

【数字·人文】

数字人文视角下古籍资源知识挖掘与可视化研究*

——以《隋唐演义》为例

●董守轩

中国海洋大学文学与新闻传播学院,青岛,266100

[摘要]文章从数字人文角度出发,以图书馆馆藏资源及各类电子数据库为主要来源,运用 SNS、Python、Gephi、QGIS 等技术和软件,通过数据模型和传统图表结合的形式,绘制历史演义小说《隋唐演义》人物关系及时空分布的可视化图谱。相关的分析结果有助于完成古籍的深入挖掘,丰富人文研究的范式,为古籍资源的整合与知识数据的联结提供新的思路。

[关键词]数字人文 《隋唐演义》 人物关系 时空分布

[中图分类号]G353

[文献标识码]A

[文章编号]1003-7845(2024)01-0056-08

[引用本文格式]董守轩.数字人文视角下古籍资源知识挖掘与可视化研究——以《隋唐演义》为例[J].高校图书馆工作,2024(1):56-63.

引言

数字人文(Digital Humanities),是以数据为基础,以平台为支撑,运用数字技术方法来研究人文学科的一门学科^[1]。这一术语最初来自人文计算(Humanities Computing)领域,早期的人文计算仅是一种纯粹的技术工具。此后,随着技术的进步及时代的发展,人文计算理论与实践不断完善,逐渐被用于一些人文社科领域。20世纪90年代初,人文计算开始被当作一个独立的交叉学科^[2]。2001年,约翰纳·德鲁克(Johanna Drucker)计算机与人文协会和在文学与语言计算协会(ACH/ALLC)联合会议上作的报告《现状核实:数字人文项目与前景》(Reality Check: Projects and Prospects in Digital Humanities)是“数字人文”在学术领域的首次出现。2004年,苏珊·施莱布曼等人编纂的《数字人文指南》使用数字人文概念,标志着这一术语从人文计算到数字人文的转变^[3]。

事实上,数字技术的发展对传统人文学科产生了极为重要的影响,比如各类古籍数据库的出现,如爱如生中国基本古籍库^[4]、鼎秀古籍全文检索平台^[5]、中华经典古籍库^[6]等。这些数据库往往存储了海量的古籍文献,研究者可以通过其检索功能迅

速、精确、全面地找出目标古籍的相关内容,从而开展研究。部分高校图书馆还提供书籍的电子扫描、文本识别、资源共享等服务,对文献资料的保存、整理、研究起到重要作用。除数据库外,数字人文还促进了学术地图平台及各类人物世系数据资源的建构,如王兆鹏教授团队开发的“唐宋文学编年地图”平台^[7],徐永明教授团队开发的“学术地图发布平台”^[8],美国哈佛大学包弼德教授联合北京大学等机构合作建立的“中国历代人物传记资料库”^[9]等。目前,部分高校图书馆也将这些非文献类的新型数据库纳入馆藏,作为师生学术研究的重要依据及来源。

除上述宏观层面的实践外,部分学者也开始使用 SNS、GIS、NLP 等数字技术及相关软件对文学作品进行数字化呈现。如赵薇^[10]借助 SNS 对《大波》三部曲人物关系及社会网络进行可视化研究。徐永明等^[11]利用 GIS 软件将《全元诗》作者的地理分布进行可视化呈现。孙建旺等^[12]借助 SVM 作为机器学习模型,分析微博文本的不同情感类型。传统文学作品的早期研究只能通过例证的形式对其中的主要人物或篇目进行分析,难以做到全面覆盖。而利用数字技术处理文本信息,则可以帮助研究者进行

* 本文系山东省高等学校青年创新科技计划“儒家文化与文学关系研究”(项目编号:2020RWC004)的研究成果之一。

作者简介:董守轩,博士研究生,研究方向为数字人文。

收稿日期:2023-10-29

责任编辑:杨千子

“全景式”阅读,有效提取目标信息,并通过软件进行可视化。这一过程实际上是借助大数据计算及科学图谱模型完成的,利用数字技术处理文本信息改变了传统人文研究的固化思维,并为其提供了新的研究路径。

《四雪草堂重订通俗隋唐演义》(以下简称《隋唐演义》)是清代文学家褚人获所编的长篇章回体小说,也是明清小说的代表之作。全书以史为经,以事件为纬,叙述了隋朝至中唐的百余年历史,塑造了众多的人物形象,故事情节中蕴含大量地名。书中的人物涉及社会各个阶层,既有王侯将相,也有平民百姓,这些人物以时间为序依次出场,构成了交错复杂的网状结构。因此,本文以书中的人物关系、空间分布为例,采用数字人文技术及相关理论,运用各类数据库、数字软件,以数据模型和图表相结合的形式,对书中人物的社交网络及地理空间分布等要素进行可视化呈现,将书中的人名、地名等实体具象化,并挖掘其内在的联系,从而为数字人文研究提供借鉴。书中的地名涵盖南北,基本囊括了隋唐时期的重要城市,具有重要的文学与史学价值,对了解明清小说家的创作倾向、隋唐的历史发展及图书资源知识的发现有着深刻的意义。

1 《隋唐演义》数据来源及数据清洗

本文选取鼎秀古籍全文检索平台扫描识别的《隋唐演义》电子版为原始数据来源,利用 Python、Gephi、QGIS 等工具,通过数据采集、数据清洗、数据处理、数据可视化、数据分析等流程对《隋唐演义》进行全面而系统的研究,实现对小说知识的挖掘和发现。其中,数据采集阶段主要是对书的文本进行爬取,并以 txt 文件的形式保存爬取出来的数据。数据清洗阶段是利用 Python 中的 jieba 库进行分词,提取书中的人物、地名等内容,并绘制相应的表格。数据处理阶段是运用 Gephi、QGIS 等数字软件对书中的社交网络、地理分布等进行分析。数据可视化阶段是运用 Gephi 软件对数据进行可视化呈现,形成可视化图谱。数据分析阶段主要是对可视化图谱内容展开分析发掘,实现文本内容与数字技术的结合,具体流程如图 1 所示。

2 《隋唐演义》社会网络与人物共现关系

首先,利用 Python 中的 jieba 库对《隋唐演义》的 txt 文本进行分词,通过 jieba 库的 posseg 模块提取人名,排除非人名词语。然后,结合书的具体文本内容及相关的历史文献,对同一个人物的不同指代称谓进行设置,建立映射词表。通过软件得到具体

数据后,再结合书的文本内容进行人工比对和增补,由此得出《隋唐演义》所有人物的人名数据。再次,使用 Python 对《隋唐演义》txt 文本进行识别,如果两个人物同时出现在同一段落中,则可认为二者存在某种联系。经数据分析与人工核对统计,最终得到书中的 302 位人物及 7 220 组共现关系。最后,将人物与共现关系数据导入 Gephi 软件进行处理,获得人物共现关系的可视化图谱,并进行网络直径、网络平均度等网络特征指标的计算。

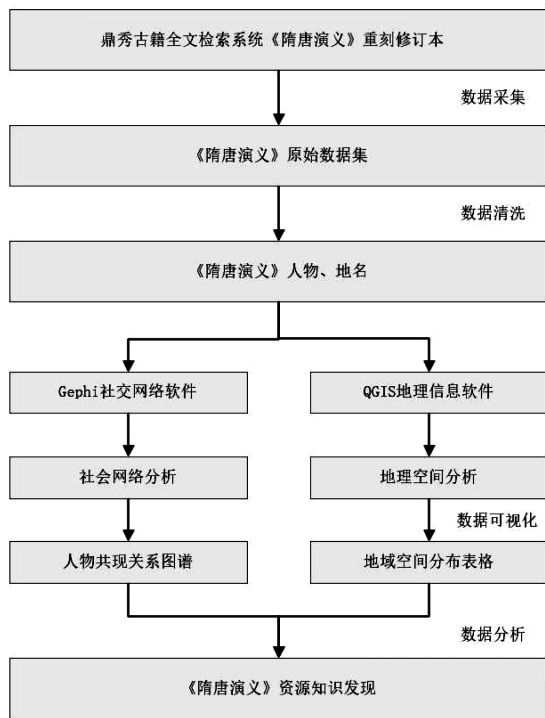


图 1 《隋唐演义》资源知识发现流程

2.1 《隋唐演义》人物关系挖掘

2.1.1 共现网络关系图谱建构

图,也叫网络,表示为 $G=(V,E)$,其中 V 表示图中的节点集合, E 表示图中边的集合^[13]。在图论中,度数表示一个节点的连接数,即与该节点相连接的边的数目,一个节点的度数越大就意味着这个节点在某种意义上越“重要”^[14]。节点的度与节点的大小相关,节点度值越大则节点面积越大。

将所得人物与共现关系数据绘制到 Excel 表格,以 CSV 格式保存,利用 Gephi 软件读取表格数据,并进行人物矩阵网络关系图绘制。在操作过程中设置节点的大小,节点大小与度相关,度越大则节点面积越大。此外,连接两节点的边的粗细程度反映了两节点联系的紧密程度,两节点间的边越粗则代表联系越紧密,可视化结果如图 2 所示。

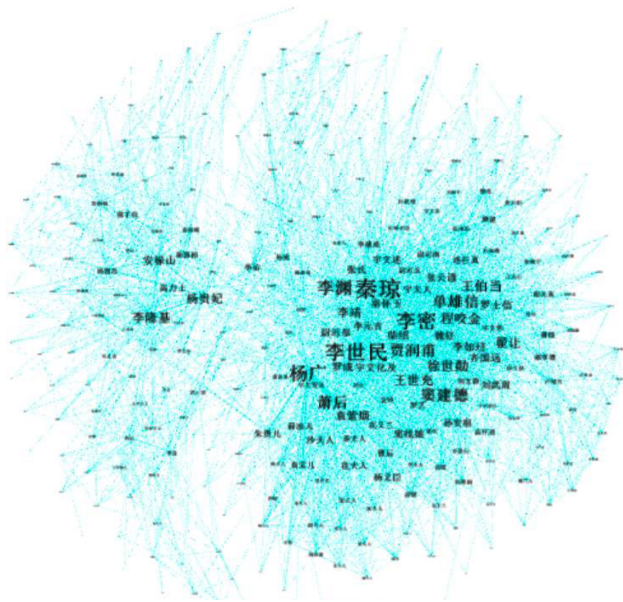


图 2 《隋唐演义》人物共现网络关系

2.1.2 人物共现关系的网络特征指标分析

人物共现关系网络的拓扑结构可以通过网络平均度、平均路径长度、平均聚类系数等指标进行分析,《隋唐演义》人物共现关系的网络特征指标数据如表 1 所示。

表 1 《隋唐演义》人物共现关系的网络特征指标

数据集	节点数	边数	网络直径	网络平均度	平均聚类系数	平均路径长度
人物共现关系网络	302	7 220	6	23. 91	0. 684	2. 4

通过表 1 可以看出,本次录入的人物共有 302 位,他们相互之间构成的社交关系共 7 220 条。共现关系网络中的人物,平均需要 2—3 步即可建立联系,而最长则需要 6 步才能建立。网络中的每个人物平均与 23—24 个其他人物存在共现关系。每个人物之间的平均聚类系数为 0. 684,即共现概率为

68. 40%。综上所述,《隋唐演义》中的人物联系较为紧密,具有较为明显的小世界特性。

2.1.3 基于中心度的网络人物中心性分析

计算出人物共现关系网络中每个人物的中心度,统计出《隋唐演义》前 40 位核心人物,按照中心度的大小依次排列,结果如表 2 所示。

表 2 基于中心度的《隋唐演义》前 40 位核心人物

人物	中心度	人物	中心度	人物	中心度	人物	中心度
秦琼	134	徐世勣	77	柴绍	61	尉迟恭	55
李世民	119	程咬金	75	罗士信	61	刘武周	54
李密	117	王世充	73	罗成	61	李元吉	54
杨广	110	李隆基	72	齐国远	59	宁夫人	53
李渊	101	翟让	72	宇文文化及	58	宇文述	53
窦建德	90	李靖	69	秦怀玉	58	孙安祖	53
萧后	88	杨贵妃	68	魏征	57	沙夫人	52
单雄信	88	袁紫烟	63	张公谨	56	杨义臣	52
贾润甫	87	安禄山	62	窦线娘	56	朱贵儿	50
王伯当	77	李如珪	62	张氏	55	花夫人	49

结合图2和表2可以看出,秦琼的度中心性最高,在人物关系网络中处于中心地位。秦琼出场时间早、社交网络广,全书以其为中心连接了众多人物。从小说来看,秦琼出身没落的将门,混迹于市井草莽之中,因而结交了各路绿林豪杰。本书的前半部分以秦琼为中心,塑造了秦母宁夫人、秦妻张氏等女性形象,以及单雄信、王伯当、徐世勳等英雄形象。随着事件的发展引出瓦岗寨与李唐的线索。李渊和李世民作为唐王朝的缔造者,广泛联络豪杰、讨平各方势力,自然是书中的核心人物。而徐世勳、程咬金、李靖等人都是经历隋末动乱和大唐初建的功臣,与唐密切相关,也是较为重要之人。李密、窦建德、王世充及与之相关的单雄信、王伯当等人作为起兵反隋的重要领袖,在小说前半部分有较多的戏份,因此地位也比较突出。至于隋炀帝杨广、萧后,及与之相关的宇文文化及、沙夫人、花夫人等,在小说中作为反面形象出现,在书里是连接隋唐历史交替的重要纽带,具有不可或缺的作用。此外,还有李隆基、杨贵妃、安禄山等历史人物,虽在小说中出场较晚,但也具有多的戏份,处于重要的社交网络位置,因此地位也较高。

2.2 《隋唐演义》社交网络人物群体分析

2.2.1 基于模块化的凝聚子群分解

一般认为,模块化指数大于0.3,即代表网络具有较明显的社区结构,真实世界社区的模块化指数通常介于0.3—0.7之间^[15]。经过计算,《隋唐演义》网络的模块化指数为0.482,表明其具有较为明显的网络社区结构。利用 Gephi 模块化功能对《隋唐演义》中的不同群体进行着色,以图谱的形式呈现人物的共现关系网络,从而较为清晰地区分互动相对频繁的群体,结果如图3所示。

从图3可以看出,《隋唐演义》人物的共现关系网络大致可分为4个群体,分别是以秦琼等为中心的A群体,占比43.05%;以杨广为中心的B群体,占比33.11%;以李隆基为中心的C群体,占比22.52%;以王勃为中心的D群体,占比1.32%,该群体完全处于边缘位置。除群体间的不同外,人物的共现关系网络还存在“中心—边缘”的差异。人物的共现关系网络边缘存在的大量孤立节点,说明众多边缘人物共现关联较少、地位较低。

2.2.2 基于 k -core 的人物群体过滤

利用 k -core 社交网络对人物的共现关系网络进行过滤,可以更为直观地看出核心人物及其存在的共现关系。根据计算出的人物关系网络特征数

据,结合本研究的实际情况,最终以 $k=12$ 为标准对《隋唐演义》中人物的共现关系网络进行过滤,保留核心人物群体共现关系,并进行可视化呈现,结果如图4所示。

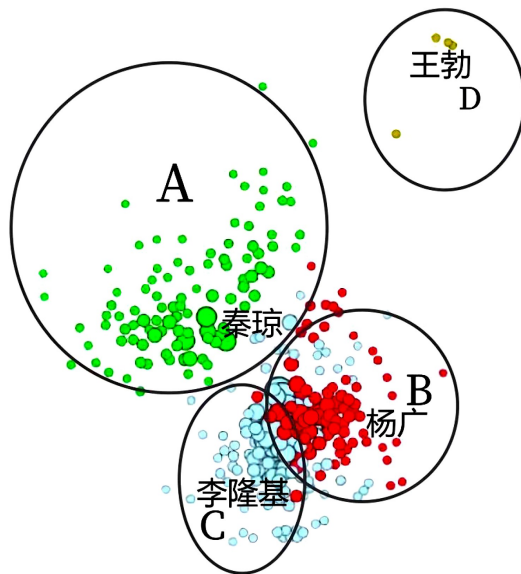


图3 《隋唐演义》人物模块化解构共现网络

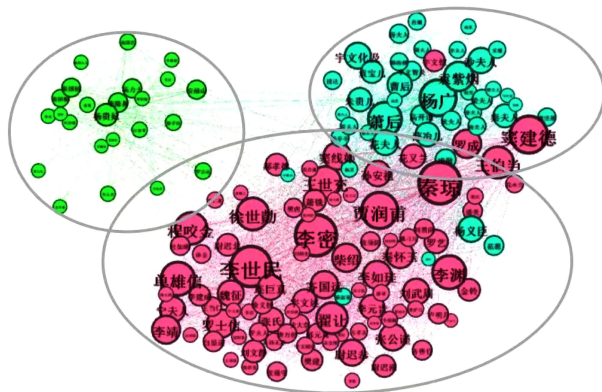


图4 《隋唐演义》人物 k -core 结构共现网络

由于设置了 $k=12$ 的参数,故图3人物模块化解构共现网络中的D群体因重要程度较低而消失。需要说明的是该群体比较特殊,仅有王勃、宇文钧、阎伯屿、吴子章4人。此4人仅出现在小说的第71回“武才人蓄发还宫 秦郡君建坊邀宠”。究其原因,是作者褚人获在撰写《隋唐演义》的过程中,有意地将王勃为滕王阁撰写《滕王阁序》的历史典故穿插进去。此章过后,书中便再未对该群体4人的生平活动进行交代。

2.2.3 核心人物群体分析

根据《隋唐演义》人物在共现关系网络中的中心度排名、基于模块化的凝聚子群分解以及基于 k -

core 的人物群体过滤,统计并划分前 40 名核心人物群体,结果如表 3 所示。

表 3 《隋唐演义》核心人物群体划分

编号	核心人物	核心人物数/所在群体数
1	秦琼、李世民、李密、李渊、窦建德、单雄信、贾润甫、王伯当、徐世勣、程咬金、王世充、翟让、李靖、李如珪、柴绍、罗士信、罗成、齐国远、秦怀玉、魏征、张公谨、窦线娘、张氏、尉迟恭、刘武周、李元吉、宁夫人、宇文述、孙安祖	29/130
2	杨广、萧后、袁紫烟、宇文化及、沙夫人、杨义臣、朱贵儿、花夫人	8/100
3	李隆基、杨贵妃、安禄山	3/68
4	无	0/4

《隋唐演义》的人物分属 4 个群体,设置编号为 1、2、3、4。结合表 2 可以看出,前 40 位核心人物在其中 3 个群体中都有分布。以秦琼、李世民、李密等人为代表的群体 1 所涵盖的核心人物最多。这些人物主要活动于小说的前期和中期,部分经历了隋唐两朝,见证了隋亡唐兴的历史进程,也是书中最为重要的一个群体。从人物构成来看,其主要是李渊父子及隋末的各路豪杰,如李密、窦建德、贾润甫等。需要注意的是,宇文述虽被归为这一群体,但其与以杨广为代表的群体 2 也有着密切联系,情况较为特殊。

以杨广、萧后、袁紫烟为代表的群体 2 主要活动在隋末这一时间段,在书中的前半部分出场。从人物构成来看,这一群体具有明显的“隋”属性,多为隋朝的武将谋臣和后宫妃子。其中,萧后、袁紫烟等都是后妃,长期陪伴在杨广周围。宇文化及、杨义臣两人一奸一忠,都是杨广较为倚重的臣子。他们常出入宫廷,为杨广出谋划策,也形成较为紧密的人物关系。

群体 3 的人数较少,只有李隆基、杨贵妃及安禄山 3 人。他们虽出场时间较晚,但却是全书后半部分的关键人物。李隆基晚年沉迷享乐,任用李林甫、杨国忠等人,导致朝政腐败,最终爆发安史之乱。在小说里,李隆基专宠杨贵妃,两人往来频繁,关系极为密切。而安禄山前期刻意讨好唐玄宗李隆基,后期则形成敌对,构成了复杂的人物关系。

上述群体不仅在内部互相联系,还在外部与其他群体保持沟通与互动。尤其是群体 1 和群体 2 的多位人物都有着双向的互动,如群体 1 的所有人原先都是群体 2 中隋炀帝杨广的臣子,李渊、李世民更是传统意义上的“皇亲国戚”,只是随着隋炀帝杨广

的倒行逆施才反叛。此外,群体 2 中的袁紫烟本是隋炀帝的妃子,隋亡后流落民间,结识了群体 1 中的贾润甫、徐世勣等人,并最终嫁给徐世勣,加入李氏父子的阵营。由此完成了从群体 2 到群体 1 的转变。而群体 3 的李隆基在小说中不仅是李渊的后代,还被设定为隋炀帝宠妃朱贵儿的托生,而杨贵妃则相应地成为杨广的转世,两人由隋入唐,经历了两世的姻缘结合,这一点带有轮回果报的宗教思想。因此小说不同群体人物之间往来频繁,联系紧密。

2.3 数字人文视角与传统人文视角人物研究对比

与传统人文视角相比,数字人文视角下的《隋唐演义》人物研究具有以下 3 个新的特点:其一,研究的人物数据更多、涉及范围更广。数字人文研究借助数字技术,对书中的全体人物进行分门别类的详细探究,研究对象更加全面,研究结论更加可信。其二,利用 Python、Gephi 等软件将书中的文本信息进行可视化呈现,还原所有人物的社交网络关系,能更为直观地识别人物之间的社会关系及联系的密切程度。其三,在图谱的基础上,运用科学的 SNS 理论分析文本,借助相关的技术理论将书中人物划分为不同的凝聚子群,了解其内部关系及外在联系,更好地把握书中的人物布局。

3 《隋唐演义》府、州、郡的时空分布

首先,利用中国历史地理信息系统(CHGIS)及中国历代人物传记资料库(CBDB)的地名沿革系统,限定时期为“隋(公元 581 年—公元 618 年)”——“唐(公元 618 年—公元 907 年)”,查找书中府、州、郡的经纬度坐标,保留 3 位小数。其次,将书中府、州、郡地名和出现频次绘制成 Excel 表格,结果如表 4 所示。

表4 《隋唐演义》中的府、州、郡空间及密度分布

古今地名	出现频次	古今地名	出现频次	古今地名	出现频次
长安(西安)	256	济阳	10	武阳(大名)	3
洛阳	133	荥阳	10	贝州(今属清河)	3
潞州(长治)	113	清河	10	汴州(开封)	3
太原	77	睦州(杭州)	10	荆州	2
扬州	72	汾州(汾阳)	8	丹阳	2
齐州(济南)	70	淮安	6	沧州	2
幽州(北京)	70	华州(渭南)	6	徐州	2
涿州(保定)	46	兖州(济宁)	6	蒲州(永济)	2
睢阳(商丘)	44	庐陵(吉安)	5	营州	2
朔方(灵武)	37	瀛州(今属沧州)	5	魏州(今属大名)	2
南阳	21	永宁(今属北京)	5	襄阳	1
熊州(宜阳)	18	成都	5	荣州(自贡)	1
饶阳	17	鄂州(武汉)	4	寿州(寿县)	1
平原(德州)	14	岷州(岷县)	4	涪州(翼城)	1
泽州(晋城)	12	安州(安陆)	4	康州(德庆)	1
青州	12	凤翔(宝鸡)	4	江州(九江)	1
常山(石家庄)	11	灵昌(滑县)	3	眉州	1
陈留(开封)	11	解州(运城)	3	广州	1
濮州(濮阳)	10	平阳(晋州)	3	朔州(朔县)	1

3.1 《隋唐演义》府、州、郡时间分布特点

作为一部历史演义小说,《隋唐演义》以时间顺序进行叙事,但时间线索相对隐晦,只能通过不同皇帝的在位情况予以辨别。从书中的文本内容来看,大体可分为3个阶段:隋文帝—隋炀帝时期,唐高

祖—唐太宗时期,唐高宗—唐代宗时期。将《隋唐演义》的txt文本按照上述3个阶段进行划分,并将其保存为3个txt文件,运用Python软件依次统计长安、洛阳、潞州等7个府、州、郡在3个时期的出现频次,绘制出河流图,结果如图5所示。

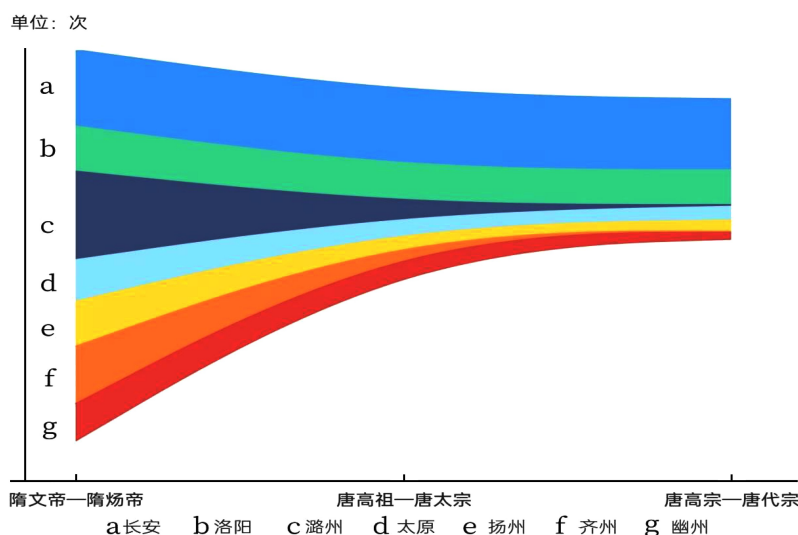


图5 《隋唐演义》代表城市各个时期频次河流分布

从图 5 中可以较为直观地看出,除西都长安、东都洛阳在 3 个时期内的分布较为均匀外,其余府、州、郡的分布都呈现出明显的下降态势。这些州、府、郡都集中在书中的第一个阶段(隋文帝—隋炀帝时期),而到了唐代以后,他们的出现频次陡然减少。结合书中的具体内容来看,隋末群雄并起、逐鹿中原的情节显然是全书的中心,也是作者着力刻画的部分。书中的主要人物秦琼、李世民、李渊等主要活跃于这一时期。而自第 47 回隋炀帝身死,至第 100 回唐代宗即位。作者用了一半的篇幅描写了唐朝百余年的历史,略显冗长散漫。通过历时变化可以看出,书中所讲述事件的后期从关注全国各地的人物事件,到只关注东、西二京的历史人物,经历了由整体地域空间到局部中心城市的转变。

3.2 《隋唐演义》府、州、郡空间分布特点

《隋唐演义》中出现的府、州、郡一级地点共 57 处,涉及较多的地名。从表 4 空间及密度分布来看,频次区间为“1—5”的府、州、郡级地名有 30 处,约占地名总数的一半。此外,区间为“6—10”的地名有 9 处,区间为“11—20”的地名有 7 处,区间为“21—50”的地名有 4 处,区间为“51—100”的地名有 4 处,区间为“101—300”的地名有 3 处。可见,即使是那些重要性不高的府、州、郡,《隋唐演义》也给予了一定的记录。而涉及重要人物、事件的地方,更是加强着墨。“记所遇之地,即系地;记相见之时,即编年。”^[1]《隋唐演义》已具备一定的“系地”意识,有意识地对地域空间进行记录,从而使文本内容更加真实可信。此外,这样的行文方式对事件的展开起到了一定的辅助作用,可以帮助受众更好地理清人物与空间的交互关系。

小说描绘的是隋唐故事,这一时期的重大历史事件往往集中在长安、洛阳、幽州等北方都会,如李渊父子起兵于太原、安史之乱爆发于范阳、张巡死守睢阳等。同时,书中前半部分出场的豪杰也多为北方人士,故小说总体以北方各州郡为中心,呈现北多南少的特点。此外,书中的地名主要集中于唐代的河东、河南、河北 3 道。其中,河东道下辖的太原府是李渊父子的根据地,聚集了众多英杰,河东道的泽州、潞州也是小说较早出现的地名,重要人物秦琼、单雄信等人早期活动于此;河南道包括今山东大部及河南一部,其下辖的齐州、青州等属现在的山东,早期王伯当、程咬金等英雄人物位于此,导致此道地名较多;与之有别,河北道各州郡出现时间较晚,多在第 14 回后出现,其所属的幽州、贝州、涿州,属反

派窦建德、安禄山等人的势力范围。

3.3 数字人文视角与传统人文视角时空研究对比

与传统人文视角相比,数字人文视角下的《隋唐演义》空间研究具有几个新的特点:其一,可以采集的地理数据更多。传统研究往往只能选取个别地域,如对关陇地区隋唐人物进行研究^[16-17]等。而数字人文研究可借助 GIS 技术及 Python、QGIS 等软件,对书中出现的地名及人物活动地域进行精确的经纬度定位,从而得出更为准确且全面的结论。其二,传统空间研究无法直观还原空间分布特点,而数字技术可以在唐代当时的历史地图上直接实现标注,形成空间及密度分布表,以图表和数据的形式使府、州、郡的区划分布变得一目了然,便于后续研究的展开。其三,传统研究不能做到深入理解书中不同时期的空间分布特点,而通过数字技术可以实现对《隋唐演义》文本的拆分,从而对不同历史时期的府、州、郡分布有更为直观的了解,由此探知书中的编排布局及创作规律,从而更好地分析书中人物与时空分布之间的联系。

4 结语

本研究充分运用数字人文技术及 SNS、GIS 的相关方法论,从非结构化的《隋唐演义》的小说文本中抽取出结构化的人物与府、州、郡实体,进行社交网络及地理空间分布的可视化分析。需要注意的是,本文仅是基于长篇历史小说《隋唐演义》的文本所展开的人物关系挖掘,网络所呈现的人物关系及地域空间与真实的历史可能会有所出入。此外,小说中的一些人物塑造得比较简略,地名存在偏差,因此一定程度上会影响研究的准确性。总之,本研究是将社会网络分析及地理信息系统用于分析古典长篇历史演义小说的实践尝试。通过数据采集、数据清洗、模型构建、数值统计、分析阐释等步骤完成从质化描述到量化分析、传统文学研究到数字人文结合的转变,在一定程度上拓宽了学术视野,为文学研究带来新的思路与方法。期待未来有更多的学者运用先进的数字技术研究传统人文学科,并推动其向前发展。

参 考 文 献

- [1] 王兆鹏,邵大为.数字人文在古代文学研究中的初步实践及学术意义[J].中国社会科学,2020(8):108-129,206-207.
- [2] 赵雪芹,党昭,李天娥.数字人文视角下的档案信息资源开发问题与对策[J].北京档案,2021(1):18-22.
- [3] 薄一航.后数字时代计算思维下的电影研究新范式[J].北京电影学院学报,2023(10):23-32.

- [4] 爱如生中国基本古籍库个人版[EB/OL]. [2023-11-17]. <http://igjk.er07.com/>.
- [5] 鼎秀古籍全文检索平台[EB/OL]. [2023-11-17]. <https://www.ding-xiu.com>.
- [6] 中华经典古籍库[EB/OL]. [2023-11-17]. <http://publish.ancientbooks.cn/docShuju/platformSublibIndex.aspx?libId=6>.
- [7] 搜韵. 唐宋文学编年地图平台[EB/OL]. [2023-11-17]. <https://sou-yun.cn/MPoetLifeMap.aspx>.
- [8] 学术地图发布平台[EB/OL]. [2023-11-17]. <http://amap.zju.edu.cn>.
- [9] 中国历代人物(CBDB)[EB/OL]. [2023-11-17]. <https://www.inindex.com/biog>.
- [10] 赵薇. 社会网络分析与“《大波》三部曲”的人物功能[J]. 山东社会科学, 2018(9): 50-64.
- [11] 徐永明, 唐云芝. 《全元诗》作者地理分布的可视化分析[J]. 浙江大学学报(人文社会科学版), 2019(1): 150-160.
- [12] 孙建旺, 吕学强, 张雷瀚. 基于词典与机器学习的中文微博情感分析研究[J]. 计算机应用与软件, 2014(7): 177-181.
- [13] Even S. Graph Algorithms[M]. Cambridge: Cambridge University Press, 2011: 1.
- [14] 张春红, 于翠波, 朱新宁, 等. 社交网络(SNS)技术基础与开发案例[M]. 北京: 人民邮电出版社, 2012: 42.
- [15] Green D G, Liu J, Abbass H A. Dual-phase evolution[M]. New York: Springer, 2014: 3-40.
- [16] 徐清廉. 论隋唐的门阀之争与关陇集团[J]. 西北大学学报(哲学社会科学版), 1981(1): 91-98, 111.
- [17] 刘彭冰. 关陇集团与关中文化本位研究综述[J]. 西部学刊, 2017(12): 36-38, 63.

Study on the Exploitation and Visualization of Knowledge in Ancient Book Resources from the Perspective of Digital Humanities: Taking *Romance of the Sui and Tang* as an Example

Dong Shouxuan

College of Liberal Arts, Journalism and Communication of Ocean University of China, Qingdao, 266100

Abstract This article, from the perspective of digital humanities, utilizes library collection resources and various electronic databases as the main basis. It employs methods and software such as SNS, Python, Gephi, and QGIS to draw the visualized charts of character relationships and temporal-spatial distribution in the historical fiction *Romance of the Sui and Tang* through a combination of data models and traditional charts. The analysis results contribute to the in-depth exploration of ancient books, enrich the paradigm of humanities research, and provide new ideas for the integration of ancient book resources and the linkage of knowledge data.

Keywords Digital humanities; *Romance of the Sui and Tang*; Character relationship; Temporal-spatial distribution

(上接第 48 页)

A Systematic Review of Research Data Management Capability Maturity Assessment Models

Jin Yegang

Fudan University Library, Shanghai, 200433

Abstract The research data management capability maturity assessment model identifies and improves deficiencies in practices by comparing current research data management practices with best practices. This study aims to describe the current state of development of research data management capability maturity assessment models in the context of “Big Science.” To this end, a systematic mapping study was conducted on a series of high-quality research results published in five related databases. The analysis of 30 research outcomes highlights the deficiencies and challenges in the field. Despite the gradual refinement of the concept of research data management, there is a lack of good consensus on the domain ontology among scholars, and the absence of objective factors in the models leads to insufficient innovation motivation in the field. In addition, there are issues with the unclear role positioning of management institutions in the assessment models and a significant lack of interaction between model construction and entities outside academia. The article hopes to provide new research directions and suggestions for the future study of research data management capability maturity assessment models.

Keywords Research data management; Capability maturity; Maturity assessment; Systematic review; Systematic mapping study