

# 社会网络分析方法在金庸小说 人物关系中的应用研究\*

姚睿琦 张 辉 姚云洪

(山东大学管理学院, 济南 250100)

**摘 要:** [目的/意义] 探讨运用社会网络分析方法研究小说中的人物影响力和关系变化特征, 分析小说的写作意图与写作特点。[方法/过程] 以金庸小说《射雕英雄传》和《神雕侠侣》为研究对象, 基于中文信息处理与社会网络分析法, 抽取小说中的人物关系网络, 通过社区探测、k-核分解、中心性分析研究小说人物的重要程度, 通过时间演化和人物关系网络结构测度分析人物影响力的演变。[结果/结论] 人物关系网络具有“小世界效应”, 人物影响力与亲密关系的变化规律显著, 能够揭示作者的写作风格。

**关键词:** 小说社会网络 人物关系 中文信息处理 演化分析

**分类号:** TP391

**DOI:** 10.31193/SSAP.J.ISSN.2096-6695.2021.03.06

## 0 引言

自然语言是人类传递信息、获取信息的重要手段, 随着信息时代的到来, 自然语言处理 (Natural Language Processing, 简称 NLP) 应运而生, 它将自然语言转化为机器所能理解的机器语言, 以帮助机器理解内容和提取有意义的信息<sup>[1]</sup>。中文信息处理 (Chinese Information Processing, 简称 CIP) 是针对中国的语言文字开展相关研究的一个专属领域, 是自然语言处理的一个具体分支<sup>[2]</sup>。

小说作为自然语言的表达形式, 是以描述人物为中心、反映社会生活的文学体裁。人物作为小说三要素之一, 识别人物的重要性、剖析人物关系是理解小说情节和背景的关键。主角具有揭示小说主题、突出中心的作用, 而重要配角对情节和主角的成长往往有推动作用。传统研究小说

\* 本文系教育部人文社科青年基金项目“社会化链接信息共享行为机制及条件组合研究: 基于关系强度视角”(项目编号: 19YJC870019)的研究成果之一。

**[作者简介]** 姚睿琦 (ORCID: 0000-0001-8754-143X), 女, 硕士研究生, 研究方向为中文信息处理、数据挖掘, Email: sd3yyh@163.com; 张辉, 女, 副教授, 研究方向为信息组织与检索、知识管理与创新, Email: huizh@sdu.edu.cn; 姚云洪, 男, 副教授, 研究方向为数据挖掘、信息管理, Email: yaoyh@sdu.edu.cn。



姚睿琦, 张辉, 姚云洪. 社会网络分析方法在金庸小说人物关系中的应用研究 [J]. 文献与数据学报, 2021, 3(3): 068-080.

的方法往往耗时较长, 依赖于主观地解读小说内容、理解作者意图, 导致了“一千个人眼中就有一千个哈姆雷特”。随着信息技术的发展, 定性地分析小说内容已经无法满足小说研究的需要, 因此, 近年来利用计算机技术对小说文本进行定量研究, 以弥补定性分析的不足, 吸引了许多学者的关注。

本文将中文信息处理与社会网络分析法应用于分析金庸武侠小说人物, 以《射雕英雄传》和《神雕侠侣》(简称“射雕二部”)作为语料训练对象。首先, 将小说文本分章回存储, 建立人物词典和指代消解词典; 其次, 按照段落抽取人物关系, 构建“射雕二部”关系网络; 然后, 用社区探测分析、k-核分解和中心性分析, 识别小说各人物的重要程度; 最后, 将小说按照时间序列绘制人物关系网络, 计算各阶段网络结构与人物特征向量中心性, 从而达到研究人物影响力和关系变化特征、揭示小说写作特点的目的, 不但可以推动中文信息处理与文学研究的融合, 还能促进小说的个性化推荐。

## 1 相关研究概述

目前, 国内已开展了许多用中文信息处理分析小说文本的研究, 涌现了一批优秀研究成果, 例如, 通过研究语料库的语言结构特征来判定作者<sup>[3]</sup>, 通过统计词频和聚类分析辨析小说的风格<sup>[4]</sup>。近年来, 还出现了利用社会网络法对小说人物分析的研究。社会网络分析是对社会关系结构和属性加以分析的一套规范和方法, 其意义在于可以对多种关系进行量化分析, 从而为某种理论或者框架的构建和实证命题检验提供量化工具, 甚至可以建立宏观与微观之间的桥梁。将社会网络法用于分析小说文本, 不仅可以发现网络特征, 还能挖掘人物关系, 例如, 刘海燕等提取了菲茨杰拉德的四部小说人物角色, 分析人物关系网络的结构性质<sup>[5]</sup>; 赵京胜等学者运用节点的度、中心性和凝聚子群, 分析《三国演义》的角色分布特性<sup>[6]</sup>; 王一博等结合战略坐标图与核心边缘结构对聚类结果进行阐述<sup>[7]</sup>; 钮亮以《水浒传》中 108 位梁山好汉为研究对象建立社会网络, 揭示人物排名和人际关系形成的原因<sup>[8]</sup>; 杨鑫等以金庸小说《倚天屠龙记》为分析对象, 实现了对小说人物的共词分析、聚类分析和情节演化分析<sup>[9]</sup>; 胡云鹏以《三国志》为研究载体, 结合三国时期专用的文言文词典, 深度挖掘各人物关系<sup>[10]</sup>。魏会洋、袁曦临分析了《白鹿原》中的人物关系结构, 对社会网络分析法在文学阅读研究中的适用性问题加以探讨<sup>[11]</sup>。

金庸武侠小说被誉为武侠小说中的“殿堂级作品”, 作者查良镛用浪漫的叙事、丰富的剧情为读者树立了鲜活饱满的武侠人物形象, 创造了独一无二的武侠世界, 深受广大读者喜爱。《射雕英雄传》和《神雕侠侣》作为“射雕三部曲”中的前两部, 是金庸武侠小说中吸引读者数量最多的作品, 自问世以来被多次翻拍为影视作品, 在中文武侠小说中具有代表性。“射雕二部”具有时间序列特征<sup>[12]</sup>, 其复杂的人物关系、繁多的人物别称、文字规模庞大的章回结构等造成了比单本小说更大的噪音, 增加了利用计算机提取有价值的信息、过滤掉冗余的难度。

虽然上述研究取得了一定的成果, 但是大多数局限于以小说主角为引导的关系研究, 没有将配角的重要性纳入考虑, 其研究重点在于设计小说人物关系的宏观分析框架。然而, 随着情节的发展, 小说人物关系并不是一成不变的。按照时间序列进行人物关系网络演化, 并将人物影响力



量化，不仅能在总体上把握人物关系，更能动态地揭示人物的影响力和亲密关系的变化，达到将宏观小说人物研究细化的目的，为文学作品的定量分析提供参考。

## 2 人物关系网络的构建

### 2.1 数据采集与处理

本文数据源自金庸网，金庸网收录了全部版本的金庸武侠小说，构建了较为全面的金庸小说数据库，被众多金庸小说爱好者所认可。金庸武侠小说均经过两次改版，分为旧版、新版、新修版。由于新修版做出了更合理的删改，为保证数据的严谨性与一致性，本文均采用小说的新修版。

首先，用爬虫软件提取金庸网中《射雕英雄传》与《神雕侠侣》新修版的全部内容以及数据库中小说的人物资料，包括人物别称和经历，并下载百度百科中的小说人物列表。其次，由于很多小说都没有成熟的语料库，而且人物拥有丰富的别称，如果不校正多个版本的人物名单再建立词典，不统计人物的别称对其指代消解，就会削弱人物识别的精度，增加人物深层分析的困难性；因此，本文构建了“射雕二部”语料库，包括将文本格式转换、小说分章回存取、构建人物词典与人物别称映射的指代消解词典。然后，用 python 遍历金庸网下载的人物数据与百度百科人物列表进行校对，检查二者的偏差，再根据需要进行增添与删减，如按照人物重要性删去“江南七怪”集合代词，保留七人名字，删去如“店伴”等在两本小说中指代不明确的词，完成共计 181 人物词典。接下来，结合金庸网的人物数据库和小说内容构建指代消解词典，串联人物姓名与别称，如：黄蓉，郭靖、洪七公等人称之为“蓉儿”，后期被众英雄称为“黄帮主”；完颜洪烈前期化名“颜烈”等。最后，用 python 对“射雕二部”文本 jieba 分词、增加自定义人物词典与过滤停用词，提取分词结果，判断分词结果是否在人物词典和指代消解词典中，完成指代消解操作。

### 2.2 生成人物关系矩阵

降序统计 181 位人物在两部小说内出现的频次，得到了图 1 的人物频数分布直方图。横坐标为部分人物的姓名，纵坐标代表人物出现的频数。由图可知，分布情况近似服从幂律分布。

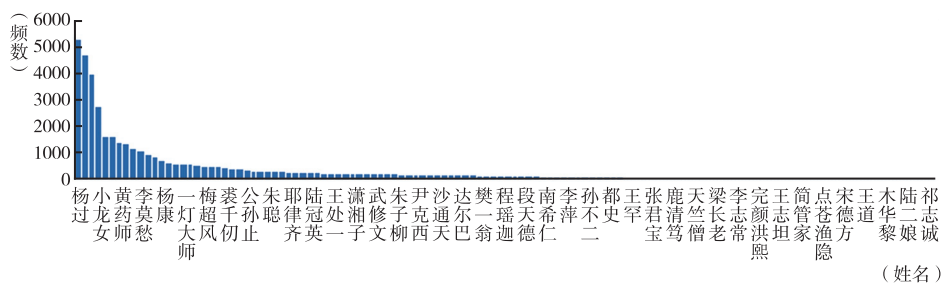
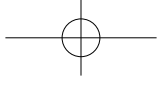


图 1 人物频数分布直方图

以人物频数排名为 x 轴，以人物频数为 y 轴，对 x、y 取双对数<sup>[13]</sup>作拟合直线，检验其是否为幂律分布，最后以误差平方和作为拟合效果评估标准，见图 2。可以看出，“射雕二部”人物频数



姚睿琦, 张辉, 姚云洪. 社会网络分析方法在金庸小说人物关系中的应用研究 [J]. 文献与数据学报, 2021, 3 (3): 068-080.

服从幂律分布, 故具有“长尾”特点, 即主角与部分配角占据较高的频次, 其他配角在文中出现次数极少且数量相近, 在小说分析中没有实际意义。这意味着绝大部分的配角频数可以忽略不计。

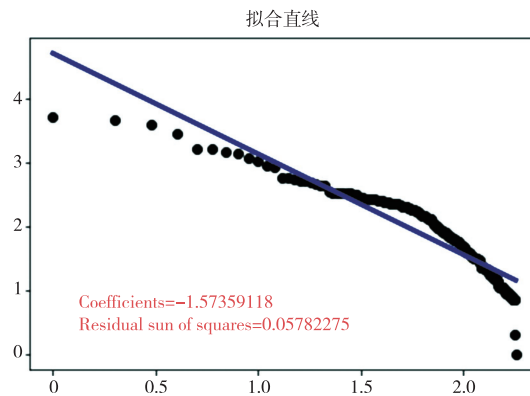


图2 幂律分布拟合直线

因此, 本文删去频次小于 20 的人物, 即在图 2 中位于“长尾尾巴”处的人物, 得到“射雕二部”124 位人物的频数。在小说中, 某人出现的次数越多, 越可以肯定此人的地位, 小说中的主人公杨过、郭靖、黄蓉与小龙女, 配角“五绝”高手以及反派角色金轮国师和李莫愁排名位于前十位, 说明他们在“射雕二部”中处于重要地位。

通过人物频数只能初步了解人物的重要性, 若要定义人物关系的亲密程度, 还需进一步分析人物之间的关系, 并通过关系权重加以判定。对于关系权重, 本文将其定义为: 若任意两人同时出现在某一段落, 则权重 +1, 反之为 0; 总计权重越大, 二人关系越亲密。根据这个定义进行测算, 得到人物邻接矩阵如表 1。

表 1 “射雕二部”人物邻接矩阵 (部分)

	杨过	郭靖	黄蓉	小龙女	周伯通	欧阳锋	黄药师	金轮国师
杨过	5243	862	775	3530	353	271	218	789
郭靖	862	4666	3670	112	588	785	612	258
黄蓉	775	3670	3954	228	360	694	423	179
小龙女	3530	112	228	2734	229	87	21	375
周伯通	353	588	360	229	1647	294	227	198
欧阳锋	271	785	694	87	294	1645	349	1
黄药师	218	612	423	21	227	349	1415	36
金轮国师	789	258	179	375	198	1	36	1159

### 3 人物关系网络分析

#### 3.1 模块化社区探测

将人物关系数据导入 Gephi, 鉴于在网络中存在多个权重很小且较为独立的节点, 而这种节



点对网络的影响很低，因此，为了提高分析效果，先进行初步过滤得到100个节点，再通过社区划分<sup>[14]</sup>得到五个社区，有助于读者初步理解小说人物亲疏关系。人物社区分布见表2。

表2 “射雕二部”人物社区分布

社区一	包惜弱、杨铁心、丘处机、完颜洪烈、杨康、陈玄风、韩小莹、张阿生、柯镇恶、段天德、李萍、朱聪、韩宝驹、全金发、南希仁、马钰、王处一、穆念慈、孙不二、郝大通、郭啸天、刘处玄、子聪、谭处端
社区二	博尔忽、拖雷、都史、铁木真、桑昆、王罕、窝阔台、札木合
社区三	郭靖、梅超风、黄药师、黄蓉、侯通海、欧阳克、华筝、周伯通、鲁有脚、沙通天、彭连虎、梁子翁、灵智上人、简管家、洪七公、欧阳锋、渔人、裘千仞、陆乘风、黎生、傻姑、简长老、瑛姑、一灯、彭长老、樵子
社区四	程瑶迦、陆冠英、尹志平、赵志敬、甄志丙、李志常、王志坦、鹿清笃、祁志诚
社区五	小龙女、杨过、金轮国师、达尔巴、霍都、郭芙、郭襄、樊一翁、大头鬼、潇湘子、裘千尺、冯默风、李莫愁、公孙绿萼、公孙止、完颜萍、武三通、武修文、武敦儒、耶律齐、程英、陆无双、耶律燕、朱子柳、尼摩星、王重阳、忽必烈、尹克西、孙婆婆、洪凌波、觉远、张君宝、天竺僧、林朝英

在社区划分的基础上绘制“射雕二部”的人物关系网络图谱（见图3），用小说人物作为节点，节点颜色代表所分布的社区，其中，绿色、橙色、红色、粉色、紫色分别代表了社区一到社区五，节点大小代表该人物在网络中的度。度的概念来自社会计量学中的“明星”概念，是与该点相连的其他点的度数之和，一个点的度数越大，就越居于中心，而该点对应的行动者的中心程度越高<sup>[15]</sup>。节点连线的粗细代表着人物关系的亲密程度，连线越粗，表示其共现次数越多，关系越亲密。

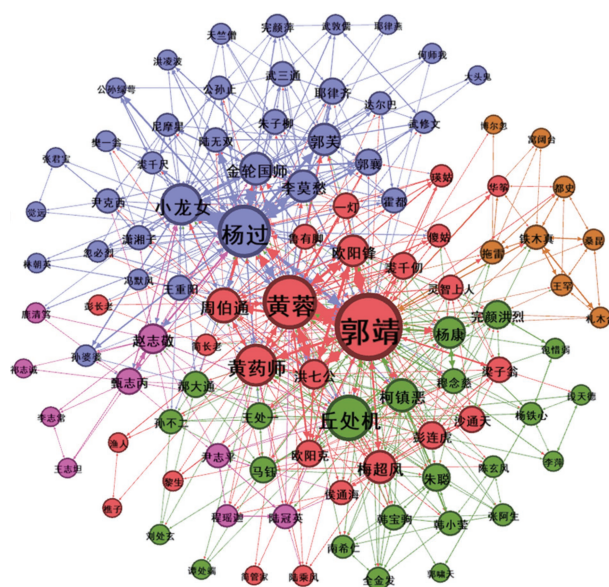
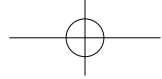


图3 “射雕二部”人物关系网络图谱



### 3.2 k-核分解

为能更好地判断节点的影响力, 本文采用 k-核分解法。k-核是由赛德曼提出的社会网络分析方法, 通过迭代不断减去度小于等于 k 的节点, 即运用最小度标准去确定团聚度高和低的区域, 从而研究网络中的成分结构<sup>[16]</sup>。利用 k-核分解将节点按照边缘层到核心层划分为不同的层次, 核心节点 (Ks 值大的节点) 才是影响力大的节点<sup>[17]</sup>。笔者将 Ks 值设置为最大, 过滤掉边缘层和中间层, 迭代分解出核心层, 得到图 4, 初步确定了重要人物的范围。

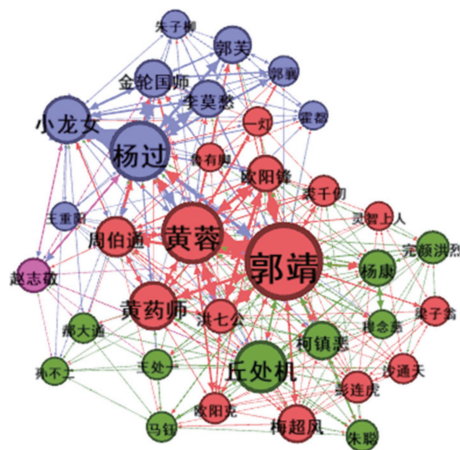


图 4 核心层关系网络图

结合表 2、图 3、图 4 可知, 红色节点组成的社区三和紫色节点组成的社区五的规模最大、平均度最高, 两社区内的部分节点是 k-核分解后的核心层主体, 他们不仅与社区内部的成员亲密程度高, 而且与社区外部人物也保持密切联系, 所以这两个社区可以被看作为“明星社区”。相比之下, 虽然社区一的规模较大, 但是仅个别度较大的成员位于核心层, 成员与成员之间的亲密程度不高, 次于明星社区, 因此称之为“普通社区”。社区二和社区四的规模最小, 社区内外联系都较为稀松, 社区内成员均不在核心层中, 是网络中的边缘节点, 这说明它们只在书中的部分情节中出现, 影响力低, 可称之为“偏远社区”。通过对比发现, 明星社区之间的联系均大于普通社区、偏远社区之间的联系, 说明了作者在人物关系写作中侧重影响力较高的人物互动, 他们之间的联系能够在一定程度上反映小说的写作意图与风格特点。

### 3.3 中心性分析

中心性分析是社会网络分析中的重要环节, 其中, 中介中心性以经过某节点的最短路径数量来判断该节点对信息的控制程度。本文利用中介中心性识别在多个具有亲密关系的人物集体中扮演桥梁作用的节点, 也就是说, 中介中心性越大的节点, 影响其他节点并建立联系的能力就越大。其计算公式为:

$$C_B = \sum_{j < k} \frac{g_{jk}(n_i)}{g_{jk}} \quad (1)$$

式(1)中,  $g_{jk}$  表示节点  $j$  到节点  $k$  之间的最短路径总数,  $g_{jk}(n_i)$  表示节点  $j$  到节点  $k$  的最短路径中经过节点  $n_i$  的数量, 故节点  $n_i$  的中介中心性就是  $g_{jk}(n_i)/g_{jk}$  的概率之和。

接近中心性意味着每个节点到其他节点最短路径的平均长度, 然而, 接近中心性对远距离或缺失的连接非常敏感, 当处理较为复杂的网络时, 断开一个节点会设置到无穷大的距离, 无法判断节点的重要性。因此, 本文采用价值中心性<sup>[18]</sup>代替接近中心性, 它是接近中心度倒数和求和运算的逆运算。价值中心性仅受到断开节点的轻微影响, 能够更好地处理复杂网络, 获得有价值的中心。在小说中, 价值中心性越高的人物越可能处于中心地位, 在信息传递的过程中承担着重要作用<sup>[19]</sup>。式(2)为价值中心性的计算公式:

$$V_c(A) = \frac{\sum \frac{1}{d(A,B)}}{N-1} \quad (2)$$

此式中,  $N$  是人数,  $d(A,B)$  指的是  $A$  和  $B$  之间的短程线距离。

特征向量中心性是基于节点连接来衡量节点重要程度的指标, 一个有高特征向量中心性的行动者, 与他建立连接的行动者往往也被其他行动者所连接, 节点的价值既取决于临近节点的度, 也取决于邻近节点的重要性<sup>[20]</sup>。在小说人物关系网络中, 节点的特征向量中心性越高, 所对应的人物影响力越大。其计算公式如式(3):

$$E_c(i) = x_i = c \sum_{j=1}^n a_{ij} x_j \quad (3)$$

其中,  $c$  是一个比例常数,  $x_i$  被记为节点  $v_i$  的重要性度量值,  $E_c(i)$  代表节点  $v_i$  的特征向量中心性, 它与临近节点的中心性得分总和成正比。

基于此, 本文制定一个衡量人物重要性的标准: 若一个节点属于明星社区, 经  $k$ -核分解后属于核心层, 其度和中心性都位居前列, 那么该节点对应的人物就是小说主角, 其地位最高、对情节控制能力最强; 若一个节点不在偏远社区, 经  $k$ -核分解后属于核心层, 具有较高的中心性, 与主角亲密程度高, 那么可以断定这个节点代表的人物具有较高的影响力和地位, 为小说的重要配角。计算“射雕二部”人物的中心性并删除缺失值, 最终按照数值大小降序排列, 得出郭靖、黄蓉、杨过、小龙女在内的四位主角和十位重要配角(见表3), 均符合小说内容, 说明该方法识别效果较好。

表3 “射雕二部”人物中心性

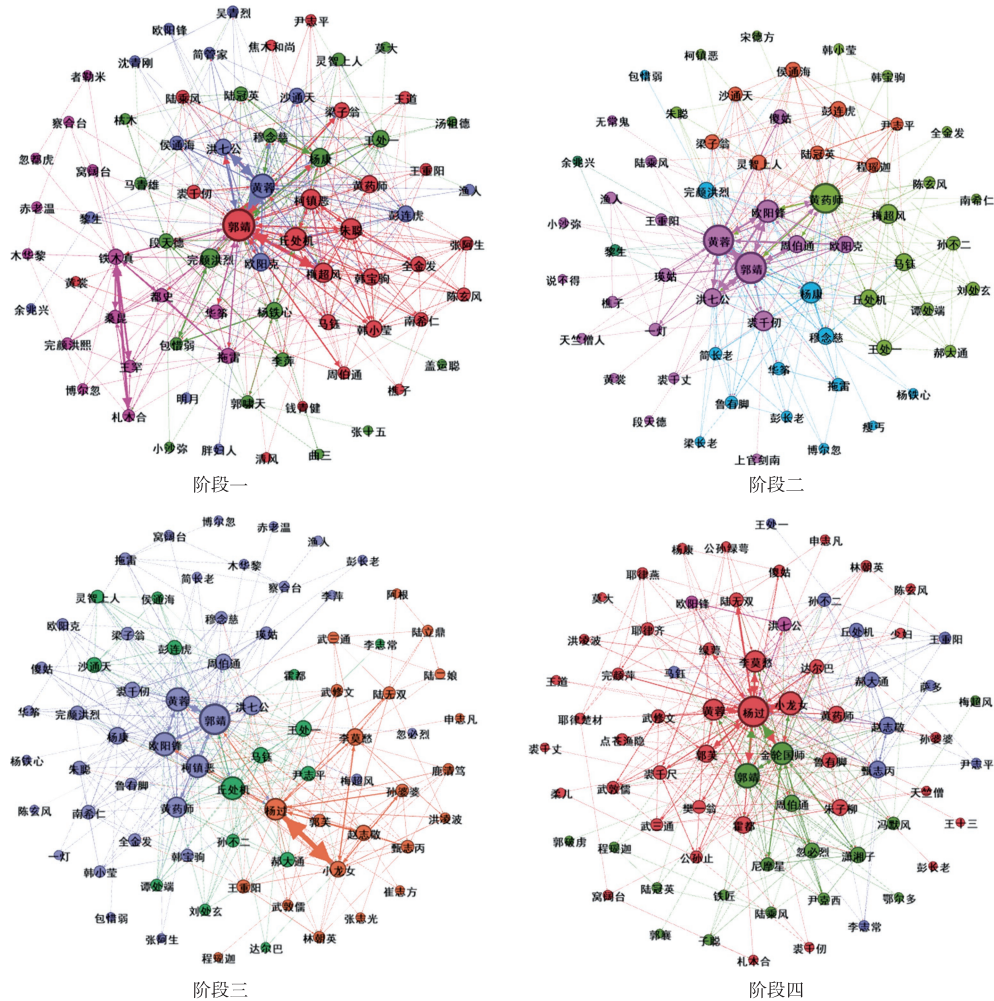
姓名	所属社区	特征向量中心性	价值中心性	中介中心性
郭靖	社区3	1.000	0.810	2487.094
黄蓉	社区5	0.836	0.719	1133.552
杨过	社区3	0.749	0.711	1154.319
小龙女	社区5	0.677	0.667	408.753
丘处机	社区1	0.675	0.661	687.071
黄药师	社区3	0.655	0.628	245.844
周伯通	社区3	0.525	0.598	235.359
欧阳锋	社区3	0.473	0.562	65.447

续表

姓名	所属社区	特征向量中心性	价值中心性	中介中心性
柯镇恶	社区 1	0.469	0.572	78.046
金轮国师	社区 5	0.434	0.583	160.172
李莫愁	社区 5	0.432	0.576	218.646
杨康	社区 1	0.427	0.596	68.589
郭芙	社区 5	0.377	0.568	160.002
郭襄	社区 5	0.334	0.536	169.497

## 4 人物关系演化分析

为了观察主角与重要配角影响力和亲密关系的动态变化, 本文绘制了人物关系网络演化图 (见图 5), 图中将“射雕二部”共 80 章回划分为五个阶段, 分别是: 阶段一, 1~16 回; 阶段二, 17~32 回; 阶段三, 33~48 回; 阶段四, 49~64 回; 阶段五, 65~80 回。





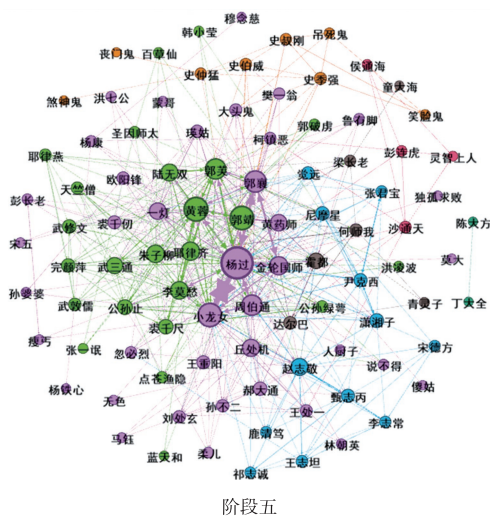


图5 人物关系网络演化图

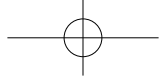
表4 小说人物关系网络结构测度表

阶段	一	二	三	四	五
节点数	74	62	76	72	95
边数	882	616	764	746	1046
密度	0.163	0.159	0.134	0.146	0.117
网络直径	4	4	5	4	5
聚类系数	0.723	0.756	0.719	0.721	0.703
路径长度	1.99	2.0	2.19	2.1	2.2

计算各阶段的网络结构（见表4），结合演化图谱可得出总体情况：

（1）图的密度随着情节的演化有下降趋势，但网络规模（节点、边数）有上升趋势，这说明随着情节的发展，越来越多的角色加入串联。（2）聚类系数是一个点的邻接点之间相互连接的程度，量化了节点间的聚集倾向<sup>[21]</sup>。网络的平均路径长度是所有的网络节点两两建立联系所需最短路径的平均步数，其大小能够说明网络的六度分离属性，即具有小世界特征。各阶段的聚类系数均在0.7~0.8之间，平均路径长度在1.99~2.2之间，与结构相近的随机网络（见表5）相比，“射雕二部”人物关系网的总体聚类系数明显较高，平均路径长度相近，这说明“射雕二部”的各阶段具有明显的“小世界效应”，与真实社会中的人物社交网络具有高度相似性，这意味着小说人物间没有完全的陌生人，均能通过中介人物建立起联系。（3）第一阶段的网络密度最大，各节点之间亲密程度较高；随着时间的推移，二、三、四阶段密度变小，人物的亲密关系更集中在网络核心位置，且存在多个孤立的边缘节点；第五阶段密度最小，但网络规模最大。这说明，“射雕二部”具有时间序列特性和“总-分-总”的结构，小说初期主线情节与支线情节穿插叙述，中期重点叙述主角经历并留下伏笔，后期核心人物逐渐转移，直至最后收尾总结全部情节和伏笔。





姚睿琦, 张辉, 姚云洪. 社会网络分析方法在金庸小说人物关系中的应用研究 [J]. 文献与数据学报, 2021, 3 (3): 068-080.

表 5 随机网络结构测度表

随机网络	1	2	3	4	5
节点数	74	62	76	72	95
边数	881	610	753	712	1048
聚类系数	0.334	0.324	0.265	0.29	0.237
路径长度	2.03	1.7	1.8	1.92	2.1

根据上文确定的主角和重要配角, 以五个阶段为时间切片, 采用特征向量中心性将人物影响力量化, 变化趋势如图 6 所示。

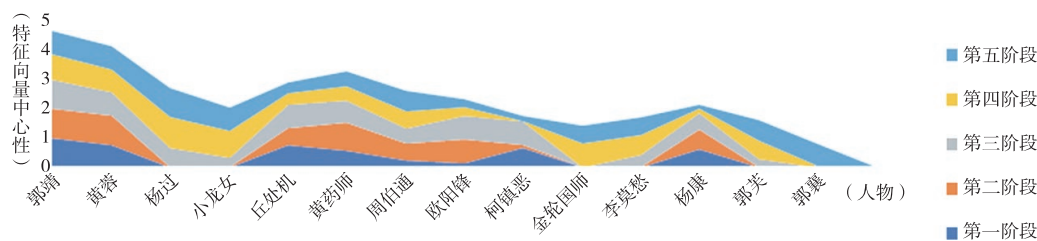


图 6 “射雕二部”人物影响力变化图

第一阶段: 主角阵营中, 郭靖、黄蓉位于网络中心位置, 其特征向量中心性最高, 是这一阶段的核心人物。重要配角中仅有丘处机、柯镇恶、黄药师、杨康的数值排在前列, 其他人物尚未登场, 特征向量中心性为 0, 在本阶段没有影响力。网络中出现大规模有亲密连接的节点对, 不仅集中在重要人物, 还体现在网络边缘的人物中, 说明本阶段着重介绍故事发生的背景与人物关系, 为后续情节埋下伏笔。

第二阶段: 郭靖、黄蓉、黄药师、欧阳锋、周伯通位于网络的中心位置, 网络中的亲密关系集中在他们之间, 意味着他们在本阶段地位最高、控制其他角色行动的能力最强, 主导着小说情节的发展。本阶段丘处机与柯镇恶的中心性下降, 被移到了网络边缘位置, 说明本阶段他们对重要情节的参与度较低。

第三阶段: 本阶段是两部小说的衔接阶段, 意味着《射雕英雄传》的完结与《神雕侠侣》的开篇,《射雕英雄传》完结于郭靖、黄蓉携手离开蒙古前往桃花岛完婚,《神雕侠侣》始于李莫愁灭门陆家及杨过拜欧阳锋为义父, 这期间许多新的人物登场, 因此网络中具有明显而独立的社区分布, 关系的亲密程度也通过社区体现, 分为“射雕人物社区”与“神雕人物社区”。“射雕人物社区”中, 柯镇恶和丘处机的影响力回升, 处于高影响力的欧阳锋、黄药师承担着信息传递的作用, 旨在对《射雕英雄传》故事的首尾呼应与《神雕侠侣》故事的开始; 杨康在《射雕英雄传》结尾惨死, 故影响力下降; “神雕人物社区”结点数量较多但网络结构松散, 核心人物未明确, 各人物均不具备高影响力, 这说明了他们的存在均为小说背景的描写而服务。通过对比第一阶段网络, 同样是背景描述, 作者对《射雕英雄传》更侧重于国家民族的范畴, 小说起笔即反映了宋朝的腐朽、金国的贪婪和蒙古的野心; 而《神雕侠侣》始于人与人之间的恩怨情仇, 全篇更着重



刻画男女主的凄美爱恋，后期才将家国大爱与个人小爱结合描写。

第四阶段：“射雕”部分角色退出故事舞台，“神雕”人物初步登场。本阶段标志着小说核心人物开始从郭靖、黄蓉向杨过、小龙女转变，首先体现在杨过和小龙女取代郭靖、黄蓉成为网络的中心，其中心性均高于郭靖和黄蓉；其次郭靖和黄蓉被分到不同社区，亲密程度降低。“射雕”社区的柯镇恶、丘处机、欧阳锋等重要配角位于网络边缘，说明他们不再对情节发展起重要作用，取而代之的是“神雕”社区的金轮国师、李莫愁和郭芙，符合小说原文中剧情的发展。

第五阶段：“射雕二部”人物集结的阶段，标志着小说故事的正式完结。两部小说的男主角郭靖、杨过位于网络中心位置，其余重要角色均有着较大的影响力，前四期中心性处于低水平的郭襄在本阶段数值大幅上升，影响力和地位提高，说明该阶段作者着力于描写她的经历，郭襄的一生，是《神雕侠侣》的终结，也是“射雕”第三部《倚天屠龙记》的开始，对她的描述方式展示了作者伏笔写作的风格。该阶段“侠情”是作者写作灵魂的升华部分，与人物和情节高度融合<sup>[22]</sup>，成为了故事发展的重要线索：一是对“情”的刻画，通过对杨过和小龙女情感的细腻处理，赋予其笔下人物鲜明生动的艺术形象；二是对“侠”的弘扬，杨过被郭靖、黄蓉保家卫国不顾生死的大义所感染，成为了百姓歌颂的“神雕侠”，最后襄阳大战时对重要人物的刻画也体现了侠之大者的家国情怀。

综上，根据重要人物演化特点分析如下：

(1) 主角阵营中，郭靖、黄蓉的累计影响力与阶段影响力最高，在小说中承担了大量描写，是“射雕二部”所刻画灵魂人物，透过二人所表达的侠义是作者在人物描写背后的精神支撑。杨过、小龙女在后两个阶段接替郭靖、黄蓉成为核心人物，二人形象的塑造表现了追求自由与极致的现代爱情意识。

(2) 在重要配角阵营中，黄药师、周伯通、欧阳锋的累计特征向量中心性仅次于郭靖、黄蓉，与主角的亲密程度保持着较高水平，他们的影响力持续时间长，并在某阶段具有和主角一般的突出地位，起贯穿小说全文的作用。他们迥异的宗师作风在小说中体现得淋漓尽致，对主角们武功和观念的养成具有较大影响。

(3) 丘处机、柯镇恶的特征向量中心性变化极为相似，有着同增共减的特点，两人皆为性格豪迈、好恶斗勇的侠客形象，作者树立他们的形象对主角成长起帮助作用。

(4) 金轮国师、李莫愁、杨康、郭芙仅在某两到三个阶段影响力高，这四人均对主角的成长产生阻碍影响，但持续性并不强。

## 5 结语

本文将中文信息处理和社会网络分析方法应用于小说人物关系研究，并添加语料库进行训练，定义了亲密关系和人物重要性的衡量标准，识别出“射雕二部”的人物重要程度，然后以时间序列绘制人物关系网络，计算网络结构和人物特征向量中心性，以此反应小说背景与人物特征，揣摩作者的写作风格，发掘人物影响力和亲密关系的变化趋势，最后对其分类总结，为大数据时代对小说的定量研究提供新思路。



姚睿琦, 张辉, 姚云洪. 社会网络分析方法在金庸小说人物关系中的应用研究 [J]. 文献与数据学报, 2021, 3(3): 068-080.

研究中, 笔者发现“射雕二部”的社会网络中人物频数呈现出幂律分布特性, 各阶段网络均具有小世界效应, 表明作者塑造的武侠世界高度符合现实社会。通过 k-核分解与中心性分析确定小说的重要人物, 通过书写他们的喜愁悲欢和精神情感来演绎武侠故事<sup>[23]</sup>, 反映整本小说的世界观。在演化过程中发现, 总体特征向量中心性高的角色在某一阶段的影响力不一定高, 但是他们主导着全书重要情节的发展; 而部分人物仅在个别阶段有高影响力, 其数值的提升代表着作者伏笔的设置或支线情节的安排, 累计特征向量中心性处于较低水平, 表明他们没有起到贯穿全文、推动情节发展的作用, 这说明了作者对人物特点及铺垫式写作的把控力。

需要说明的是, 本文在数据预处理环节设计了人物词典校对和人物指代消解词典, 虽然节省了大量时间成本和人力成本, 有助于大数据环境下的定量分析, 但是对于具体的代词和指代不明确的词处理还不够细致, 部分选择了删除。其次, 人物亲密关系的量化方法过于简单, 仅用社会网络中的人物分析方法, 未结合人物对话情感分析语料, 有待进一步深究。

#### 【参考文献】

- [1] 赵京胜, 宋梦雪, 高祥. 自然语言处理发展及应用综述 [J]. 信息技术与信息化, 2019(7):142-145.
- [2] 宗成庆. 中文信息处理研究现状分析 [J]. 语言战略研究, 2016,1(6):19-26.
- [3] 陈芯莹, 李雯雯, 王燕. 计量特征在语言风格比较及作家判定中的应用——以韩寒《三重门》与郭敬明《梦里花落知多少》为例 [J]. 计算机工程与应用, 2012,48(3):137-139,208.
- [4] 肖天久, 刘颖. 基于聚类和分类的金庸与古龙小说风格分析 [J]. 中文信息学报, 2015,29(5):167-177.
- [5] 刘海燕, 尹晓虎. 文学作品中的“小世界”——菲茨杰拉德小说人物关系网络的实证分析 [J]. 统计与信息论坛, 2015,30(12):102-107.
- [6] 赵京胜, 张丽, 朱巧明等. 中文文学作品中的社会网络抽取与分析 [J]. 中文信息学报, 2017,31(2):99-106,116.
- [7] 王一博, 俞敬松, 赵常煜. 共词方法在三国人物关系分析中的应用研究 [J]. 情报探索, 2017(7):52-56.
- [8] 钮亮. 水浒人物社会网络统计分析 [J]. 文献与数据学报, 2020,2(2):99-109.
- [9] 杨鑫, 宋卓远, 朱东霖等. 基于共词分析的小说人物关系研究 [J]. 现代计算机, 2019(35):3-7.
- [10] 胡云鹏. 中文文学作品中的社会网络构建与分析 [D]. 杭州: 浙江工商大学, 2019.
- [11] 魏会洋, 袁曦临. 社会网络分析在文学阅读研究中的适用性问题——以数字人文视角下的《白鹿原》人物关系阐释为例 [J]. 新世纪图书馆, 2019(3):30-34.
- [12] 张旋, 梁循, 李志宇等. 金庸小说中主角复杂爱情模式的识别与分析 [J]. 中文信息学报, 2019,33(4):109-119.
- [13] 李杰, 孙仁诚. 基于词频统计算法的中英文词频分布研究 [J]. 青岛大学学报(工程技术版), 2020,35(1):1-5.
- [14] VINCENT D B, JEAN-LOUP G, RENAUD L, ETIENNE L. Fast unfolding of communities in large networks [J]. Journal of Statistical Mechanics: Theory and Experiment, 2008(10):1-11.
- [15] SCOTT John. 社会网络分析法 [M]. 刘军译. 重庆: 重庆大学出版社, 2020:91-92.
- [16] 林聚任. 社会网络分析: 理论方法和应用 [M]. 北京: 北京师范大学出版社, 2009:113-115.
- [17] 韩忠明, 陈炎, 刘雯等. 社会网络节点影响力分析研究 [J]. 软件学报, 2017,28(1):84-104.
- [18] DEKKER A. Conceptual distance in social network analysis [J]. Social Structure, 2005,6(3):1-34.
- [19] 平亮, 宗利永. 基于社会网络中心性分析的微博信息传播研究——以 Sina 微博为例 [J]. 图书情报知



识,2010(6):92-97.

[20] 朱丽波. 科学合作网络的中心性分析 [J]. 图书馆学研究,2015(3):97-101.

[21] ANDREW G, MAX W, AMITAVA D. Converting a network into a small-world network: Fast algorithms for minimizing average path length through link addition [J]. Information Sciences, Volume 422, 2018:282-289.

[22] 庄国瑞. 论金庸武侠小说对“侠情”的开拓 [J]. 浙江学刊,2019(6):32-43.

[23] 吴秀明. 人学视域下的金庸武侠小说及其当下意义 [J]. 文学评论,2020(2):213-223.

## Research on Application of Social Network Analysis on Character Relationships in JinYong's Novels

YAO Ruiqi ZHANG Hui YAO Yunhong

(School of Management, Shandong University, Jinan 250100, China)

---

**Abstract:** [ **Purpose/significance** ] This paper tries to discuss the character influence and relationship change and analyze the writing intention and features of the novels. [ **Method/process** ] Based on Chinese information processing and social network analysis, this paper takes “The Legend of the Condor Heroes” and “The Return of the Condor Heroes” as research subjects and extracts the character relation networks. The importance of characters is analyzed through the community detection and k-core and centrality. The evolution of the network structure and character influence are then analyzed through the time series evolutionary analysis. [ **Result/conclusion** ] The results show that the character networks have small-world effect. The character influence and their intimate relationships are obvious. It reveals the author's writing style.

**Keywords:** Fiction social networks; Character relationship; Chinese information processing; Evolutionary analysis

---

( 本文责编: 王秀玲 )