****

**管理学院**

**2023-2024学年第一学期**

**《管理定量分析与软件应用》**

**《三体》小说的社会网络分析报告**

**小组成员及学号**：

阿依孜巴·阿力木江 U202111348

姑丽米热·买明 U202016133

张宇锋 U202116240

**班级**： 信管2101班

**时间**： 2023年12月31日

目录

[**一．问题提出与分析 3**](#_Toc154339964)

[社区结构与小说情节 3](#_Toc154339965)

[关键人物与故事发展 3](#_Toc154339966)

[关系强度与人物互动 3](#_Toc154339967)

[灰衣主教角色的分析 4](#_Toc154339968)

[**二．数据处理 4**](#_Toc154339969)

1.[数据收集和预处理 4](#_Toc154339970)

2.[数据处理脚本 5](#_Toc154339971)

[3数据处理工具 6](#_Toc154339972)

[**三．建模分析 7**](#_Toc154339977)

1.网络图构建

2.计算网络指标

3.社区发现

4.中心性分析

5.关系强度分析

6.发现灰衣主教角色

[**四．结果讨论 48**](#_Toc154339978)

[**五．总结 53**](#_Toc154339979)

[**六．小组成员及分工 54**](#_Toc154339980)

《三体》：介绍了地球文明与三体文明的首次接触，背景设置在文化大革命期间和现代，讲述了地球人类首次发现外星文明的故事。

《三体Ⅱ：黑暗森林》：进一步展开了地球与三体文明的冲突，引入了“黑暗森林法则”作为宇宙文明间交互的基本原理，并探讨了面对外星文明威胁时的人类策略。

《三体Ⅲ：死神永生》：结束了整个系列，展示了一个宏大的宇宙视野，包括对宇宙的终极命运的思考，以及人类在宇宙中的位置和未来。

# 问题提出与分析

**社区结构与小说情节：**

在《三体》三部曲中，哪些特定的历史事件或情节转折导致了明显的社区结构变化？

如何量化并解释文化大革命、三体文明发现、以及宇宙文明危机等关键事件前后的社区结构变化？

**关键人物与故事发展：**

在《三体》三部曲中，哪个人物的中心性指标（如度中心性或中介中心性）在关键情节点（如与三体文明首次接触、地球三体组织形成、宇宙封锁等）有显著变化？

这些关键人物的网络中心性如何与他们在故事中的决策和行动相对应？

**关系强度与人物互动：**

在《三体》系列中，哪些人物间的强关系对故事情节的关键转折起到了决定性作用？

弱关系网络如何促进了信息的传播或误解，特别是在三体文明与地球文明的交互中？

**灰衣主教角色的深度分析：**

灰衣主教的社交网络策略（如联盟建立、信息控制）是如何影响《三体》中的重大决策和故事走向的？

# 二．数据处理

## 2.1数据收集和预处理

文本源：网上下载的《三体》三部曲全文，分为不同章节，存储于三体.txt。

## 2.2 数据处理脚本

1. modifyTxt.py

功能：中文文本分词和去除停用词。

处理过程：

使用jieba进行文本分词。

从“txt/hit\_stopwords.txt”中读取并应用停用词列表，以清洗文本。

代码读取原始文本文件（txt/三体.txt）

对于文件中的每一行，调用seg\_depart函数进行分词和去除停用词，然后将处理后的文本写入新的文件（new三体.txt）中。

这个过程将整本《三体》小说的原始文本转换为一个经过分词且去除了常见无关词汇的清洁版本。

2.main.py

功能：提取《三体》文本中的人物名称及其出现频次。

处理过程：

读取“txt/new三体.txt”文件中的文本。

使用spaCy的中文模型对文本进行处理，识别其中的人名实体。

统计每个人名在文本中出现的频次，并将结果保存到CSV文件。

3.time.py

此脚本用于将三体小说按照纪元时间段分为魔法时代，黄金岁月，危机纪元 威慑纪元 威慑后 广播纪元 掩体纪元 银河纪元

定义了一个名为extract\_chapters的函数，用于提取特定章节范围内的文本。

3. statisticalData.py

功能：执行情感分析，并处理更复杂的数据。

处理过程：

导入库：

导入了csv、collections.defaultdict、re（正则表达式）、spacy、jieba、jieba.analyse和SnowNLP等库，用于数据处理、文本分析和情感分析。

情感分析函数 (chinese\_sentiment\_analysis)：

使用jieba对文本进行分词。

使用SnowNLP对分词结果进行情感分析，得到情感倾向（积极、消极或中性）和情感分数。

处理多个文本文件：

定义了一个时间段列表，用于指定不同的文件名。

对于每个时间段，从相应的文本文件中读取内容。

读取人物名单：

从txt/people.txt文件中读取人物名单，并存储在一个集合中。

文本分段：

使用正则表达式将每个文件的内容分割成段落。

人物关系和情感分析：

为每个人物对创建两个字典：all\_relations（基于计数）和all\_relations0（基于情感分析）。

使用jieba对整个文本进行分词，并统计每个人物名字的出现频次。

对于每个段落，如果段落包含任意两个人物的名字，则计算这两个人物的关系强度（基于出现频次的乘积）和情感分数（基于情感分析）。

写入CSV文件：

将人物间的关系强度和情感分数分别写入两个CSV文件夹：txt/person\_relationships/和txt/person\_relationships0/。

同时，将每个人物的出现频次写入另一个CSV文件夹：txt/new\_person\_counts/。对每个时间段的文本进行情感分析，并考虑文本中人物的共现情况。

此脚本的重点是在不同时间段内分析人物之间的关系强度。

## 2.3数据处理工具

1 自然语言处理

spaCy：用于人名的实体识别和词性标注。

jieba：执行中文分词，关键于文本的初步处理和情感分析的准备。

SnowNLP：对中文文本段落进行情感分析，判断情感倾向。

2 数据存储

CSV格式：将处理结果保存在CSV文件中，便于后续分析和可视化。包括人物关系数据、情感分析结果和人物出现频次。

这些脚本共同构成了一个完整的数据处理流程，用于从《三体》小说中提取关键的社会网络数据。通过分词、实体识别、情感分析等NLP技术，获得关于人物出现频次、人物关系程度的数据，分别为person\_relationships0.csv和new\_person\_counts.csv,之后的各种分析都将基于这两个数据文件。采用文本情感分析的结果作为人物的关系权重。

# 三．建模分析

## 3.1网络图构建

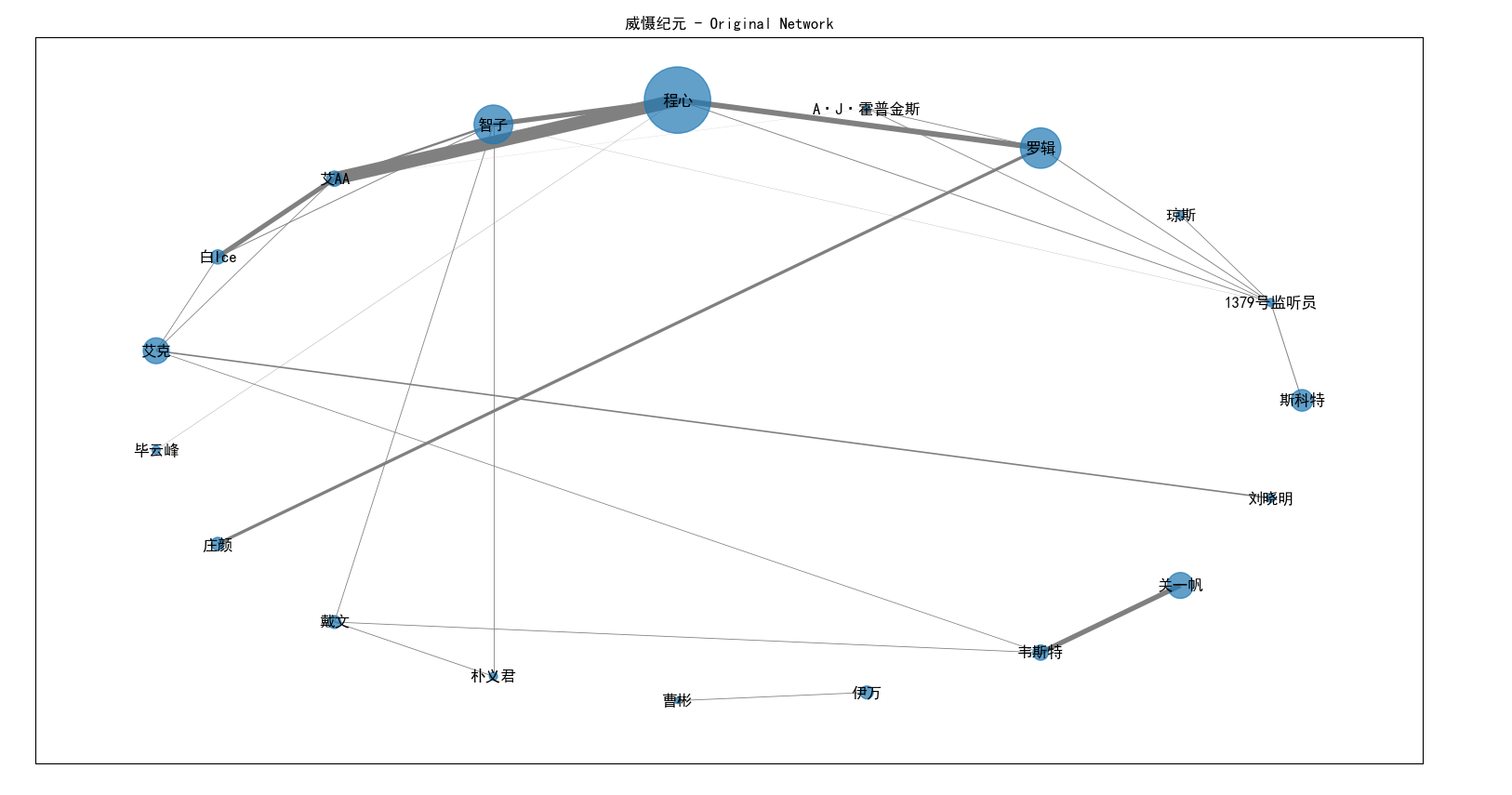
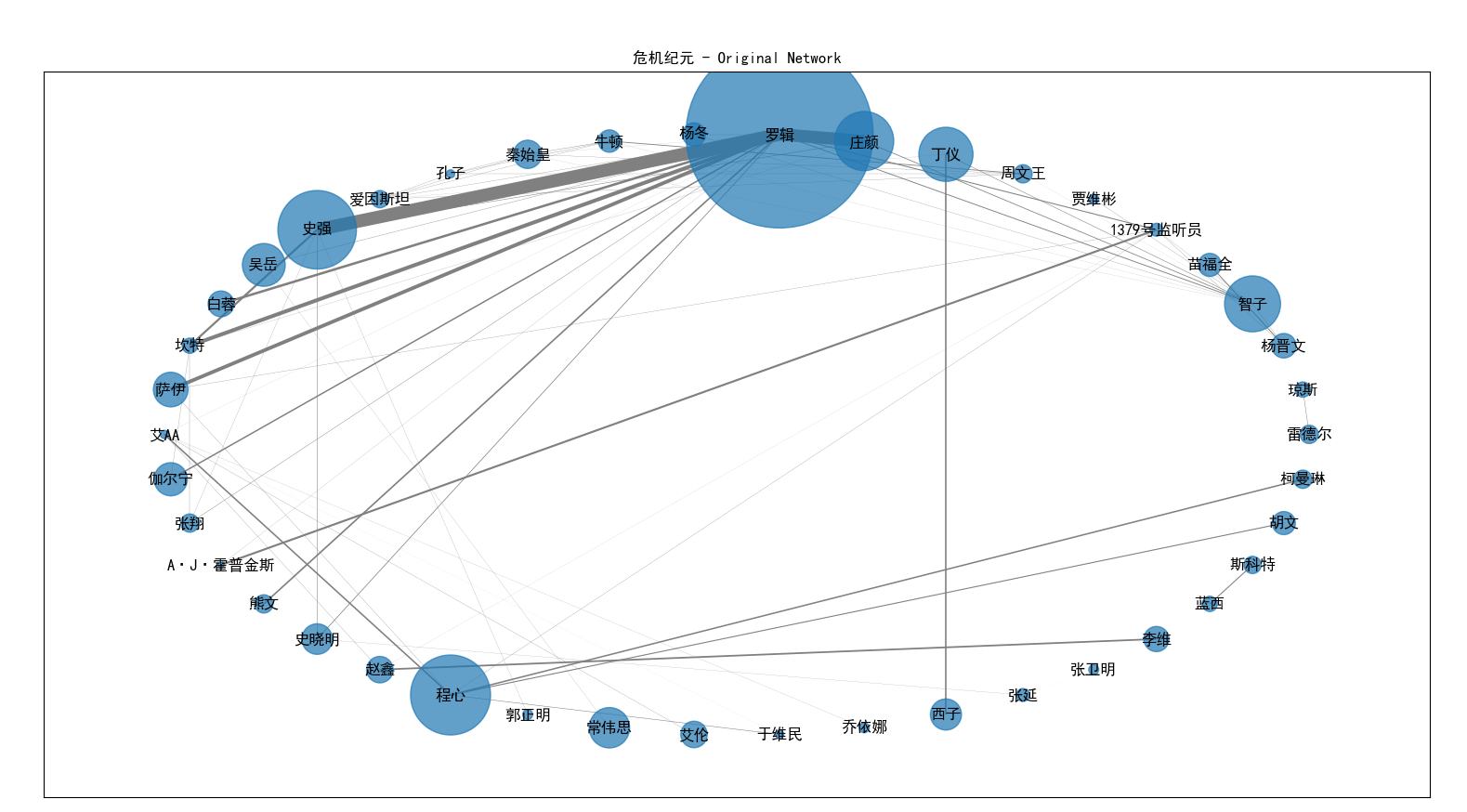
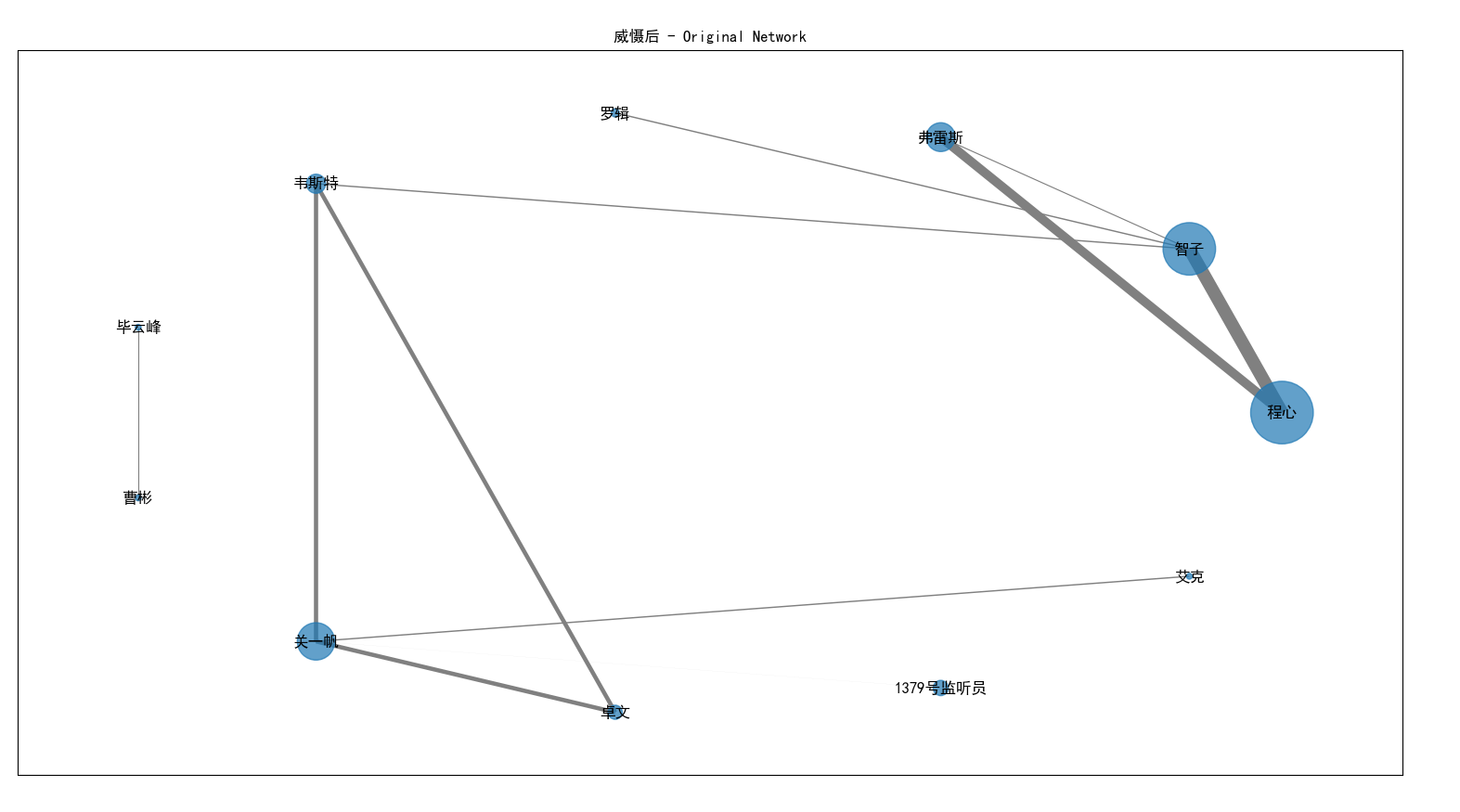
使用nx构建基于人物关系的社会网络图。

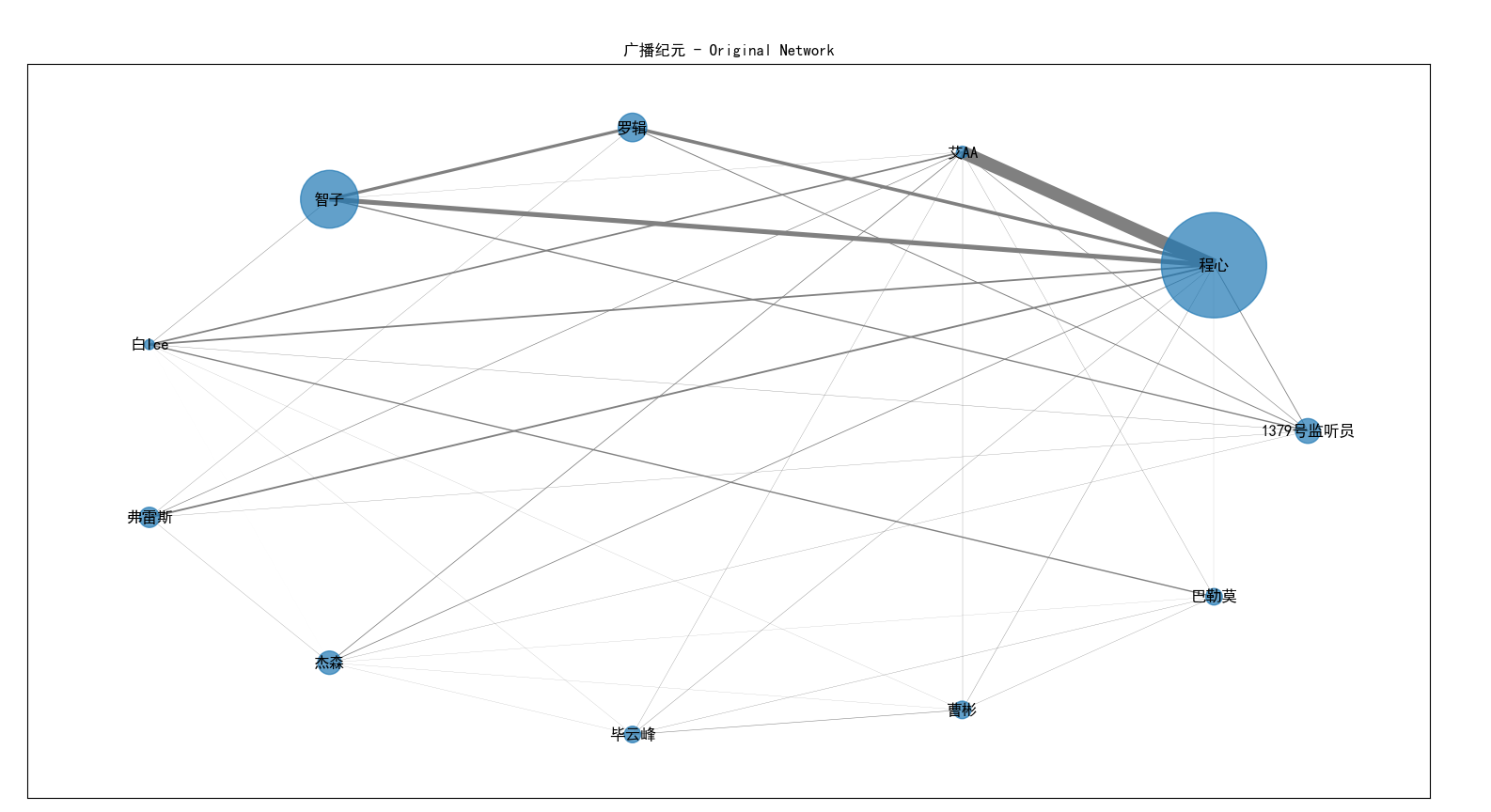
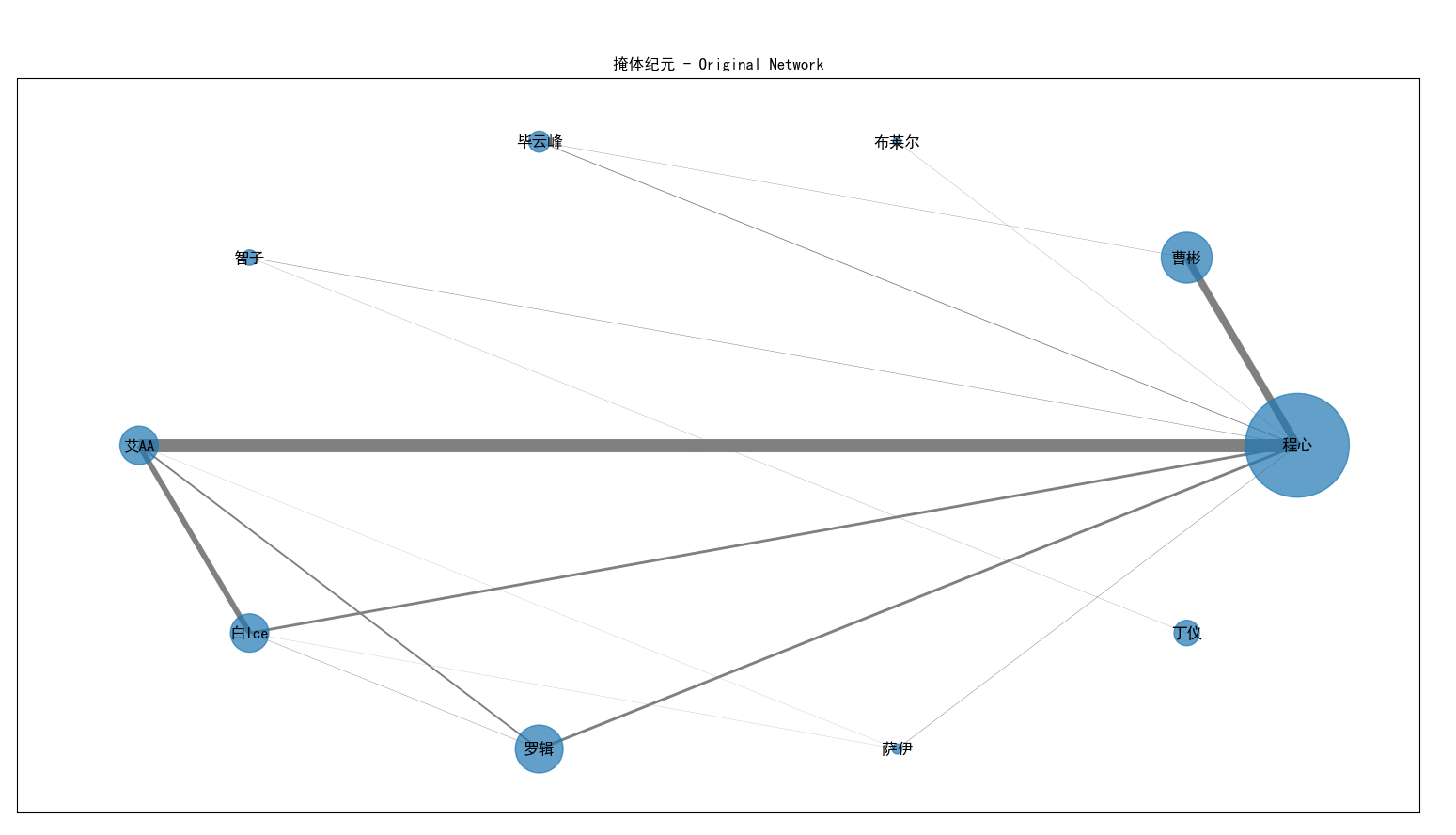
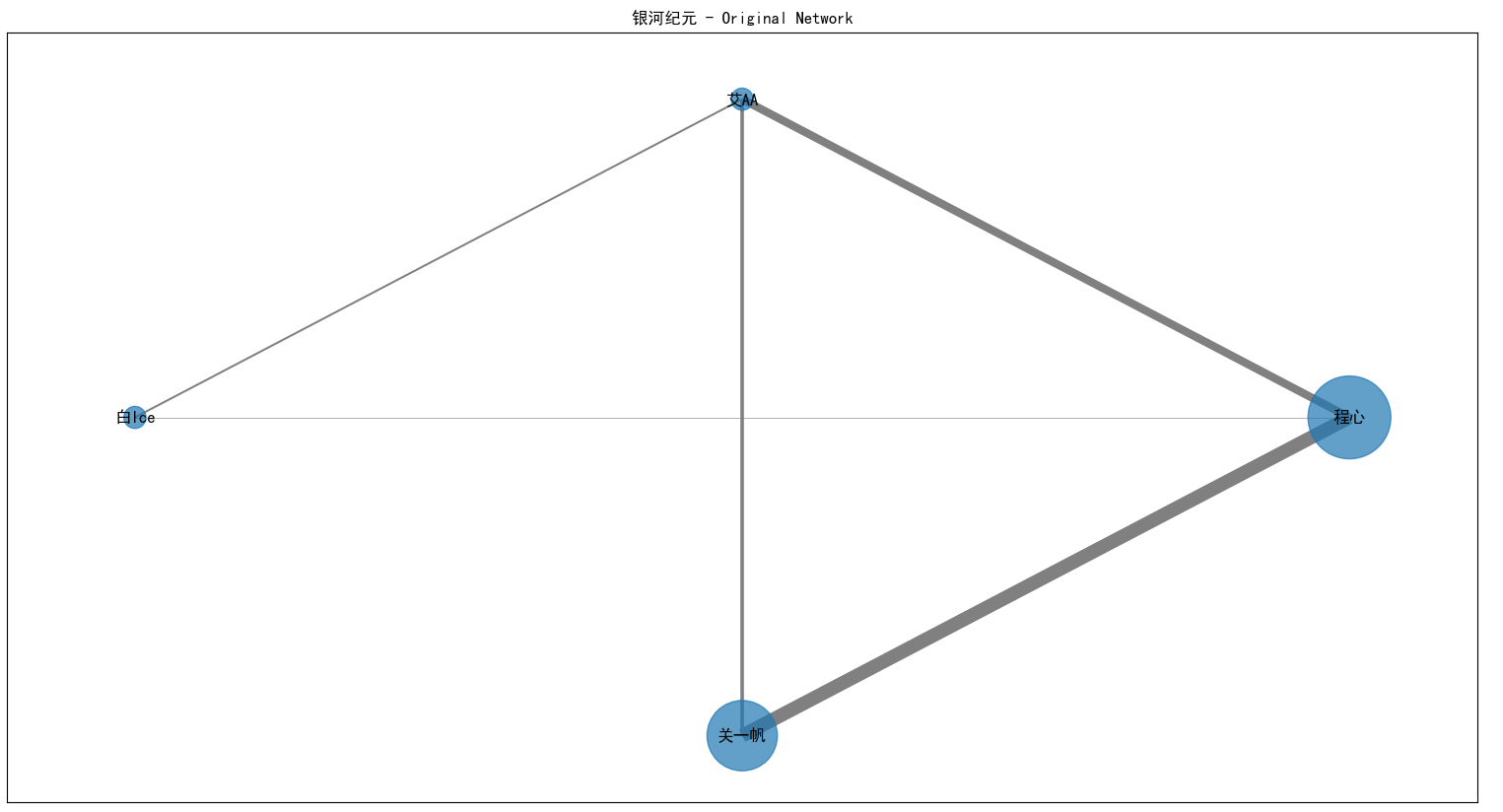
时间顺序为魔法时代→黄金年代→危机纪元→威慑纪元→威慑后→广播纪元→掩体纪元→银河纪元；由于在数据清洗之后，魔法时代只有一个人物狄奥伦娜

，且与其他纪元的人物没有任何联系，所以魔法时代没有社会网络图（整个魔法时代只有两章，作用只是为了引入四维碎块）

网络图中，节点的大小代表人物在此纪元的登场次数，边的粗细代表两人的关系强度，边画的越粗，代表两个在网络中的联系就越强

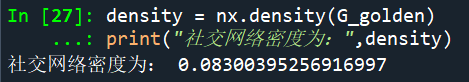


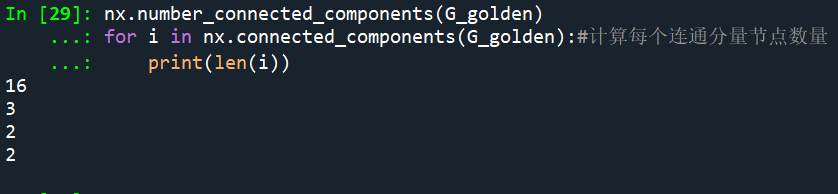
3.2计算网络指标。

黄金年代：

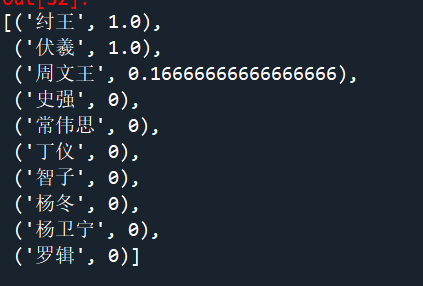


平均路径长度

连通分量



聚集系数

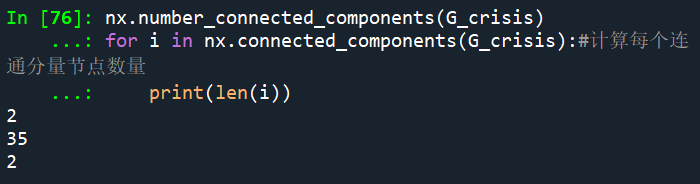


危机纪元：

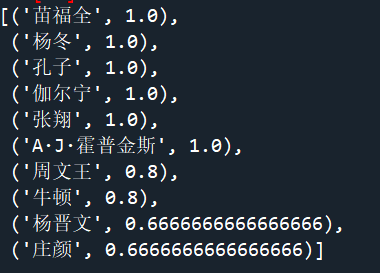
社交网络密度

社交网络密度为： 0.08502024291497975

连通分量



聚集系数

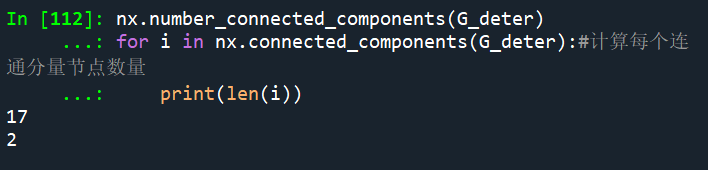


威慑纪元：

社交网络密度

社交网络密度为： 0.15204678362573099

连通分量



聚集系数

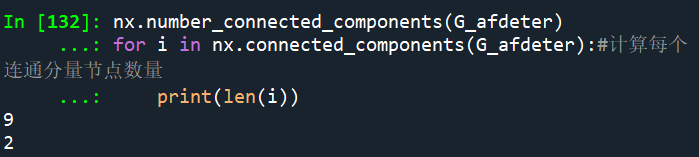


威慑后：

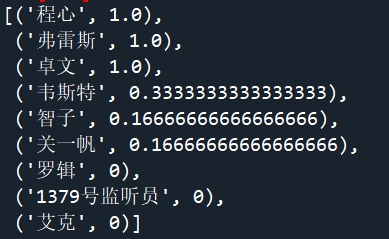
社交网络密度

社交网络密度为： 0.2

连通分量



聚集系数

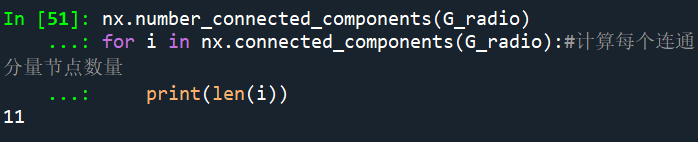


广播纪元：

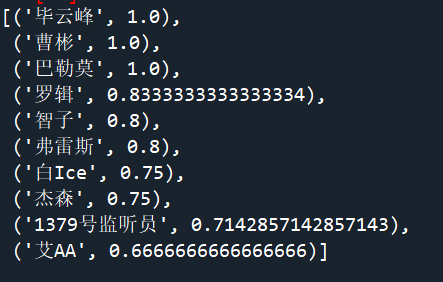
社交网络密度

社交网络密度为： 0.6727272727272727

连通分量



聚集系数

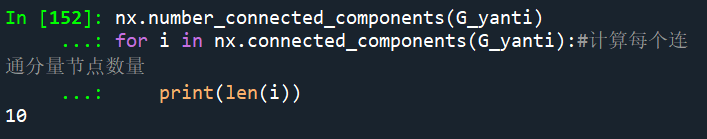


掩体纪元：

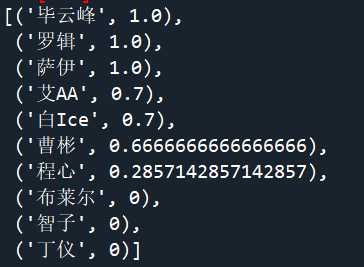
社交网络密度

社交网络密度为： 0.37777777777777777

连通分量



聚集系数

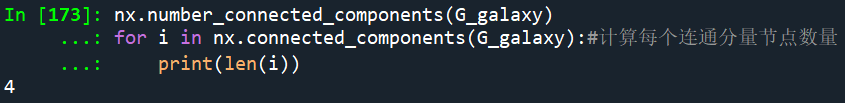


银河纪元：

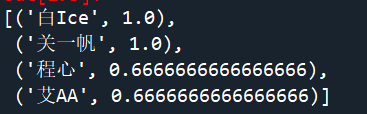
社交网络密度

社交网络密度为： 0.8333333333333334

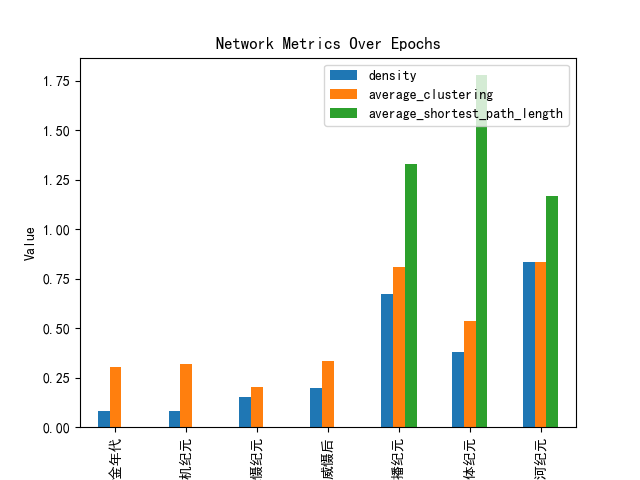
连通分量



聚集系数



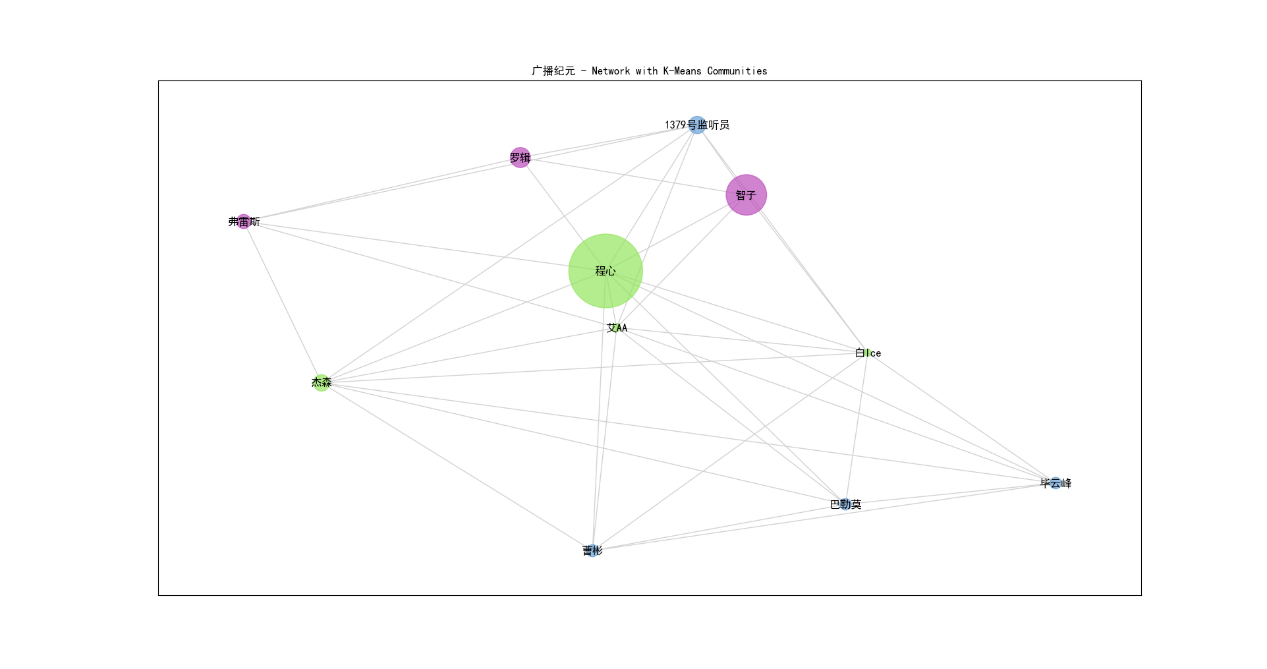
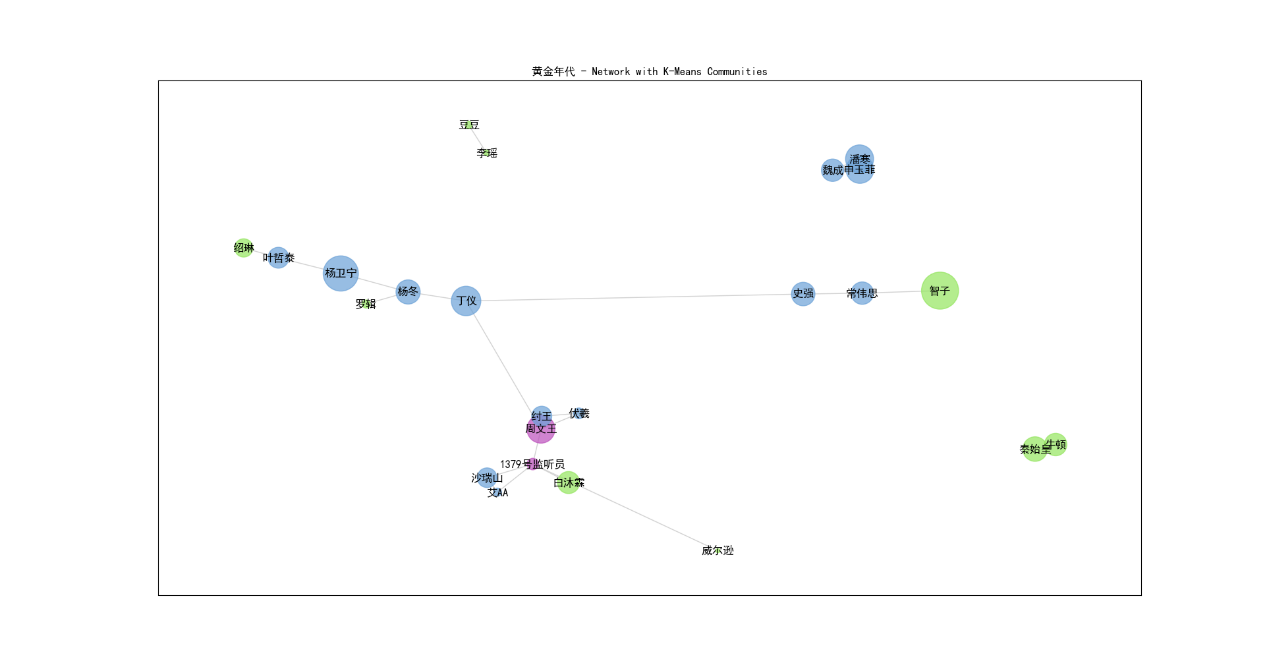
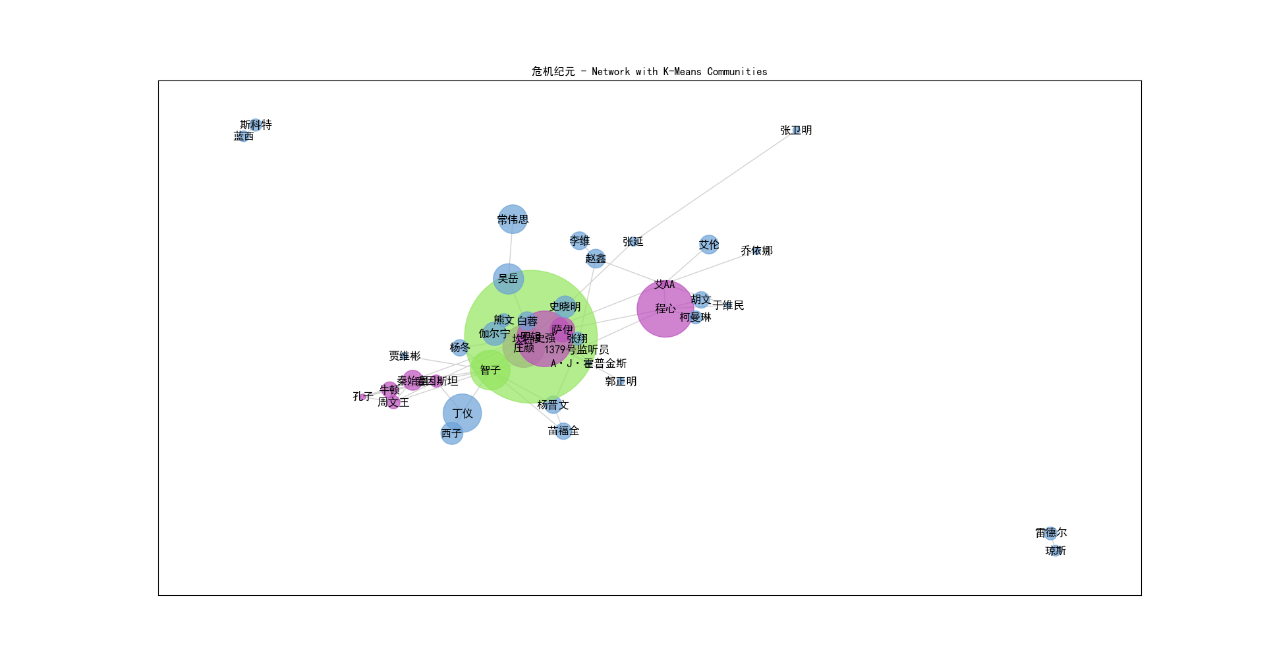
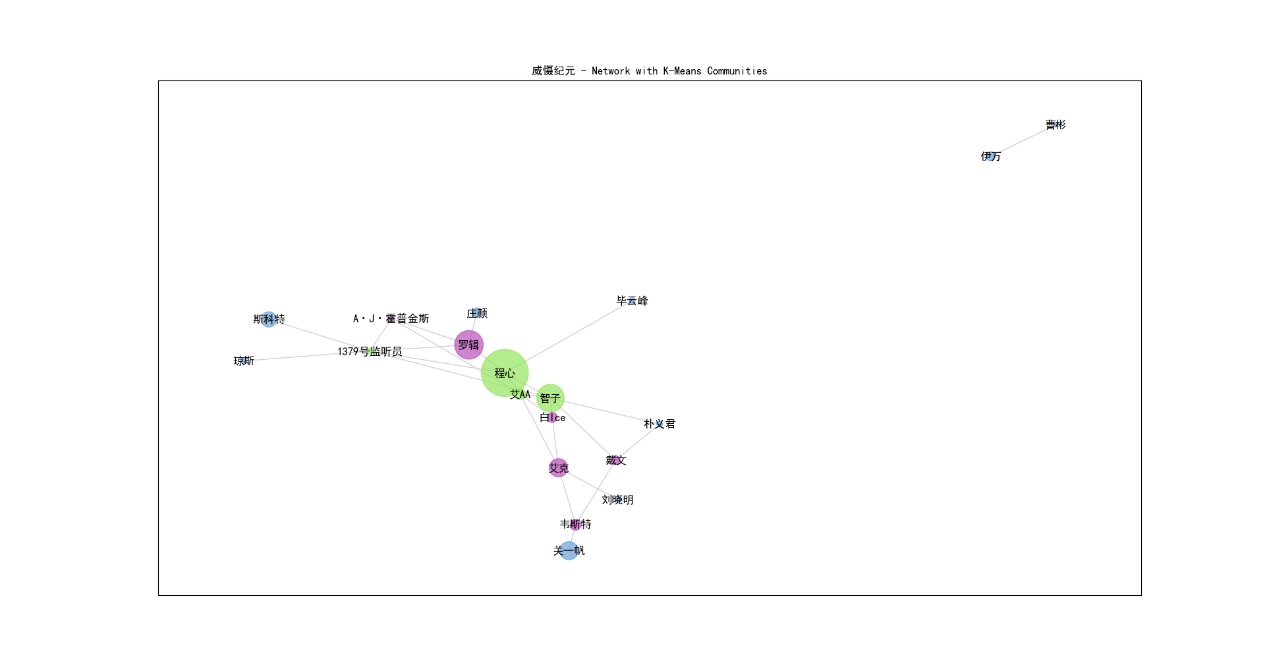
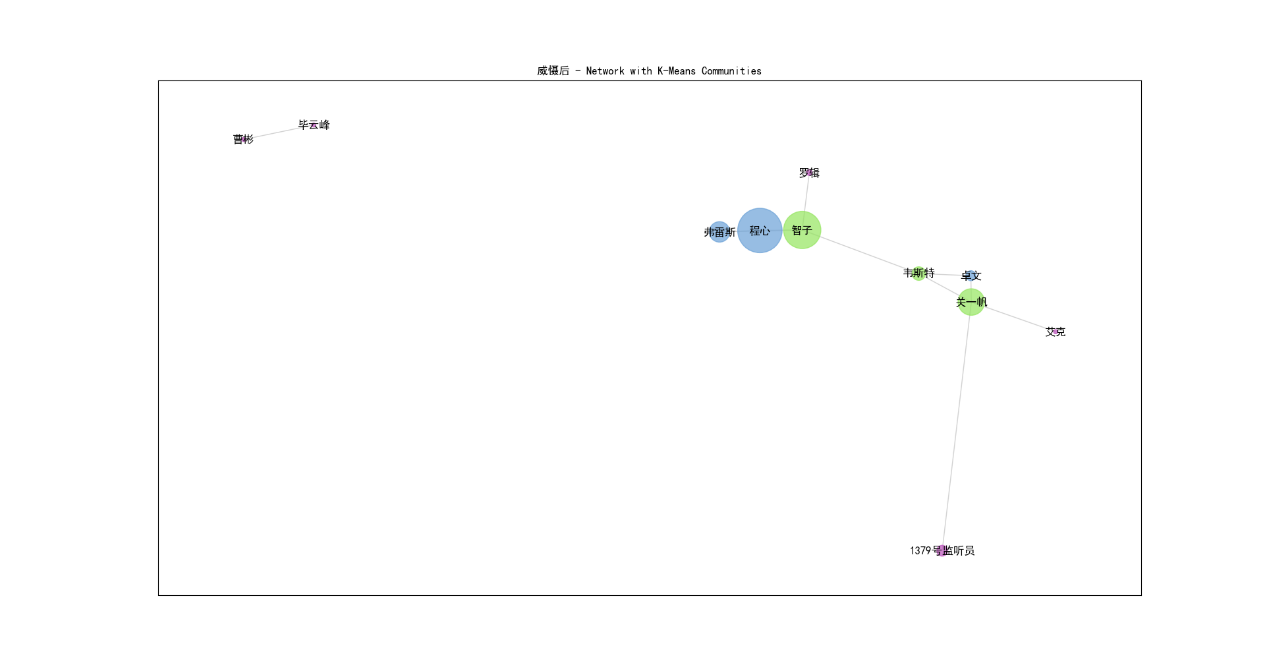
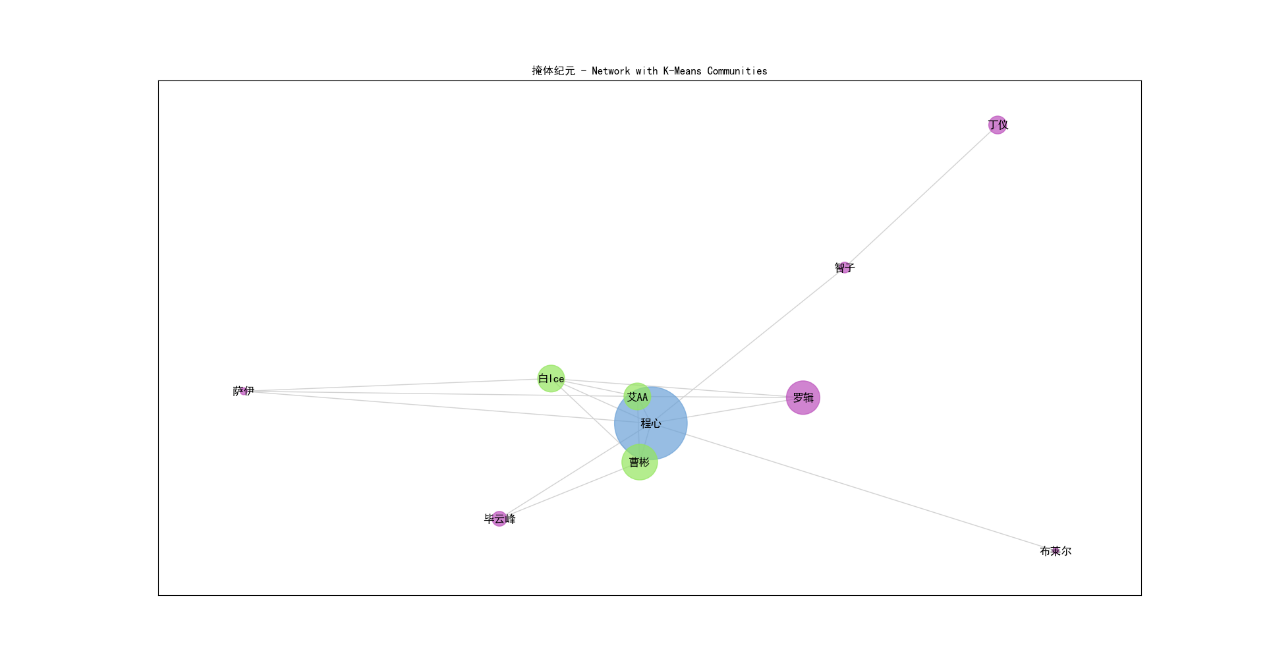
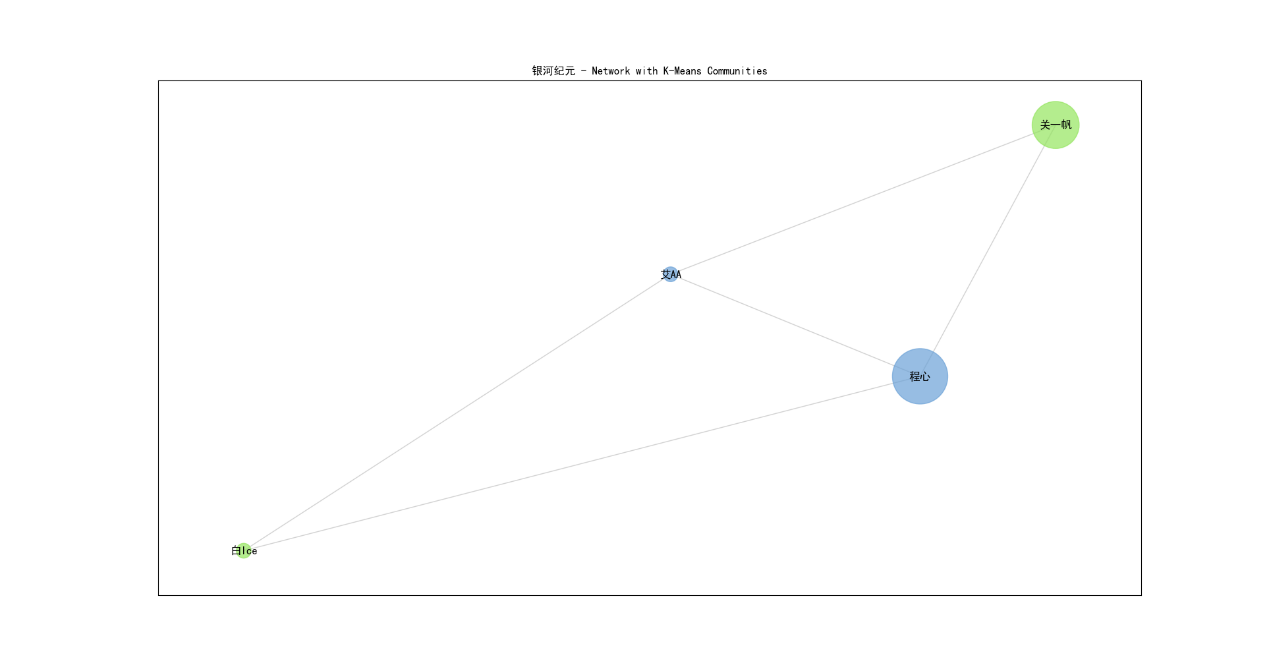
计算完每个纪元的主要网络指标后，我们定义函数calculate\_network\_metrics 来计算网络的密度、平均聚集系数和平均最短路径长度。使用pandas 和 matplotlib 将这些指标绘制成条形图。



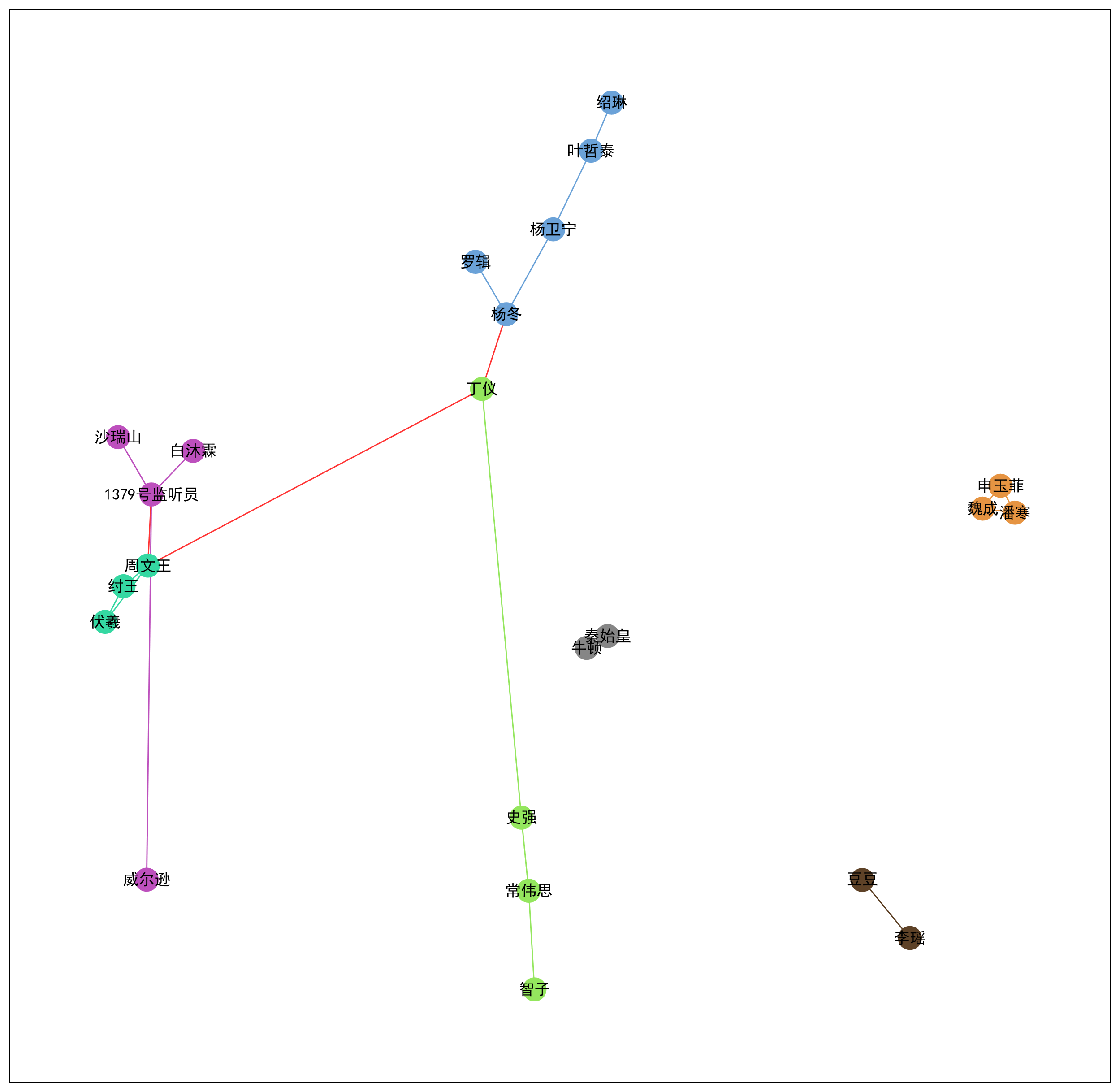
3.3社区发现:

应用 K-Means 聚类算法和贪婪模块度算法发现网络中的社区结构：

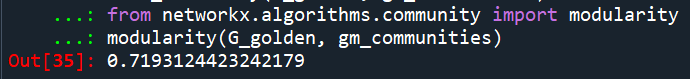
K means

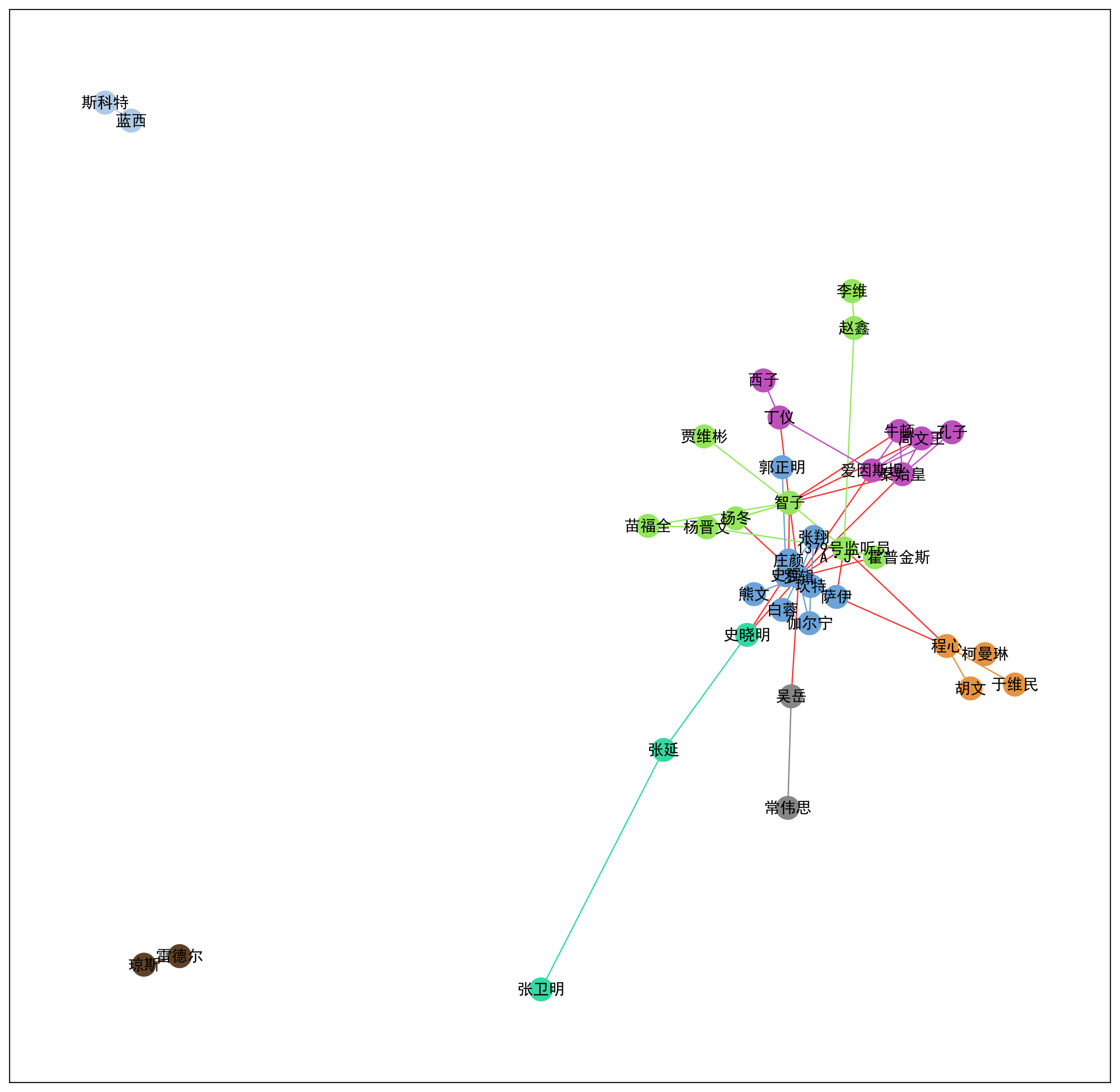


贪婪算法：

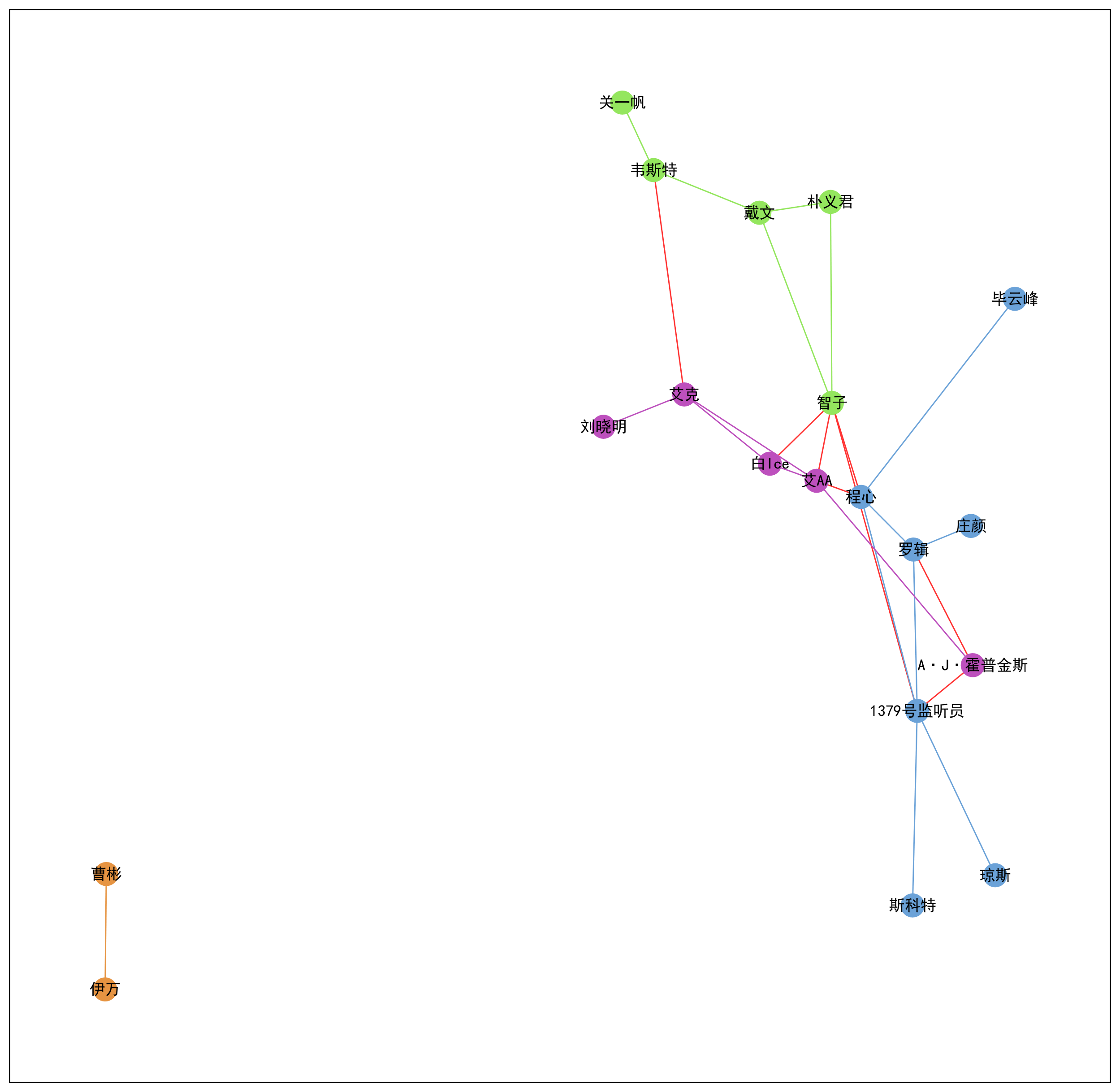


模块度

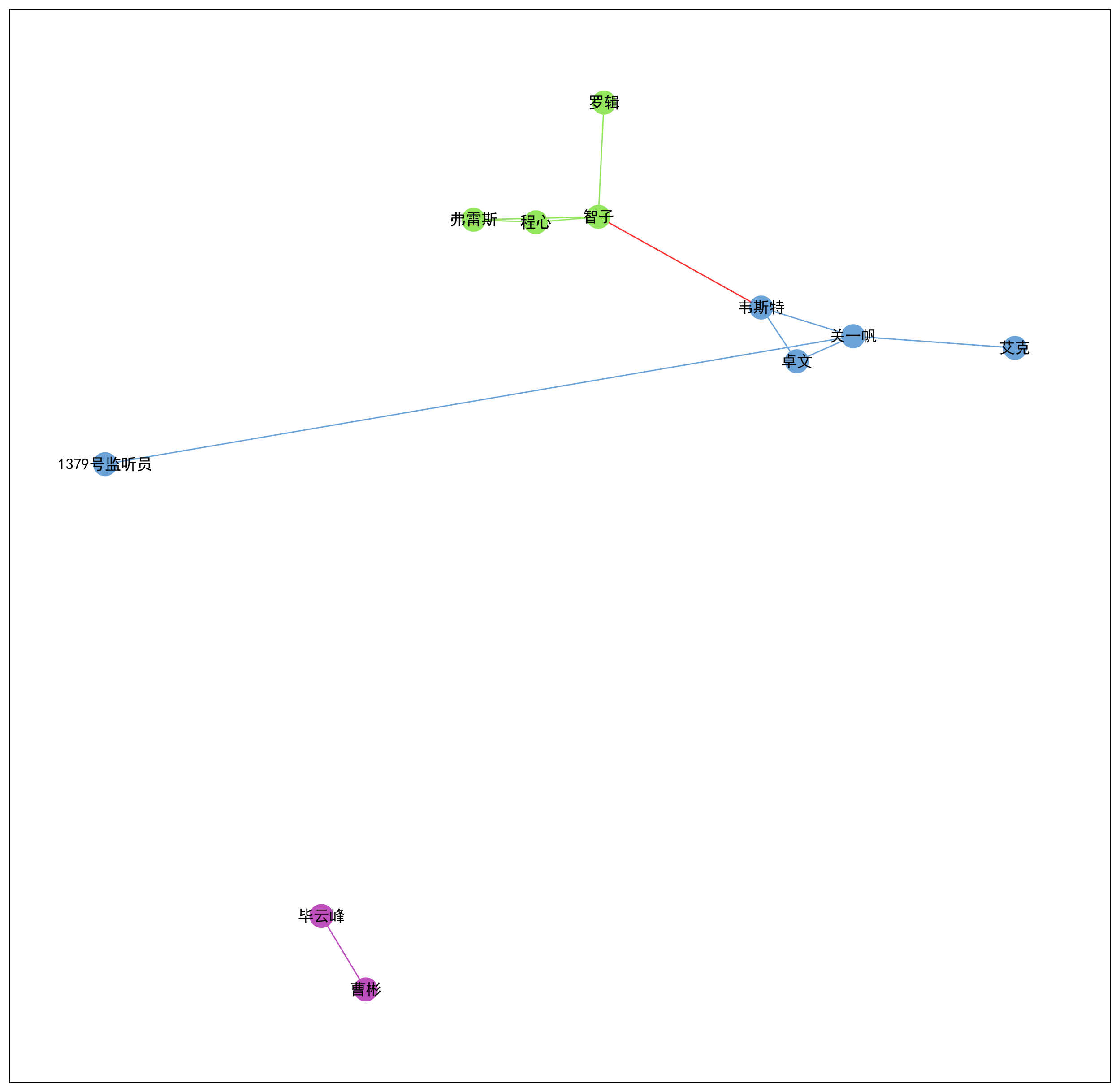




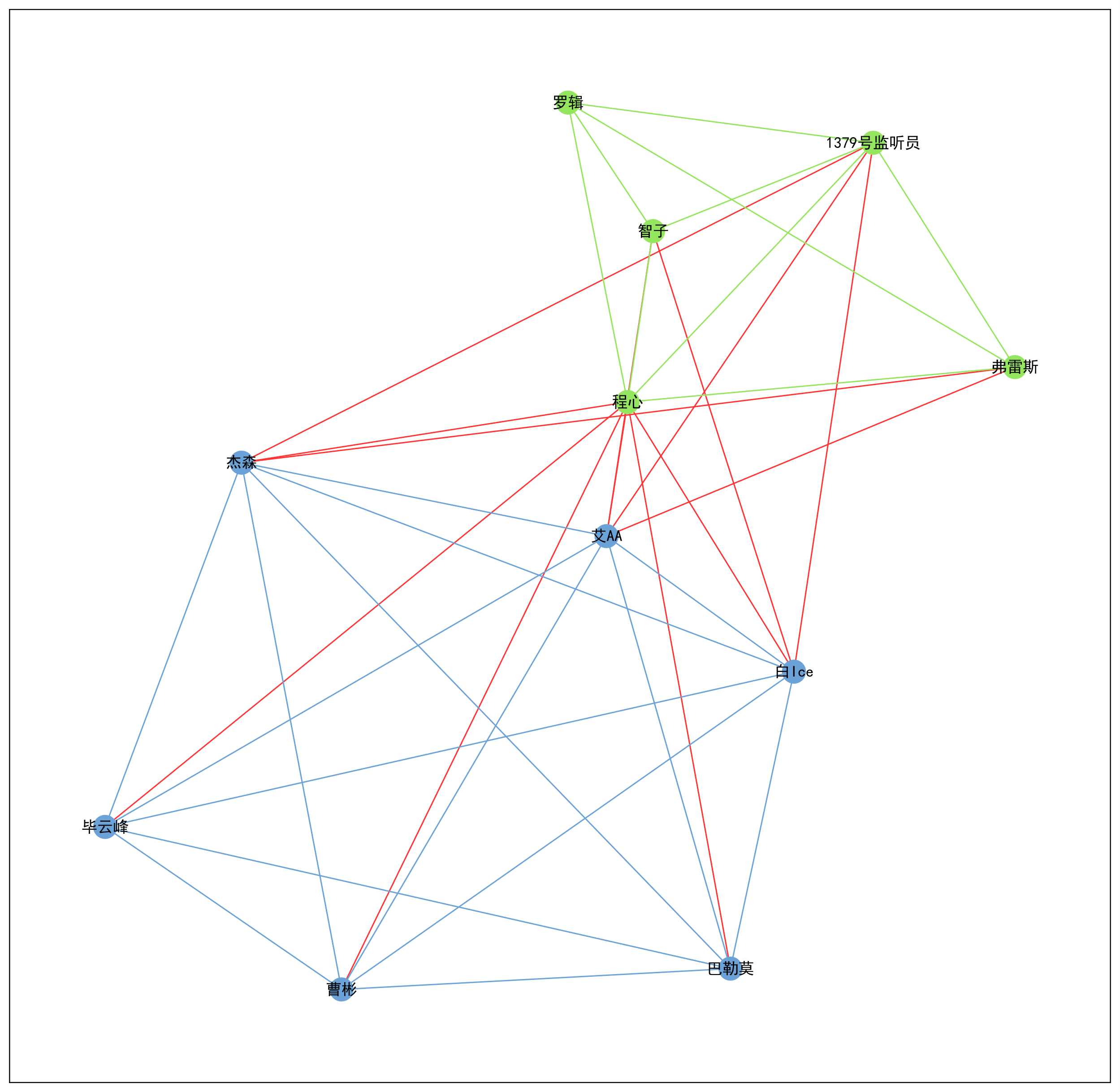
模块度：0.3793155171764883



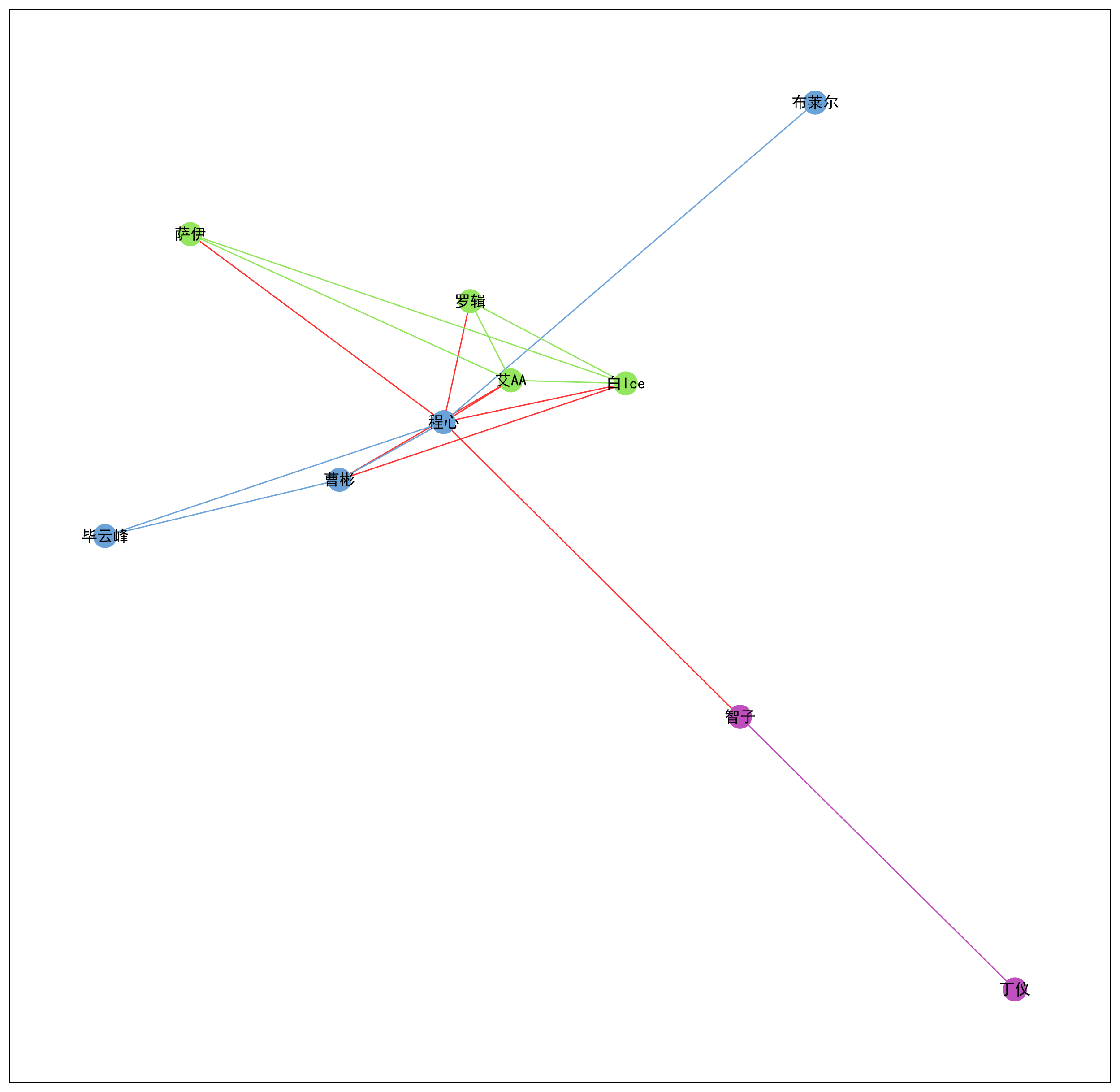
模块度：0.21294319541024403



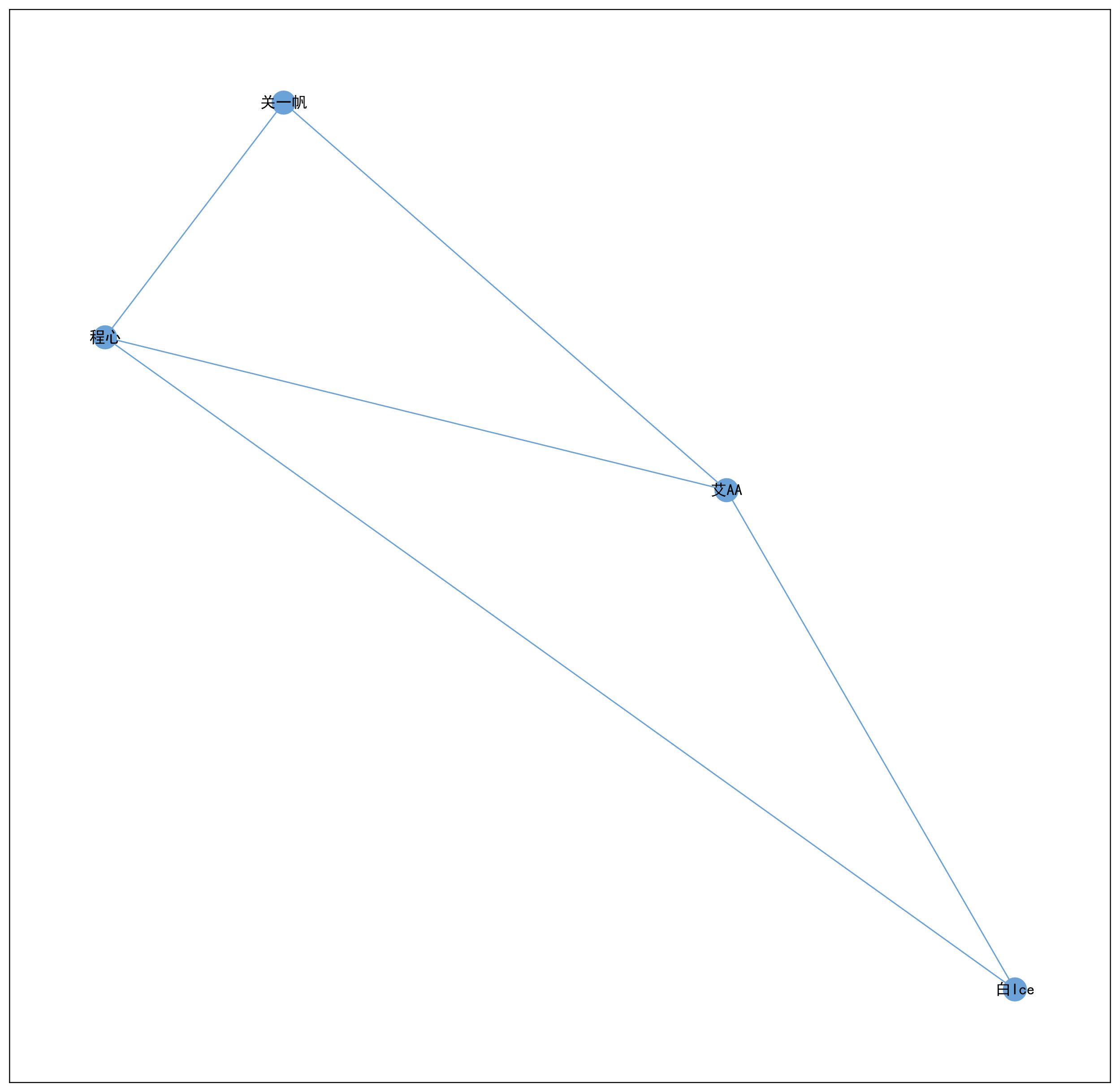
模块度：0.459259542437742



模块度：0.016409826461773747



模块度-0.018882360252913396



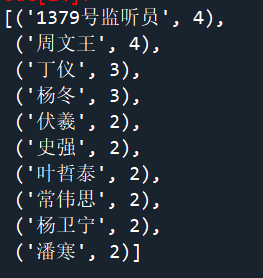
模块度0.0

3.4中心性分析:

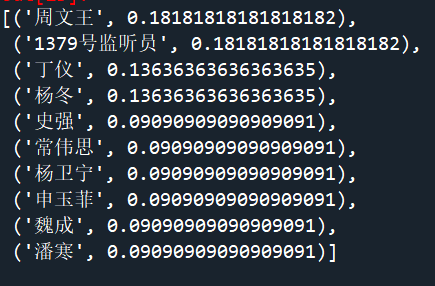
计算各网络度中心性、紧密中心性、介数中心性和特征向量中心性，选取前10名打印

黄金年代：

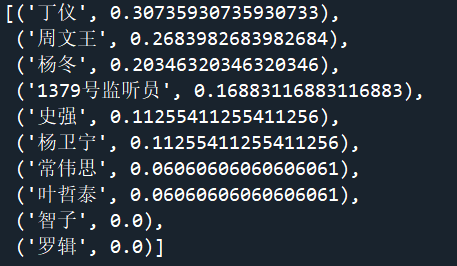
度



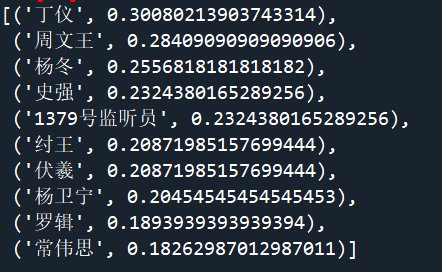
点度中心度



介数中心性



紧密中心性

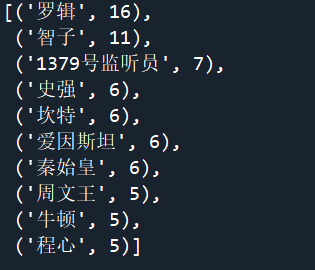


特征向量中心性

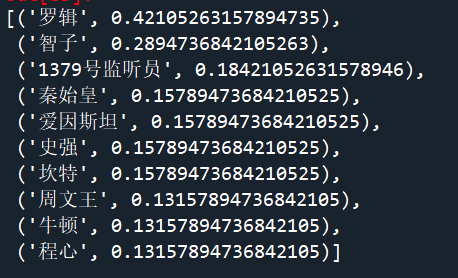


危机纪元：

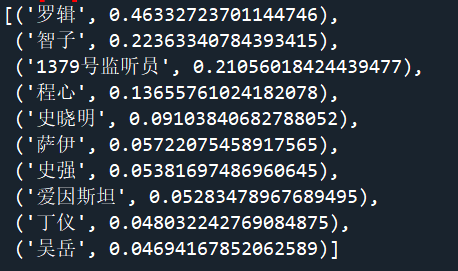
度



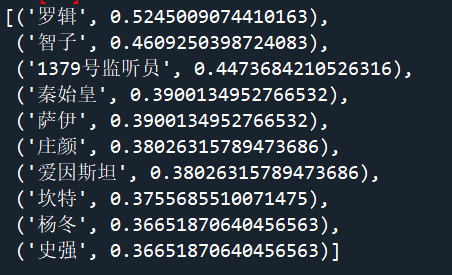
点度中心度



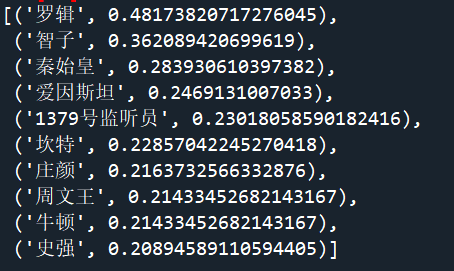
介数中心性



紧密中心性

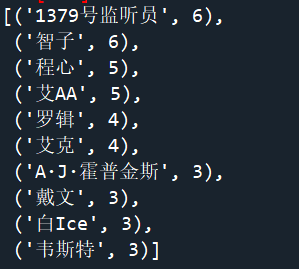


特征向量中心性

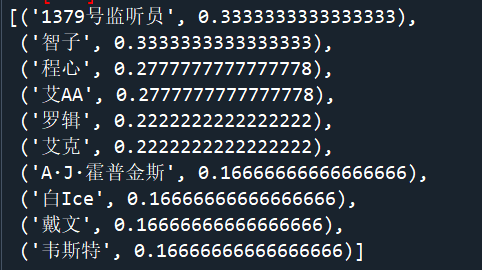


威慑纪元：

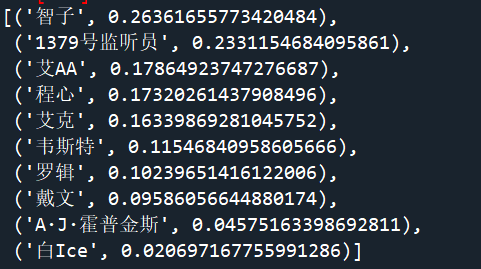
度



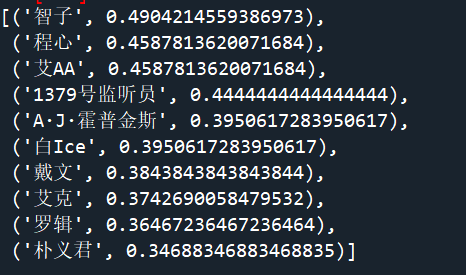
点度中心度



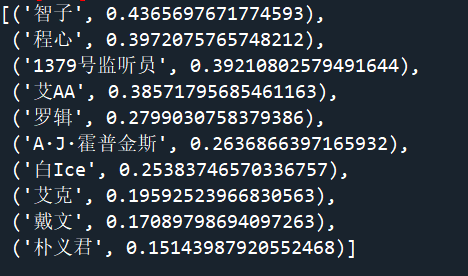
介数中心性



紧密中心性

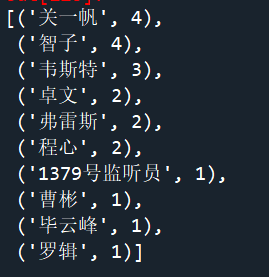


特征向量中心性



威慑后：

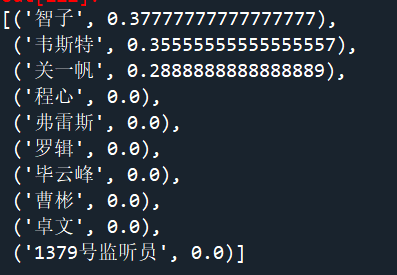
度



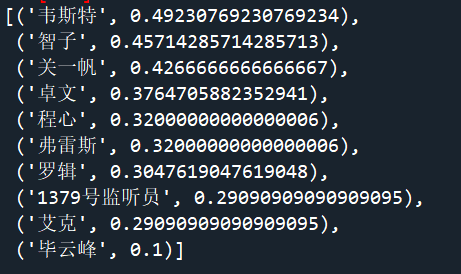
点度中心度



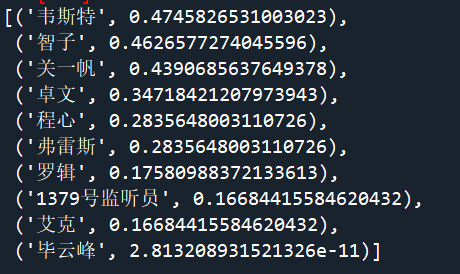
介数中心性



紧密中心性

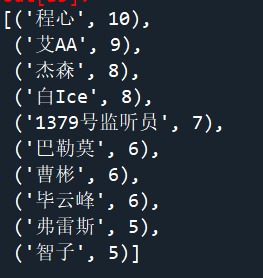


特征向量中心性

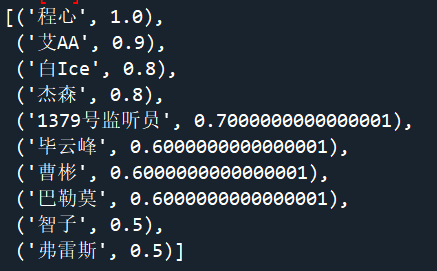


广播纪元：

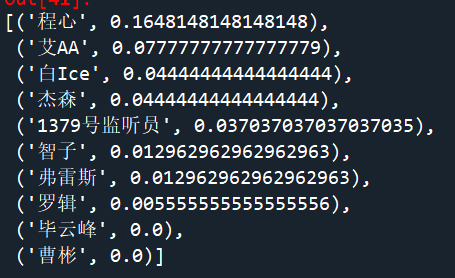
度



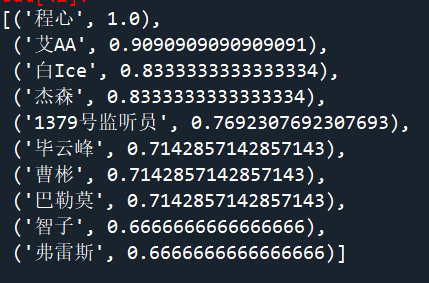
点度中心度



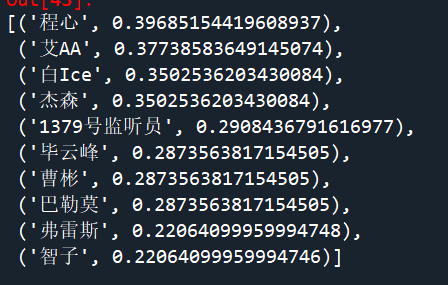
介数中心性



紧密中心性

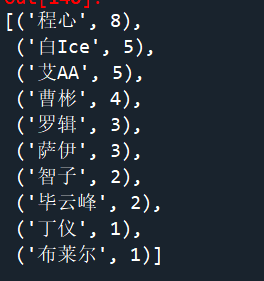


特征向量中心性

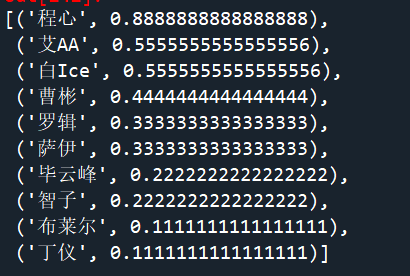


掩体纪元：

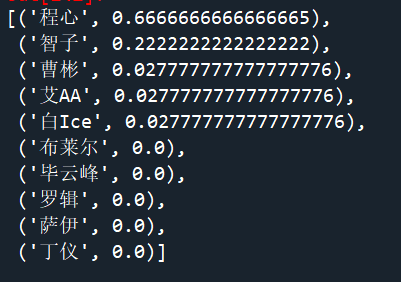
度



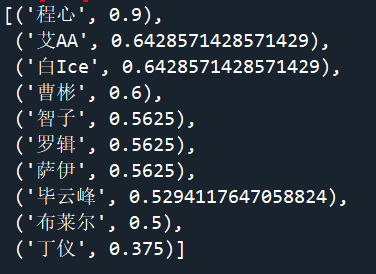
点度中心度



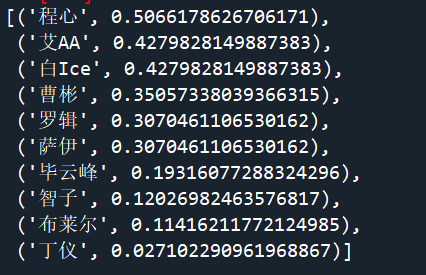
介数中心性



紧密中心性



特征向量中心性

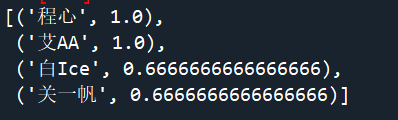


银河纪元：

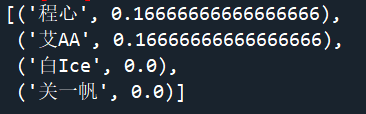
度



点度中心度



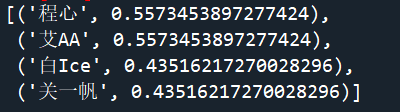
介数中心性



紧密中心性

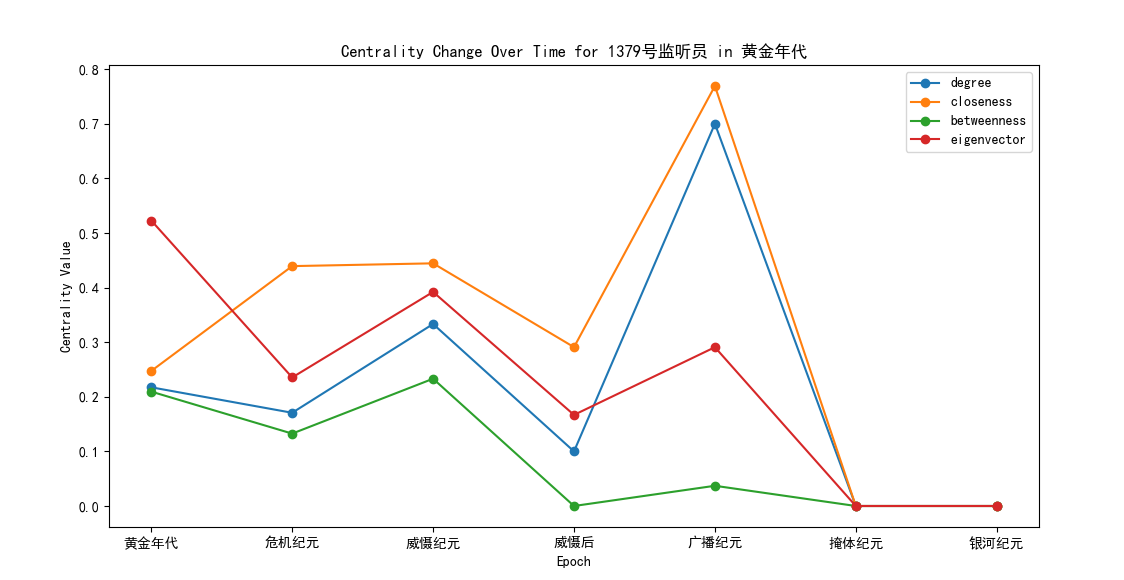


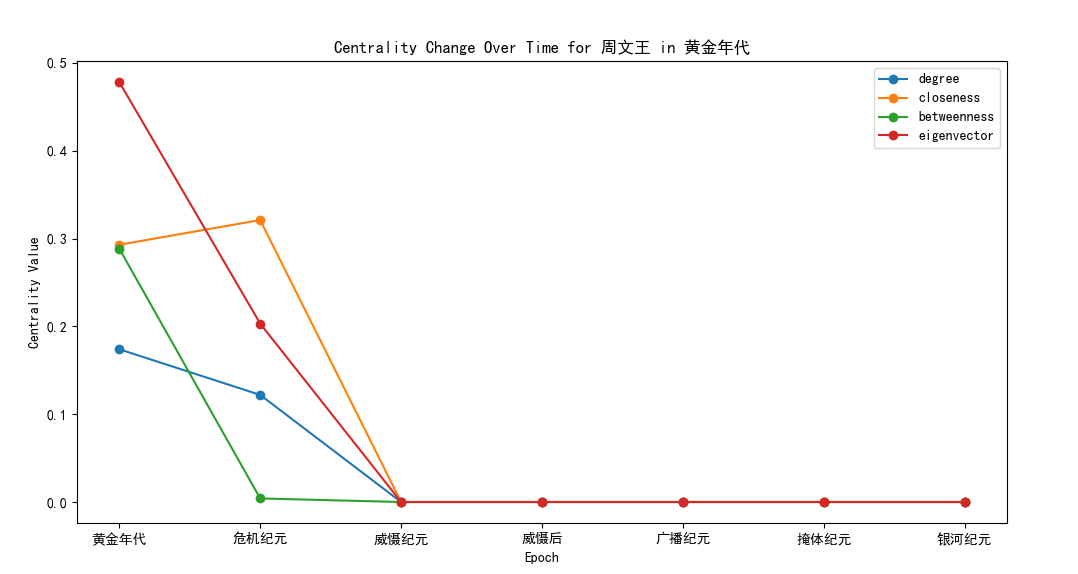
特征向量中心性

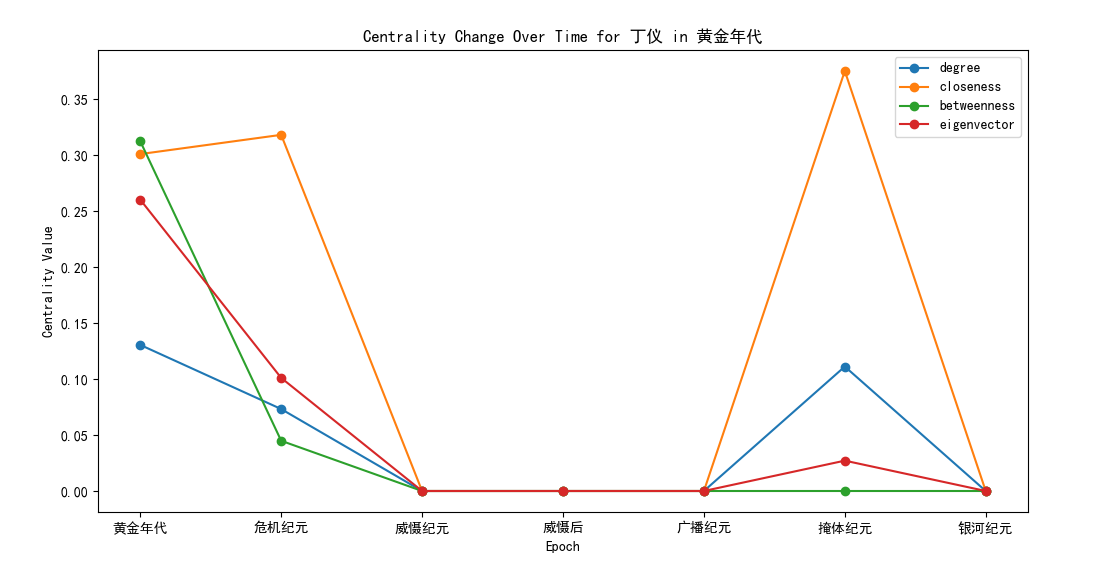


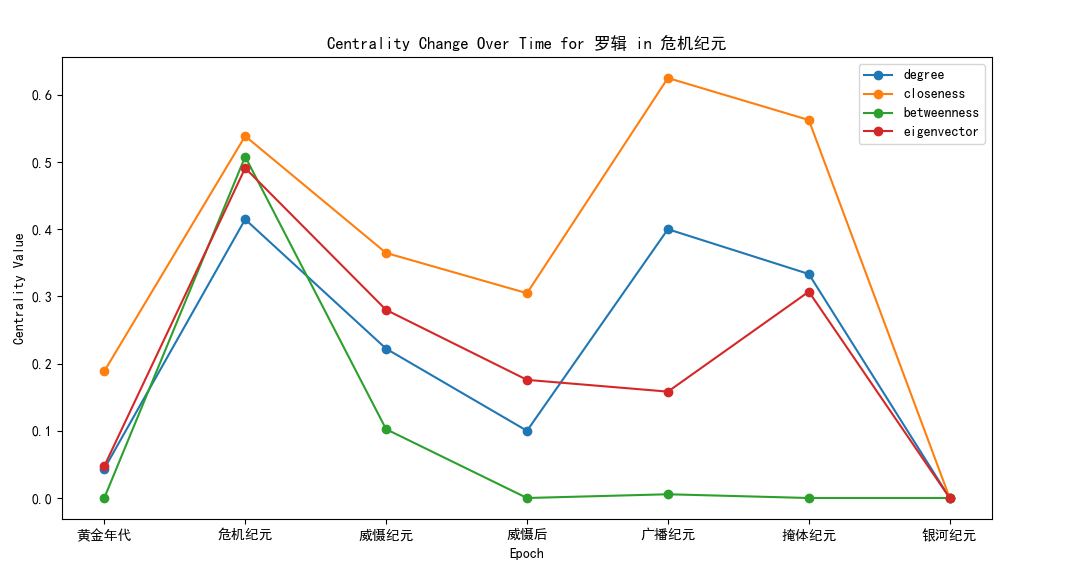
对于各个纪元，以中心性前3名的人作为关键人物，绘制关键人物的中心性变化

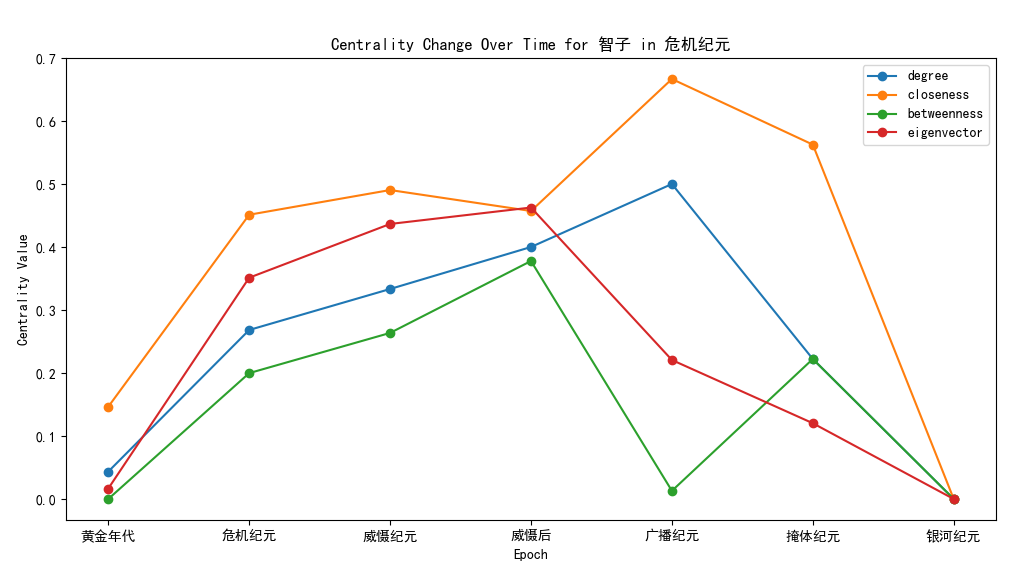
（因为有些纪元关键人物是一样的，其指标变化都是相同的，重复图片便不放出来）

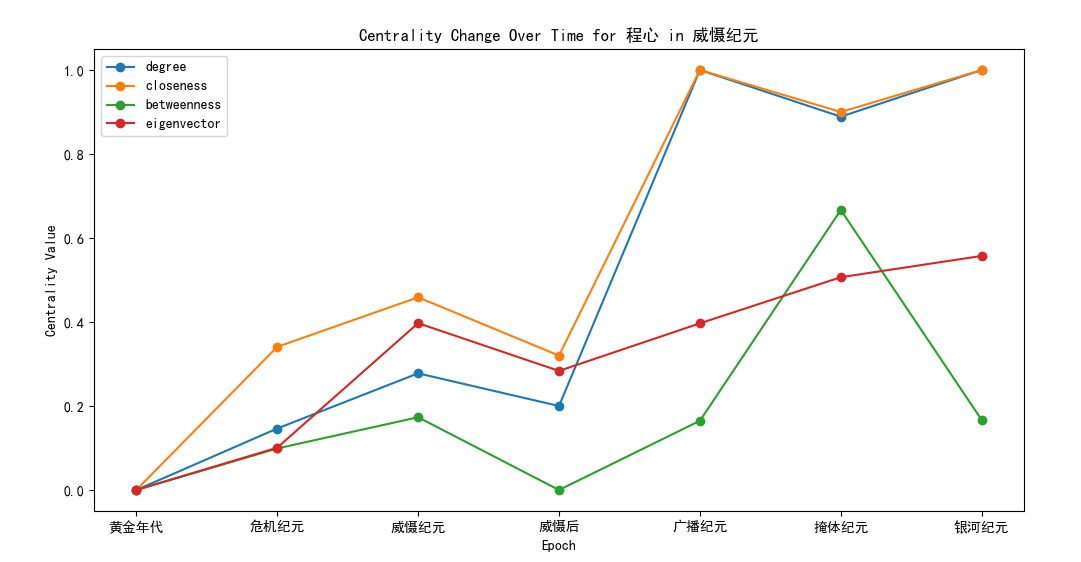


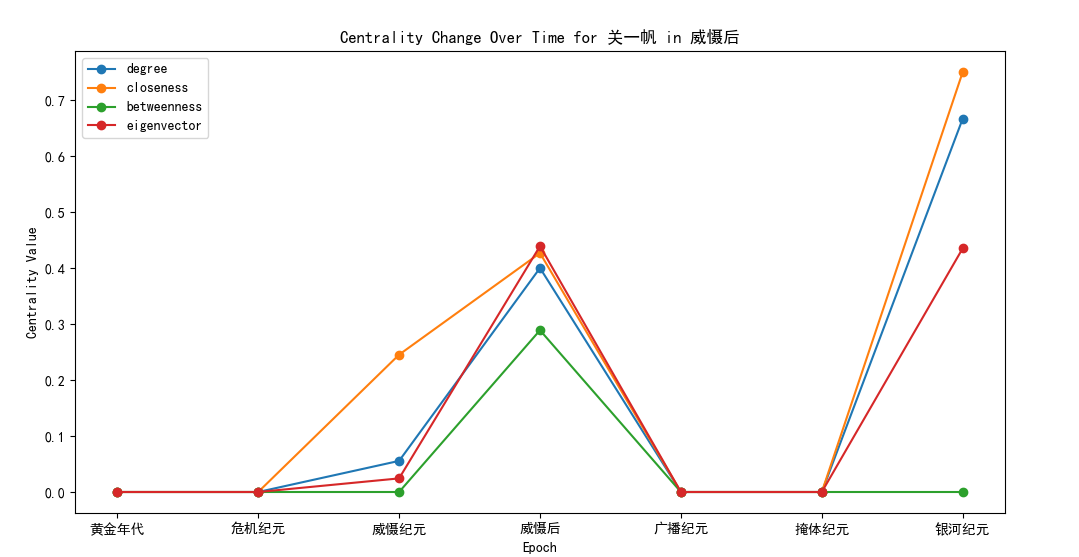


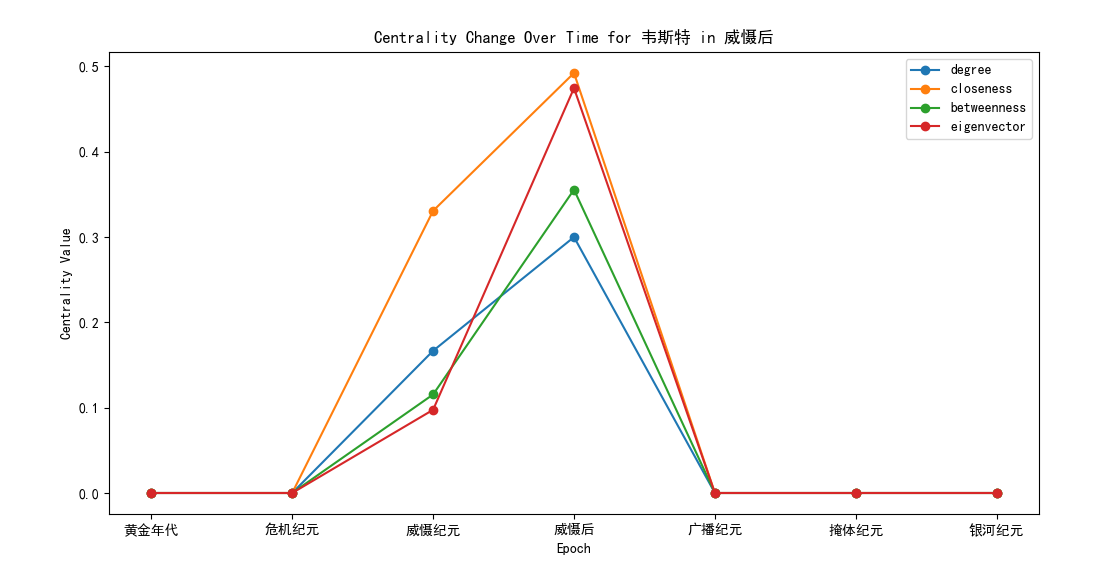


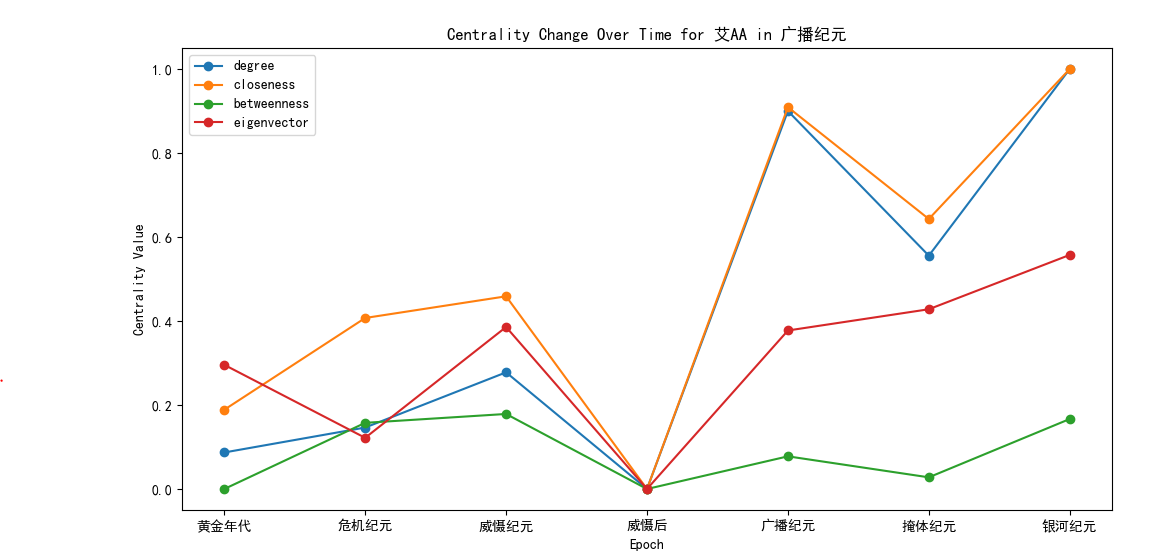


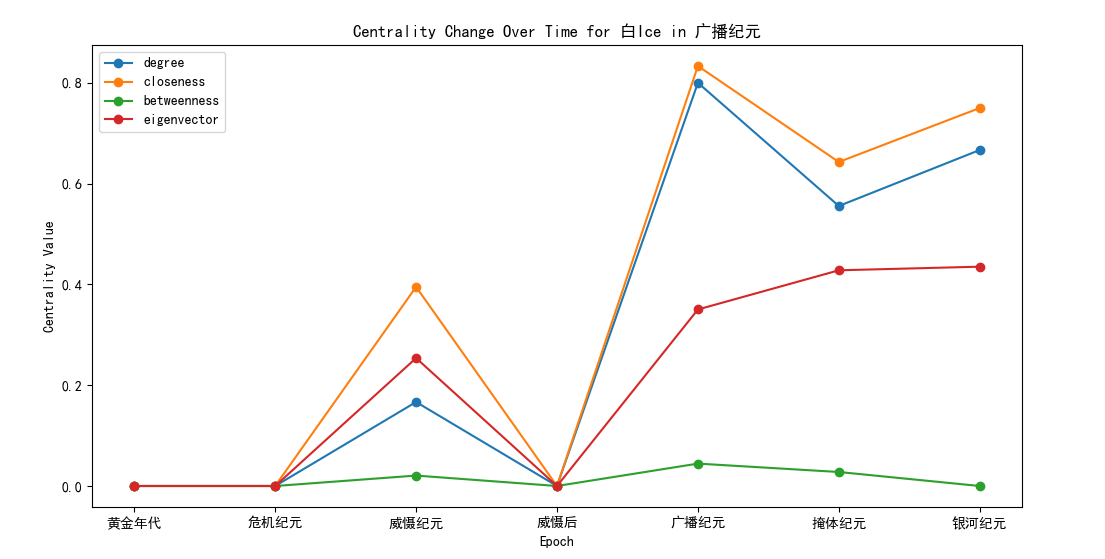






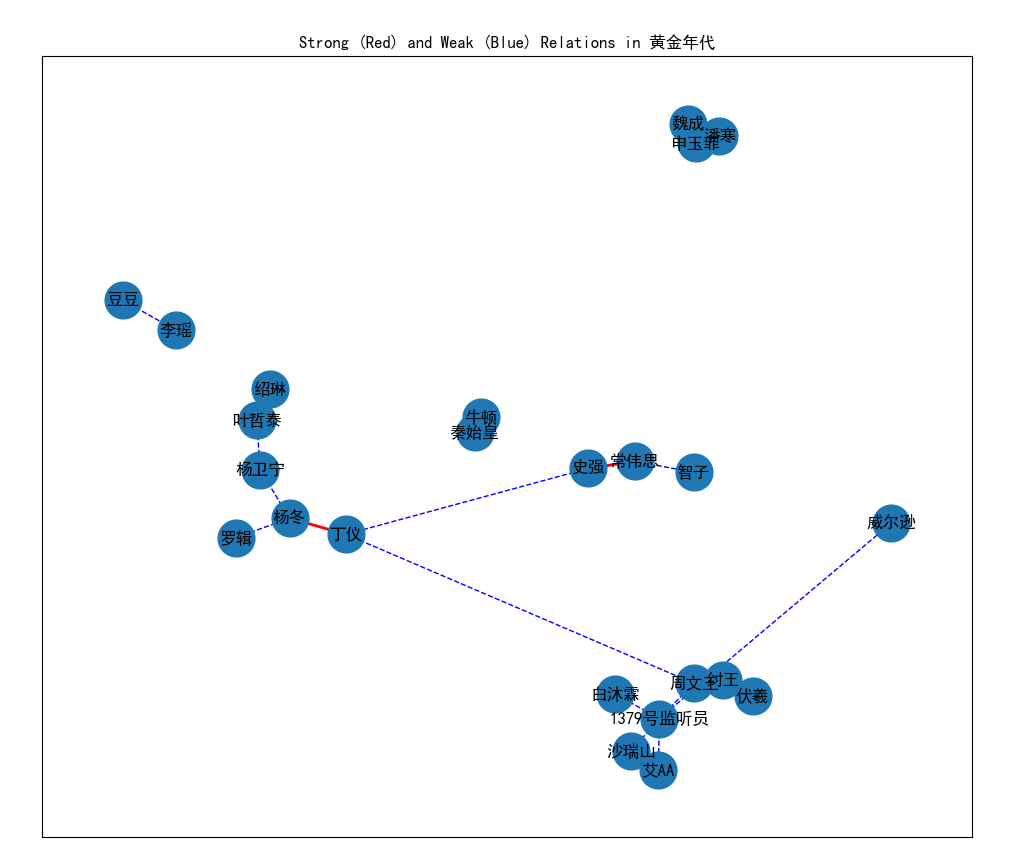


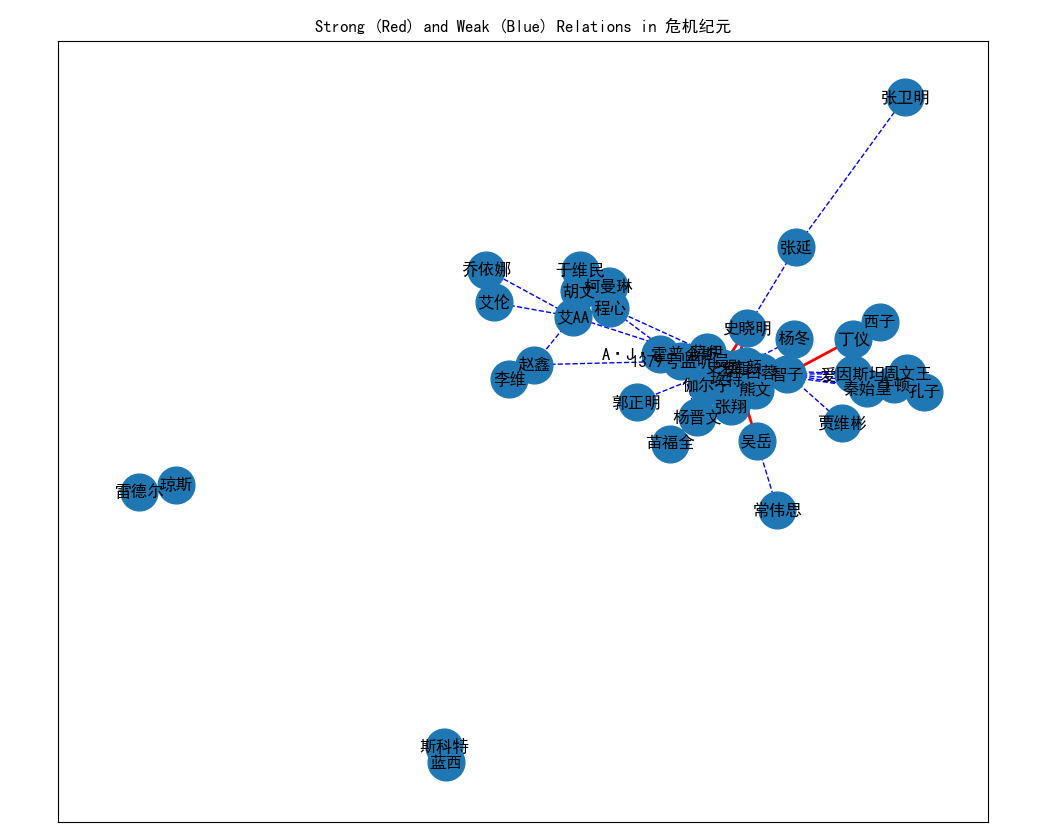


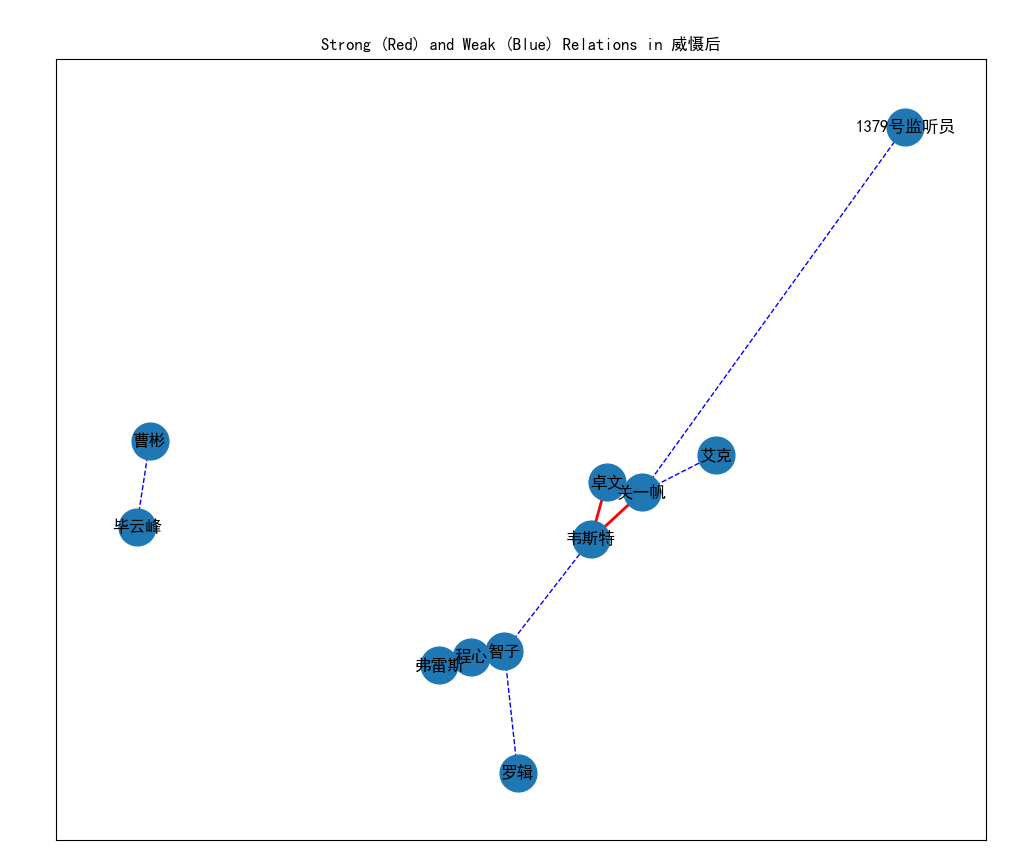


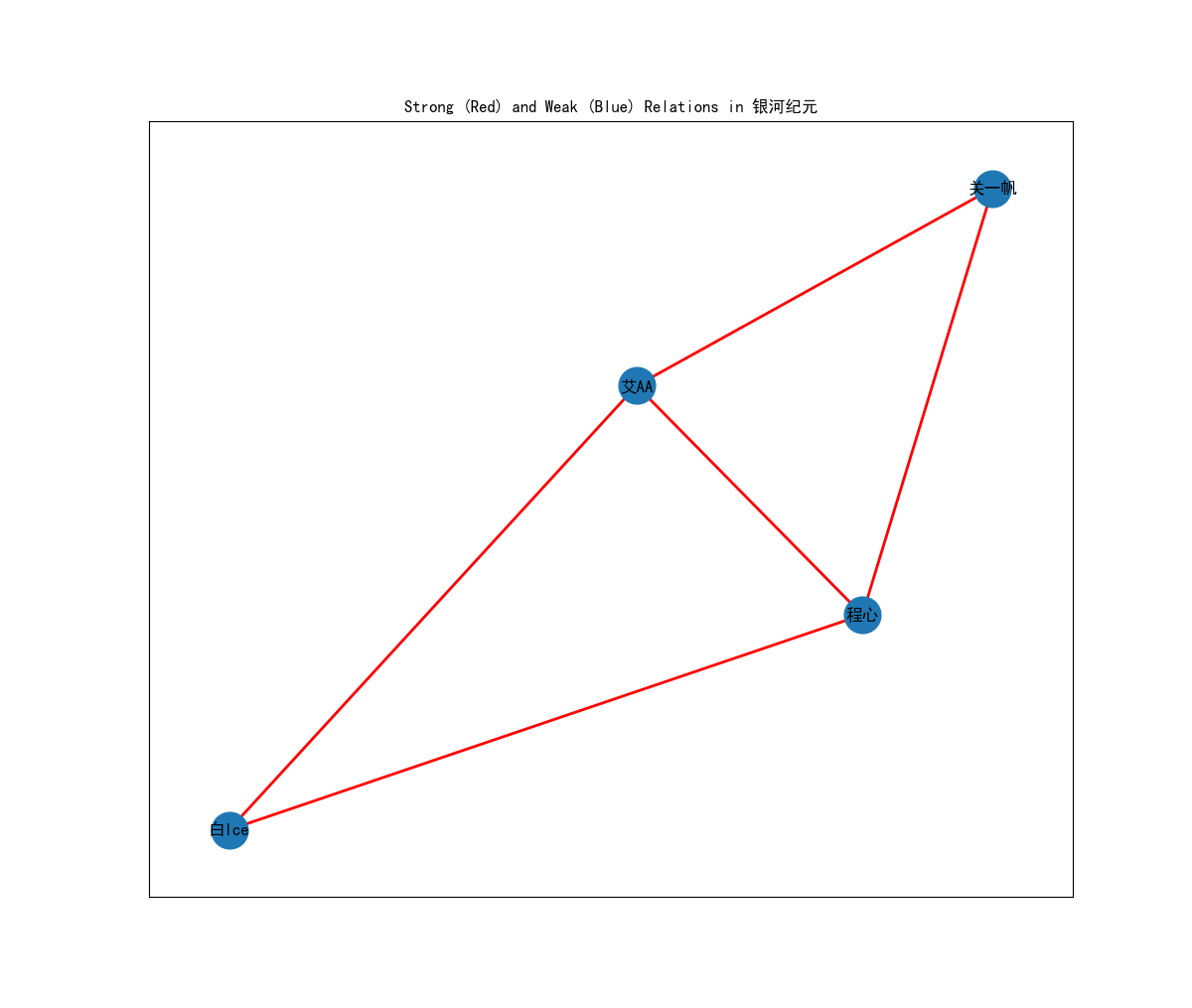
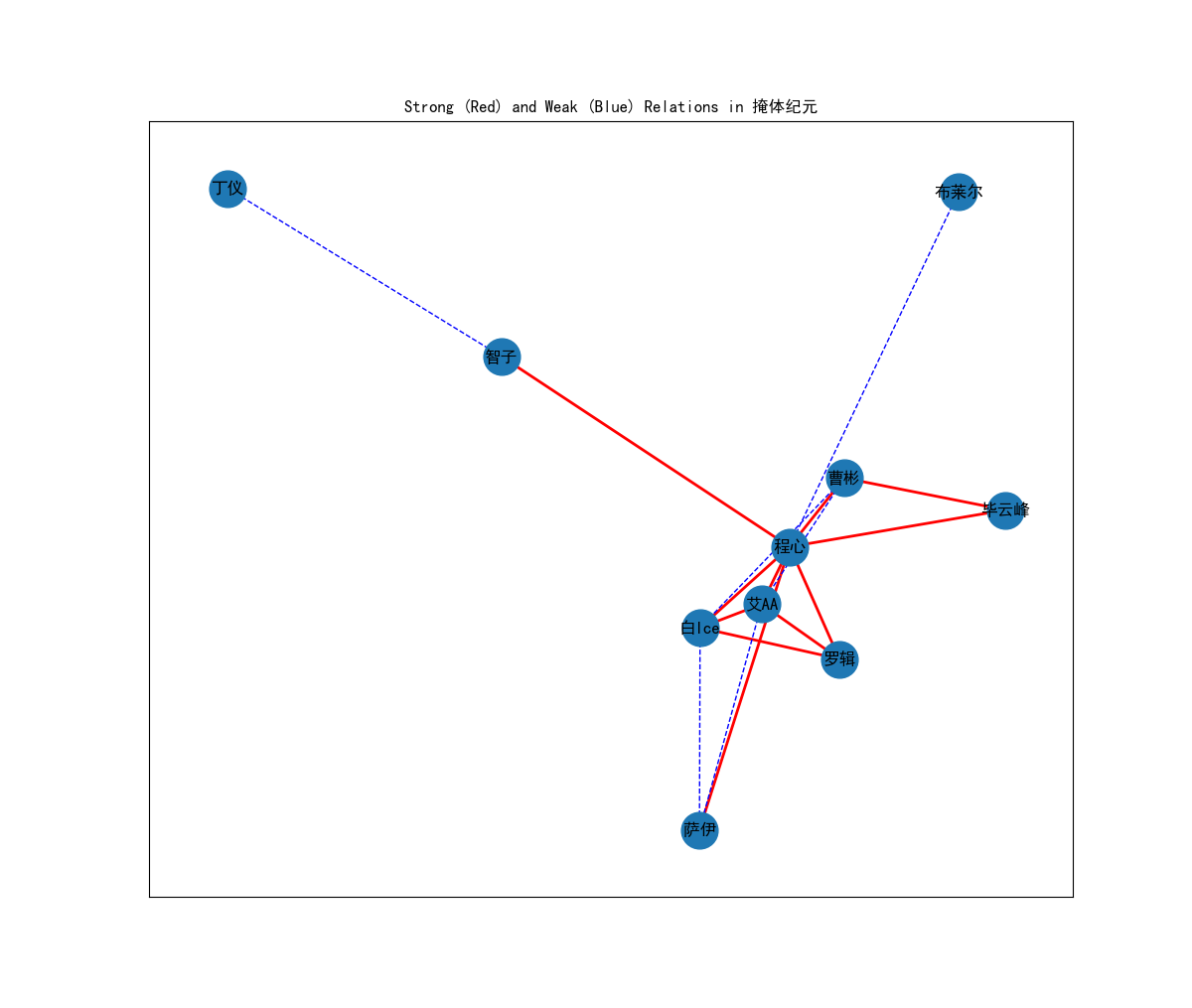
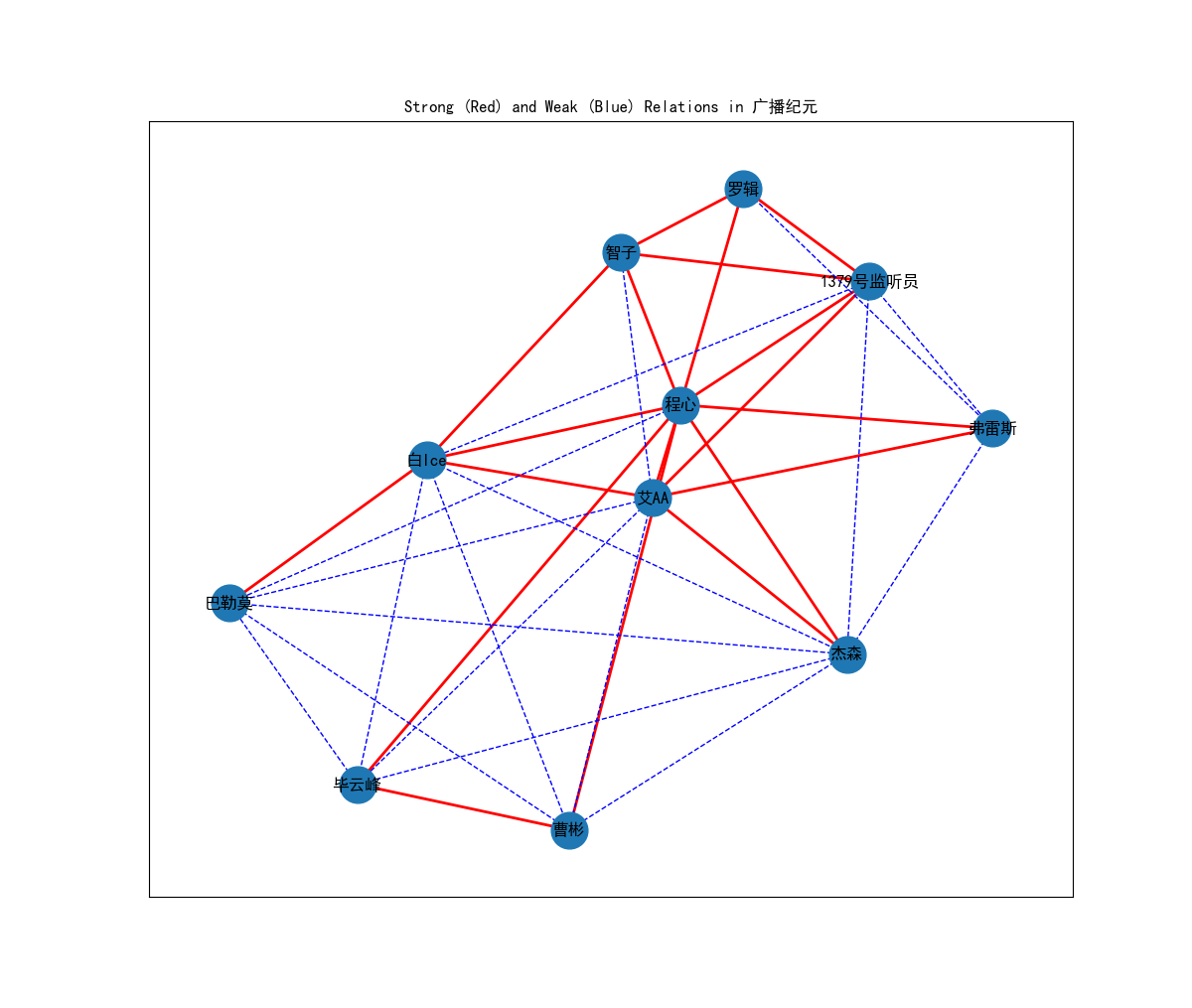
3.5关系强度分析:

分析网络中的强关系和弱关系：



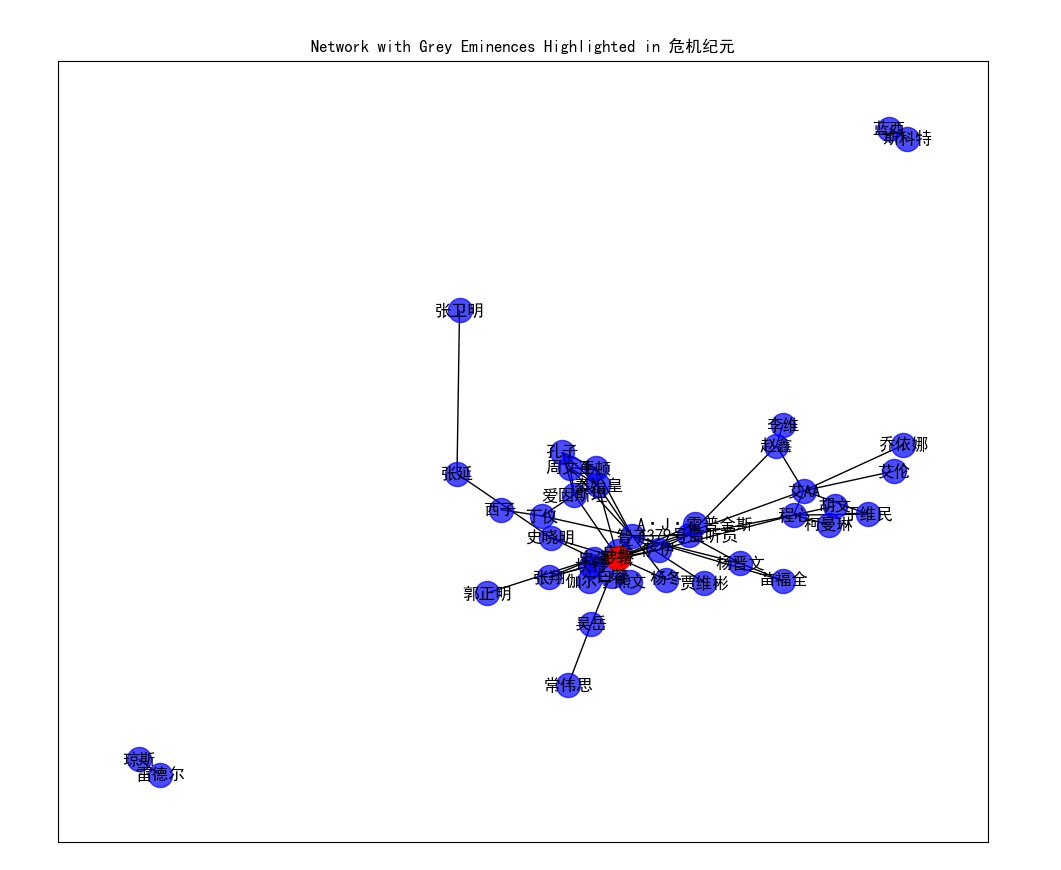
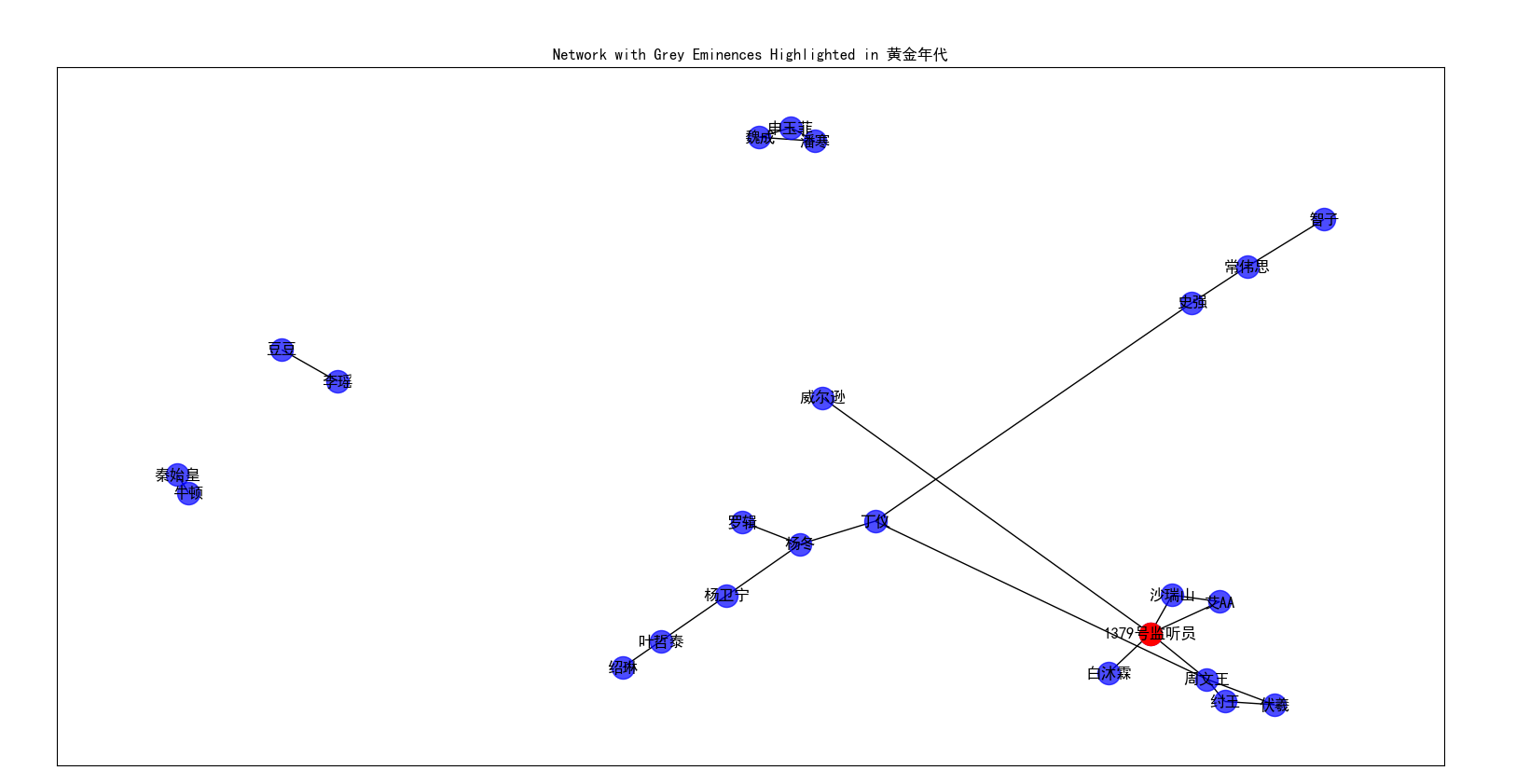


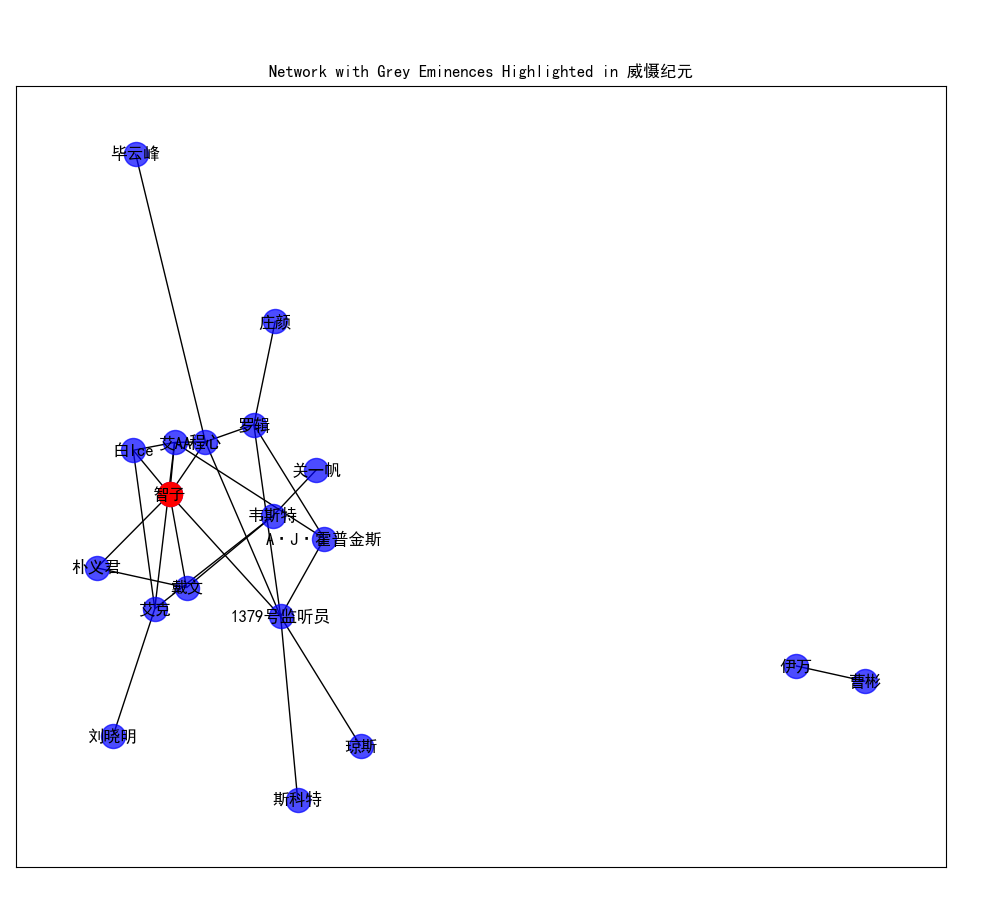


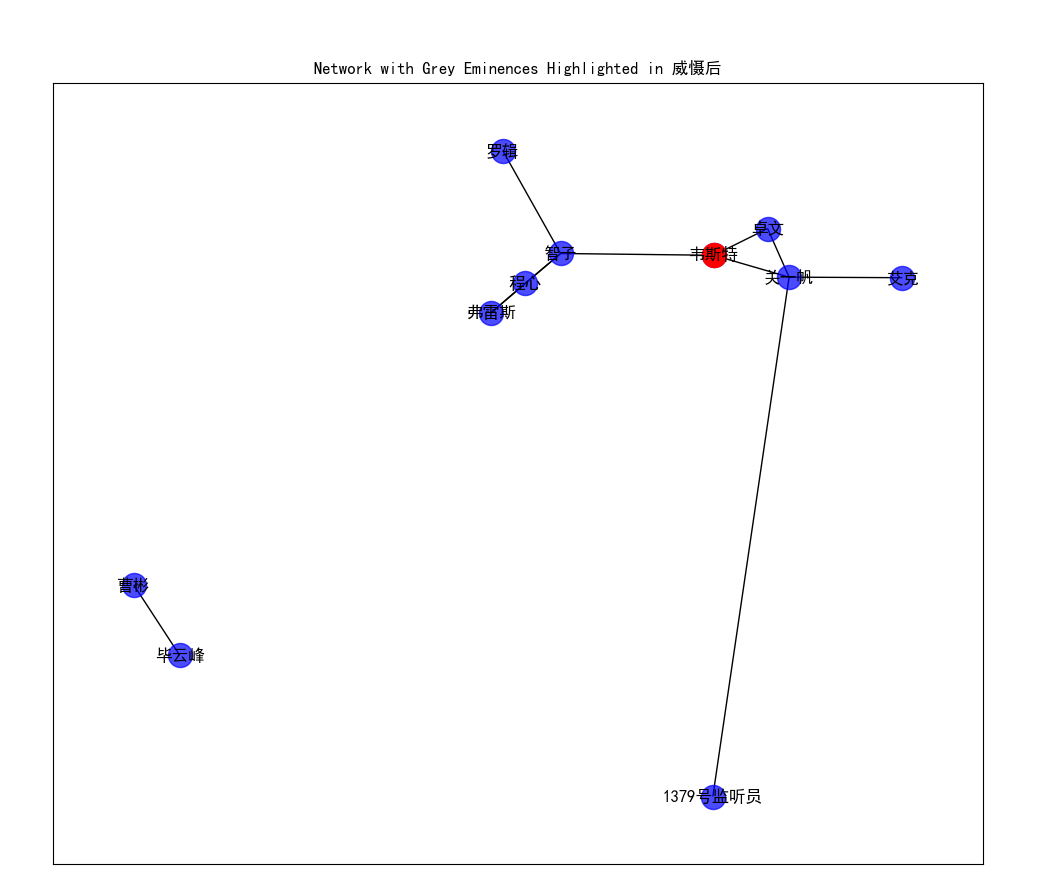


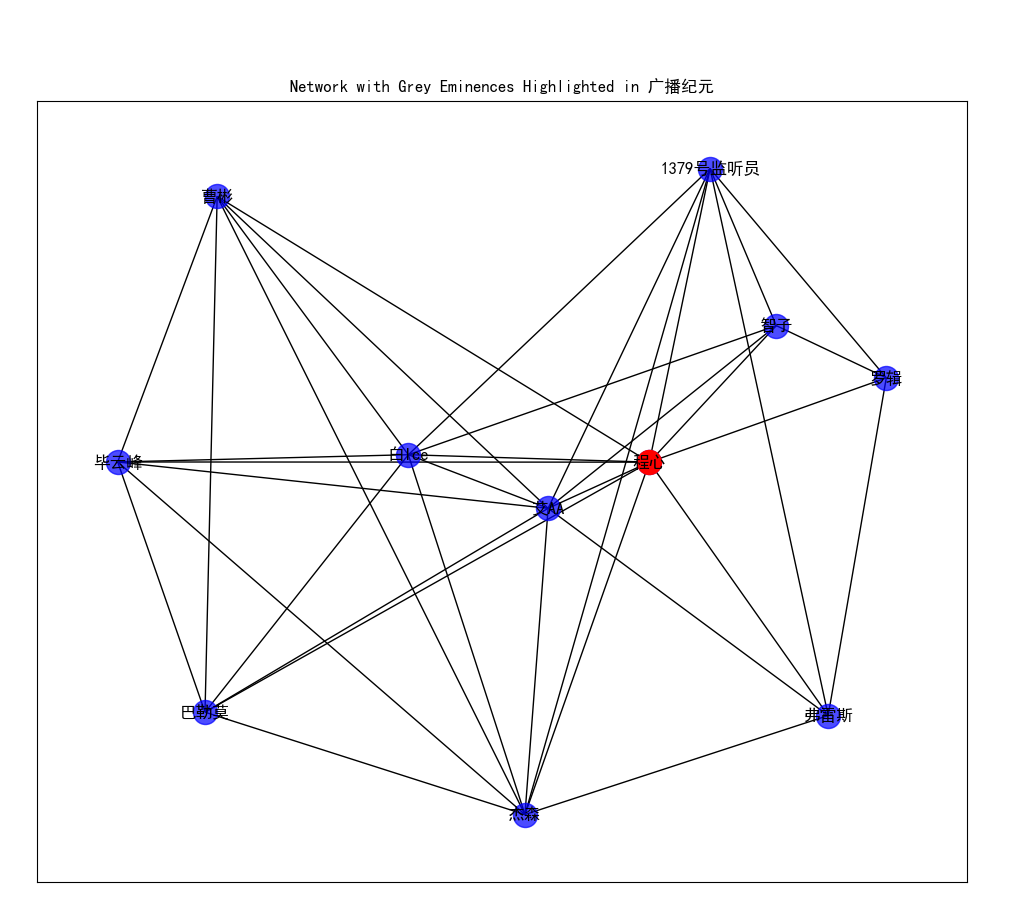
3.6灰衣主教分析

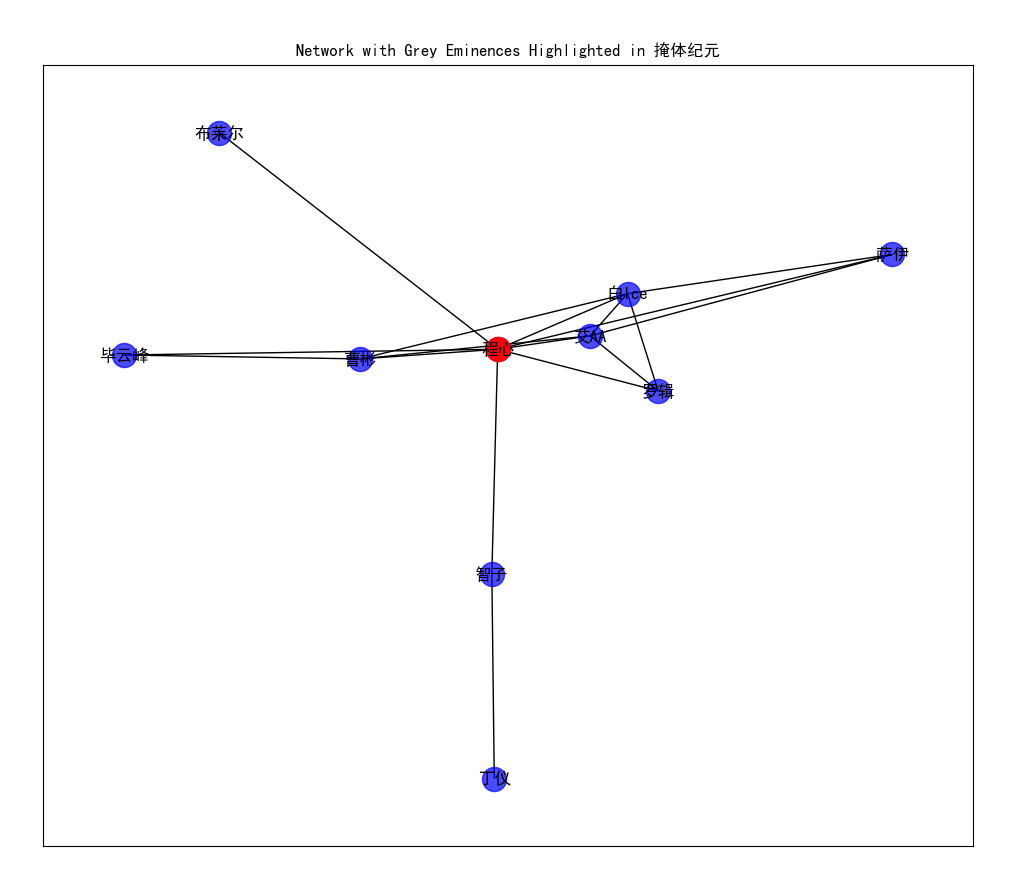
识别每个网络中特征向量中心性最高的节点，确定灰衣主教

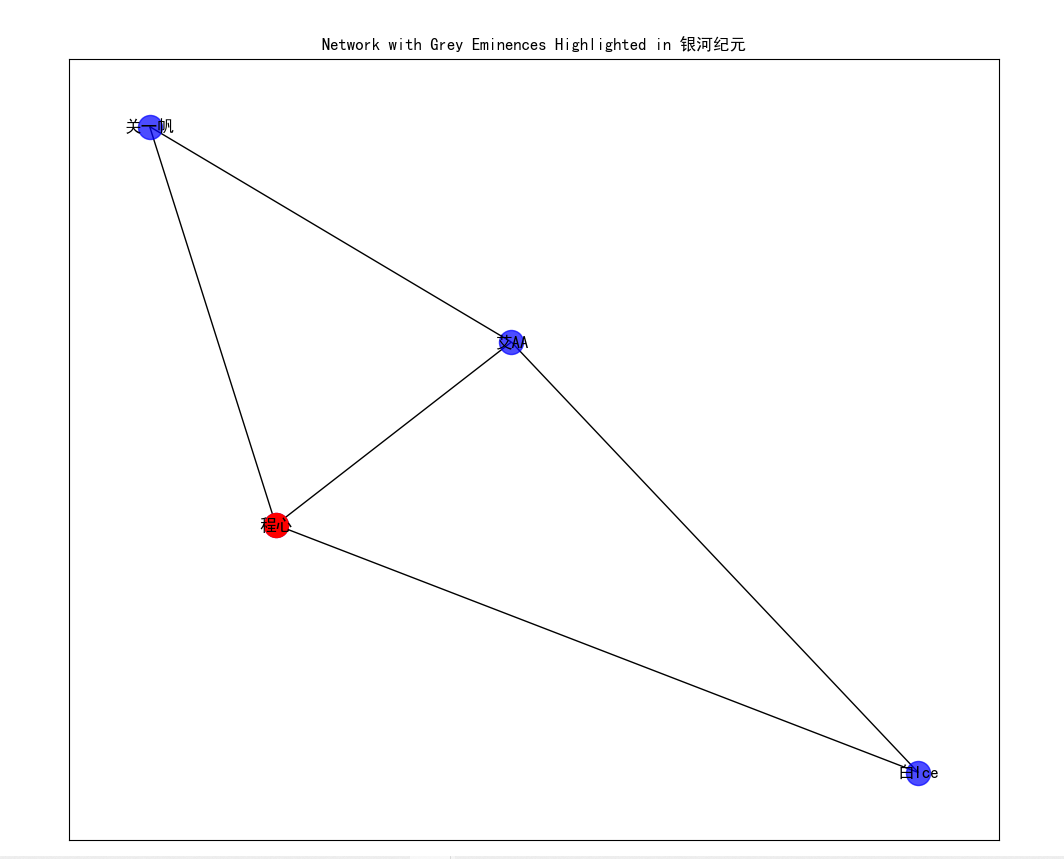












# 四．结果讨论

**4.1.**

从3.2中社会网络指标图表中可以得出：

网络密度增加：

在小说的后期，网络密度显著提高，这可能反映了人类文明在面对外星文明的威胁时，不同国家、组织和个人之间合作增加，形成了更紧密的联盟和合作网络。这可以对应于小说中地球各方势力为应对三体文明的入侵而团结起来的情节。

平均聚类系数的变化：

聚类系数在中间时期的显著提升可能表明在《三体Ⅱ：黑暗森林》中，随着人类逐渐意识到宇宙文明的复杂性，社交网络中出现了更多紧密相连的小团体。这些小团体可能代表着秘密研究组织、政府内部的策略小组，或者是三体问题的不同派系。

平均最短路径长度的增长：

在小说的结尾，平均最短路径长度的增加可能意味着故事中的人物网络变得更为分散。在《三体Ⅲ：死神永生》中，随着战争和宇宙的扩张，关键人物和组织可能被迫分散到不同的地方或星系，导致他们之间的联系变得更为间接和复杂。

**4.2**

通过3.3的社区划分图以及各级元社区结构的变化，以测《三体》三部曲中几个关键事件对社区结构变化的影响：

文化大革命背景下的社区变化：

在文化大革命期间，由于政治动荡，学术和科研机构可能呈现分散状态。不过因为黄金年代是1947-2oo7，看不出来社区结构的大的变化，不过当叶文洁开始向外星文明发送信息，我们可以预见到社区结构的变化。新的节点的引入可能代表了随着三体问题的揭露，新的科学家、政治家和军事人员开始参与到这个论题中。在社交网络图上，这可能表现为新的边和节点的增加，反映了网络中心性的提升。

三体文明的发现：

随着三体文明的发现，原先分散的节点开始凝聚成新的社区，科学界的合作可能加强，政府可能开始介入，社交网络图可能出现新的、更大的社区。这些新社区可能包含前所未有的多元化成员，包括从前未参与直接研究的政府官员和国际代表。

图中的节点增大和社区之间连接数的增加可能代表着新的合作机制的建立，如成立应对三体问题的国际组织，以及在全球范围内进行的秘密科研项目。

宇宙文明危机：

当地球文明意识到更大宇宙文明的暗森林法则和外星威胁时，原有的社交网络可能发生剧烈重组。这可能在图表上体现为原本分散的国家和组织之间的边界模糊，形成一个大规模的、紧密相连的网络，代表了全球范围内为生存而结成的联盟。

关键人物如罗辑在这个阶段可能成为连接不同社区的枢纽，他的角色扩大，可能在图表上表现为一个连接多个社区的大节点。这反映了罗辑在协调全球防御策略中的核心作用，以及他在小说中的转变，从一位普通的教授成长为关键的战略决策者。

威慑体系的建立：

“威慑纪元”开始于人类建立了面对外星文明威胁的威慑体系，这个体系依赖于地球上的一个不可预测的人类个体，即所谓的“剑持者”，他负责在地球受到攻击时启动毁灭性的反击。这种威慑状态下，社区会围绕剑持者形成（程心与罗辑），集中资源和决策力量以维护威慑的有效性。到威慑后，三体人开始攻击地球，大部分威慑纪元的角色没有在出现在威慑后纪元的网络图，程心成为了社区的一个关键角色。

**4.3. 关键人物与故事发展：**

罗辑：

罗辑在小说中经历了由一个普通大学老师到地球文明的关键决策者的转变。逻辑的影响力在危机纪元激增，说明在三体文明首次接触地球后有所上升，那么这反映了他在科学界和地球防卫计划中的中心地位的增强。罗辑的行动，如启动威慑纪元，直接影响了地球文明的生存战略。

程心：

程心在小说的后期成为重要角色，特别是在宇宙封锁之后。她的接近中心性在威慑后逐渐增加，这可能意味着她在人类社会中变得更为关键，与多个角色都保持着联系，这反映在她对人类未来命运的影响上。

智子：

智子作为三体文明的代表，如果其中介中心性在小说中有显著变化，这可能反映了她在不同势力或社区间的调解和影响力。

艾AA：

艾AA在小说中担任基础科学研究的角色。如果她的中心性指标在广播纪元的关键时刻变化，这可能反映了她的科研成果对整个人类文明产生了重要影响。

丁仪：

作为科学家和红岸基地的关键人物，丁仪在小说早期有重要作用。他的中心性变化可能与红岸基地的活动及其与三体文明的互动密切相关。

白Ice：

如果白Ice的中心性在广播纪元激增，这意味着她在小说中的角色发生了变化，从配角变成了主角团的一员

**4.4 关系强度与人物互动：**

人物间的强关系影响：

罗辑和叶文洁的关系对于开启整个系列的故事具有决定性作用。叶文洁作为“红岸项目”的关键科学家，她的选择向外宇宙发送信息，直接导致了三体文明与地球的首次接触。

在“三体”世界中，程心与艾AA的关系变得非常重要。程心继承了罗辑的意志，而艾AA的行动和决策在许多情况下都对程心产生了重大影响。

丁仪与其他科学家之间的紧密合作在探索和解决三体问题中发挥了核心作用，特别是在处理与三体文明的复杂互动时。

罗辑和智子之间的强关系起着至关重要的桥梁作用，智子作为三体文明的代表和使者，她与罗辑的互动为读者提供了关于三体世界以及其技术和文化的深入见解。他们之间的对话和冲突成为了理解两个文明冲突的关键。

弱关系网络的作用：

弱关系通常链接着社交网络中的不同群体，它们在传播新信息、构建社交桥梁以及引入社交网络外的新观点中发挥作用。在《三体》中，科学界、政治圈和军事部门之间的弱联系有助于跨领域的信息传递和策略制定。

例如，智子作为三体文明的代表，与地球文明中的多个角色建立了弱联系。这些联系虽然不如其他联系紧密，但是它们在两个文明之间的信息交流和误解的产生中起到了关键的桥梁作用。

1379监听员与很多角色间的关系则展示了弱关系网络如何在关键时刻成为信息传递的渠道。1037作为三体世界的监听员，她的身份和行动虽然在初期不为人知，但她与罗辑的最终合作揭示了三体文明内部的分歧，并对故事的结局产生了重大影响。

**4.5 灰衣主教角色的深度分析：**

**罗辑**在整个三部曲中经历了巨大的转变，从一个大学老师到成为掌握地球命运的关键人物。他的策略性行为，尤其是在威慑纪元中的角色，影响了地球对三体世界的反应。罗辑建立了与三体文明对话的联盟，控制了关键信息的发布，这些决策导致了地球文明在三体问题上的统一立场和行动。

**程心**在故事后期成为推动事件的关键人物。她继承了罗辑的遗志，建立了新的联盟，并掌握了决策的信息。她的策略性行为，尤其是对外星文明的战略规划和内部团队的管理，对人类在三体问题上的决策产生了深远影响。

**1379监听员**可能是三体文明中的一个内部人物，其角色在故事中虽然较为隐秘，但她掌握的信息和行动对于故事的发展至关重要。她的信息控制和潜在的联盟构建暗中影响了三体文明与地球文明之间的关系。

**韦斯特**是“万有引力号”战舰心理医生,负责战舰上战士们的心理健康问题。作为战舰上的心理医生，韦斯特负责处理人员的心理健康，这一职责让他与战舰上的每个人都有所接触，能够深入了解他们的想法和情绪。这样的位置赋予了他间接的影响力，可以通过心理辅导和干预来影响宇航员的决策和行动。尽管他在中可能不是最显眼的角色，他的工作和对人员心理状态的维护却是支撑“万有引力号”完成其使命的基石。他在背后的工作保证了战舰能在恶劣环境中维持运作，从而在整个故事中发挥了重要的支持作用。

# 五．总结

《三体》三部曲不仅是一部科幻巨著，它还通过细致的社会网络构建，展示了人类文明在宇宙舞台上的挑战、斗争和演化。通过本报告的分析，我们看到了人物之间的关系如何在宏大的宇宙叙事中发挥着举足轻重的作用，以及他们如何在不断变化的环境中寻求最佳的生存策略。《三体》通过其复杂的人物关系网，不仅讲述了一个关于星际冲突的故事，也深刻地揭示了人性的多面性和人类社会的复杂性。

除去报告中展现的几种社会网络分析算法，我们还做了其他计算和可视化分析，因为内容太多，无法在报告中体现，所有的可视化结果放在了名为“可视化”的文件夹中。

# 六．小组成员和分工

**张宇峰：**

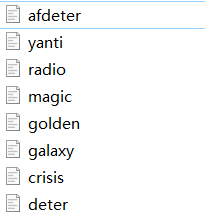
负责数据的搜集，处理，清洗，完成了所有数据处理工作

代码部分：

**姑丽米热·买明:**

负责社会网络算法分析可视化第一部分（度；点度中心度；介数中心性；紧密中心性；特征向量中心性；Pagerank算法；公共邻居；Jaccard系数；资源分配指数；Adamic-adar指数；优先链接；社交网络密度；连通分量；聚集系数；社区发GN算法模块度贪婪模块度算法Louvain算法标签传播法）

代码部分：



**阿依孜巴·阿力木江：**

负责社会网络算法分析可视化第二部分（K-Means 聚类划分社区；中心性计算社交网络图绘制；强弱关系分析：共同朋友分析；链接预测；三角关系分析；灰衣主教识别教；图表保存）

负责报告攥写（问题分析，建模分析，结果讨论）

代码部分：