又以男主角贾宝玉、其祖母贾母、其父贾政和堂兄贾琏为重要节点。通过Neo4j图数据库，我们对红楼梦中的人物关系进行了直观而清晰的展示，方便读者在此基础上更为系统地理解红楼的主要内容及其精神内核。对关系图谱进行分析，我们可以得到以下结论：

（1）贾宝玉是整部《红楼梦》毫无争议的主角和叙事核心，串联起了小说复杂的人物网络。作为着墨最重、出场最多的人物，贾宝玉是整部红楼当之无愧的主角，曹雪芹以贾宝玉的视角刻画出贾、史、王、薛四大家族的生活百态，脂砚斋批语曾言“贾宝玉为诸艳之冠”。在我们构建的关系图谱中，贾宝玉的人物关系也最为复杂和重要：

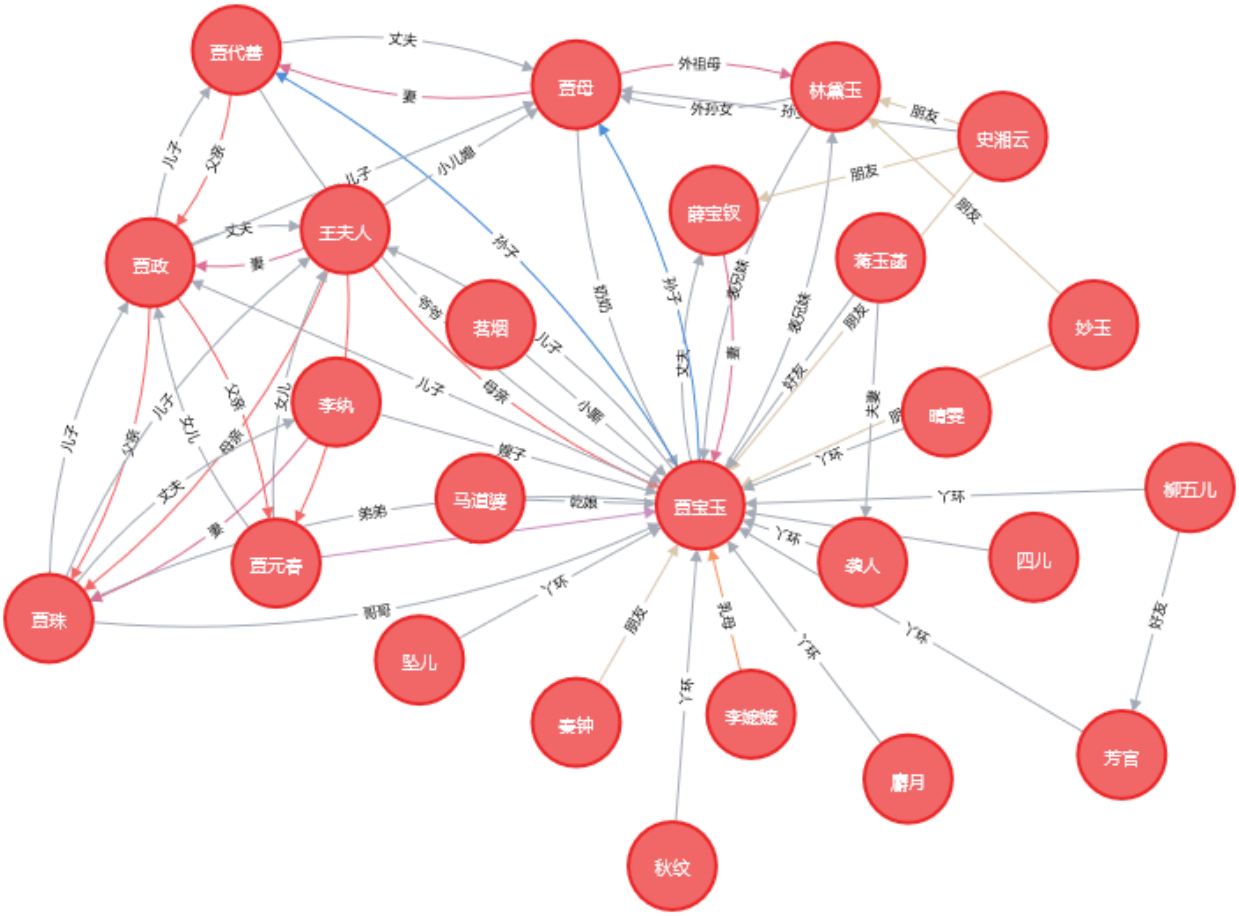

图 10 贾宝玉人物关系图

由图10可以发现，数据集中与贾宝玉存在直接关系的人物共24位，占全部人物的18.75%；而与其存在间接关系的人物则有82位，占比超六成，达到64.06%，这也与前文对数据集的描述性统计结果相吻合。

（2）部分核心人物在关系网络中的地位并不突出，如薛宝钗（4位相关人物）、王熙凤（7位相关人物）。

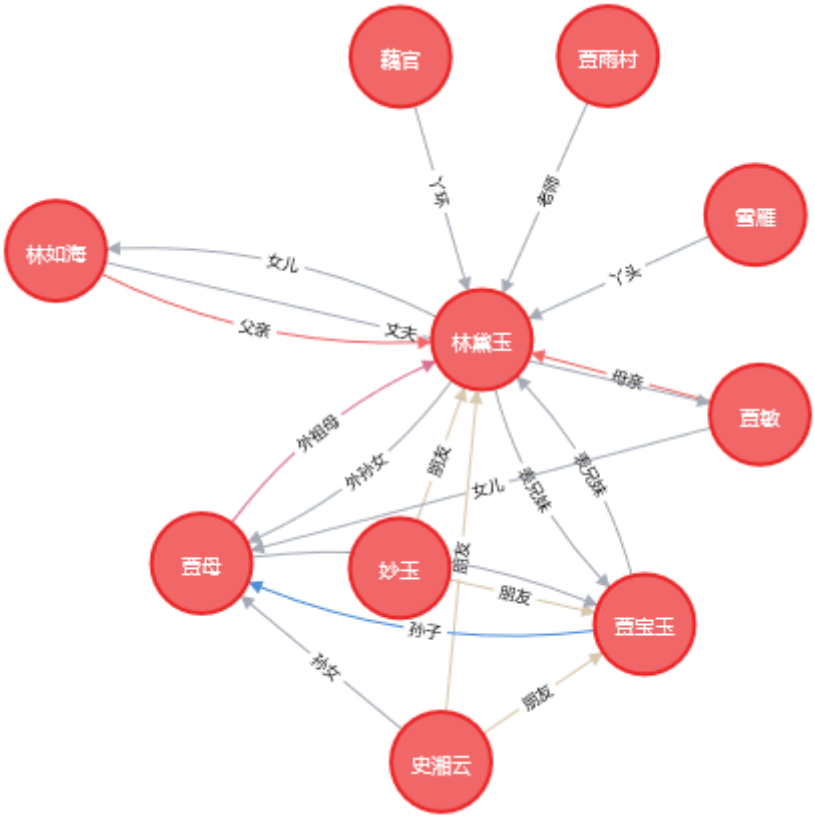

图 11 林黛玉人物关系图

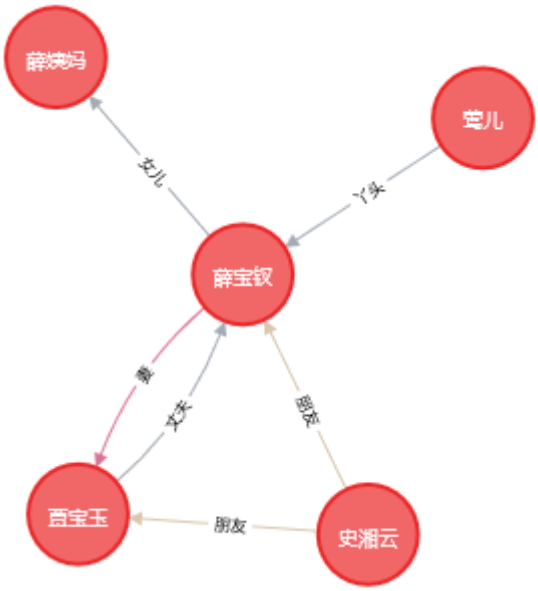

图 12 薛宝钗人物关系图

针对这一反常现象背后的原因，我们认为可能是所选取的数据集本身相对简单浅显，人物关系以血亲为主，非贾府出身的人物重要性在一定程度上被淡化。在后续工作中，可通过增加数据量的方式予以解决。

（3）从人物关系的视角来看，贾府的由盛转衰贯穿全局，是除了宝玉、黛玉、宝钗的三角爱情悲剧外，《红楼梦》的又一条主线。

2.意义

（1）将文字数据转化为知识图谱，便于读者更快理清交错复杂的人物关系，感受《红楼梦》的精神内核。

（2）通过查询功能可更为轻松地获取某个（类）特定人物的关系网络，为专门研究提供了一定便利。

（3）在关系图谱基础上建立网页平台，向大众宣传、推广《红楼梦》这一古典文学瑰宝。

排版：高宇博