

《方志物产》中物产与人物的社会网络构建与分析

李 娜

(南京林业大学 人文社会科学学院,江苏 南京 210037)

【摘要】物产包括自然产出的植物、动物、矿物和人工制造的货物,是人类赖以生存的物质基础,而人类则是生产、传播、保护和利用物产的主要力量。地方志作为中国特色典藏文献,系统而详细地记述了不同历史时期、不同地域的物产状况,其中蕴含着大量的物产与人物关系信息。作为具有个体互联关系挖掘功能的社会网络分析技术,为地方志文献中物与人之间关系网络构建研究提供了技术可能。本文在定性梳理地方志物产与人物关系的基础上,基于数字人文的研究范式,通过命名实体识别技术构建人名、物产名的智能识别模型,抽取文本中的人物信息以及物产与人物之间的关联关系,进而构建关系网络,从宏观、中观和微观的角度对方志类古籍中物与人的关系进行计量和可视化呈现,取得全新的有价值的发现,验证了从量化史学到数字人文在研究视角与方法方面的日渐成熟和独特魅力。

【关键词】社会网络分析;地方志;知识挖掘;命名实体识别;农史研究

【中图分类号】S-09;K207 **【文献标志码】**A **【文章编号】**1000-4459(2023)02-0041-11

Construction and Analysis of the Social Network between Products and Characters in *Local Chronicles: Produce*

LI Na

(Faculty of Humanities and Social Sciences, Nanjing Forestry University, Nanjing 210037)

Abstract: Property, including naturally produced plants, animals, minerals and man-made goods, is the material basis on which human beings depend, and human beings are the main forces in the production, dissemination, protection and utilization of property. As a collection of documents with Chinese characteristics, local chronicles systematically and in detail record the property conditions of different historical periods and different regions, which contain a large amount of information on the relationship between property and characters. Social network analysis technology, with the function of individual interconnection relationship mining, provides a technical possibility for the construction of relationship network between property and characters in local history literature. On the basis of qualitatively sorting out the relationship between property and characters in the Local Chronicles, this paper constructs an intelligent recognition model of person's names and property names through naming entity recognition technology, extracts the character information in the text and the correlation relationship between property and characters, and then constructs a relationship network, and measures and visualizes the relationship between property and characters in ancient books in the Local Chronicles from the macro, mesoscopic and microscopic perspectives, and makes new valuable discoveries based on the research paradigm of digital humanities. It verifies the increasing maturity and unique charm of research perspectives and methods from quantitative historiography to digital humanities.

【收稿日期】2022-03-20

【基金项目】国家社会科学基金重大项目“方志物产知识库构建及深度利用研究”(18ZDA327);江苏省社会学科基金“宋至民国时期大运河江苏段沿岸方志物产史料的整理与研究”(20LSC005)

【作者简介】李娜(1985—),女,南京林业大学人文社会科学学院讲师、硕士生导师,研究方向为农业史、数字人文。

Key words: social network analysis; local chronicles; knowledge mining; named entity recognition; agricultural history research

作为历史文献之大宗的地方志是按照一定体例全面记载特定时空内的自然、社会、政治、经济、文化等各个方面情况的文献^①,是中国乃至世界的文化遗产宝库,被誉为“一方之全史”^②,具有存史、资治、教化、兴利等重要作用^③。据《中国地方志联合目录》统计,保存至今的宋至民国时期的方志有8264种,11万余卷,约占中国古籍的十分之一。物产是历来地方志必载的内容,详细记载一时一地所出产的植物、动物、矿物等自然物产和人工货物等,是农业史、环境史等研究的重要资料来源。

地方志中记载的物产条目包括物产名称和相应的描述信息。其中,物产名称一般为通用名,也有使用别名的情况;物产描述信息则是从颜色、形状、大小、习性、生活环境、用途、产量、加工利用技术等方面对物产进行描述,不同物产的描述信息详略不一,详细的可达数千字,简略的仅有寥寥数字,甚至没有描述信息的物产也不在少数。在物产描述信息中,通常会通过旁征博引的方式丰富物产信息,使得描述内容更加完整、立体、可信。引用来源主要包括旧志类、著书类、诗词类、语言类、事迹类、背景类等,其中涉及大量人物信息。于是,物产和人物之间便形成了清晰的关联关系,如物产“稷”的描述信息为“有大小二种,视他方味美。图经曰:稷米即稌米也。吕氏春秋云:饭之美者,山阳之稌。高诱曰:关西谓之糜,冀州谓之𦏧。”(康熙四十九年《保德州志》),引用了人物“高诱”对于“稷”在关西和冀州地区的不同称谓(别名)。研究物产与人物之间的引证关系可以为史料查找、物产考证、研究拓展等提供重要的指引。

在人与物的二元关系研究中,基于哲学的视角探讨二者相互关系的成果较为常见,如剖析荀子、萨特等国内外思想与理论中的人与物的交互关系,也有探讨“庄周梦蝶”蕴含的人与物的对立、交往、融合境界^④;也有部分立足现实的实证研究,如通过《西游记》中人与物的交缠关系展现典籍中反映的丰富市民文化,分析消费社会中电视广告符号化呼吁构建更健康的媒介环境和消费环境^⑤。人与物的哲学思考有助于从本质上认识自然、人类、社会,全面建构三者之间的共生、和谐关系。社会网络分析作为关系研究的关键技术之一,理论及实践研究成果较多,其中关于人物关系的成果主要集中在人与人之间一维关系的探索,有面向文学作品中错综复杂的人物关系研究,如以《左传》《红楼梦》《项墨林墓志铭》等文学作品为例,通过抽取、构建作品中的人物关系,揭示社会结构、权利关系等现象^⑥;有揭示历史人物关系的研究,如以明代进士群体、宋代师承关系等为对象构建历史人物关系网络^⑦;也有面向古代书信文献、中文在线百科等数据中人物社交网络模型^⑧。基于人与物一元关系的社会网络研究成果不仅在研究框架和

① 来新夏:《方志学概论》,福建人民出版社,1983年,第1-5页。

② 王晖:《四论方志性质与特征》,《中国地方志》2005年第1期。

③ 来新夏:《中国地方志的史料价值及其利用》,《国家图书馆学刊》2005年第1期。

④ 矫海燕:《荀子论人一物关系》,首都师范大学硕士学位论文,2014年;高湘泽:《萨特后期人学思想中的“人一物”关系理论刍议》,《肇庆学院学报》2003年第4期;苏少伟:《“庄周梦蝶”事件的意义指向——在“人—自然”维度上的考察》,《福建江夏学院学报》2020年第4期。

⑤ 裴亚莉、石燕:《神游与痴醉:〈西游记〉中的人—物关系》,《人文杂志》2020年第3期;赵一欣:《现代广告符码对消费社会中人与物关系的解读》,西北大学硕士学位论文,2011年。

⑥ 许超、陈小荷:《〈左传〉中的春秋社会网络分析》,《南京师范大学文学院学报》2014年第1期;李万康:《董其昌〈项墨林墓志铭卷〉的人物关系与递藏链》,《南京艺术学院学报(美术与设计)》2019年第3期;余朝晖、陈蕾:《致命的三角——〈红楼梦〉主要人物关系解读》,《湖北社会科学》2010年第10期。

⑦ 潘俊:《面向数字人文的人物分布式语义表示研究——基于CBDB数据库和古籍文献》,《图书馆杂志》2020年第8期;潘俊:《探微知著:从知识库和科举文献中构建历史社会网络——以明代进士群体为中心》,《图书馆论坛》2021年第5期;杨海慈、王军:《宋代学术师承知识图谱的构建与可视化》,《数据分析与知识发现》2019年第6期。

⑧ 李惠、侯君明:《古代书信体文献的社交网络模型》,《南京师范大学文学院学报》2018年第3期;林泽斐、欧石燕:《基于在线百科的大规模人物社交网络抽取与分析》,《中国图书馆学报》2019年第6期。

技术支撑上提供了借鉴,也为从一元关系拓展到人与物二元关系研究提供了探索空间。

本文以明至民国时期山西省地方志中记载的物产史料为研究对象,系统梳理出人物与物产在种植、传播与利用等生产生活中的关系特征,从计量史学的视角,利用命名实体识别技术抽取文本中蕴含的人名实体,并基于语义关系构建物产与人物的关联关系,应用社会网络分析技术对物产与人物的关系网络进行可视化展示和结构化分析,探索数字人文技术在《方志物产》深度整理中的适用性和应用前景。

一、物产与人物关系的特征归纳

在《方志物产》的记载中,物产与人物之间的关联关系建构主要依赖物产描述信息中出现的人名的引用关系,通过系统梳理引用的具体内容,从物质、医药、经济、文化等方面归纳物产与人物之间的关系类型,呈现出明显的双向互动特征。

(一)生产生活方面:物质基础

自然界为人类提供了丰富的饮食资源,有五谷作为主食,有水果、蔬菜为辅,有豆类、麻类等榨油,有飞禽走兽鱼虾等提供蛋奶鱼肉,全面供应人们的营养摄入,保障生存的基本需求。如“白扁豆”的描述信息为“有早熟、晚熟二种,均可制豆腐及粉面。晚熟者,颗粒较大,又宜腌菜。”(民国九年《虞乡县志》),不仅可以制作面粉,还可以提供豆腐和腌菜。另外,人们充分利用自然界的原材料,制作大量生产生活用品,建造用于居住的房屋,制作用于出行的车辆和船舶,生产用于耕作的农具,开发方便生活的器皿等。“杨”因“性劲直,易长,叶圆润而尖,枝条短软扬起。”可“造船,并可作室”(清光绪六年《定襄县志》),“榆”可以“作柱梁、农器”(民国二十四年《浮山县志》),“蓖麻”“色青而黑,用以榨油、入药,并调印色、燃灯。”(民国二十三年《屯留县志补记》)可以调色印染以及燃灯照明;“苇草”则“可织席”(光绪六年《崞县志》);“牛”可直接用于“耕田或骑载”(光绪八年《寿阳县志》);“核桃树”因其“木颇类楠木”可用于“制棹椅”(乾隆十年《浮山县志》)。可见,物产提供了丰富的原始资料,人根据具体需求和物产特征,通过一定的设计、开发、利用,满足生产生活所需的物质基础。

(二)治病防病方面:康养保障

为了保障人们的生命健康,自古就有神农尝百草之举。随着人类对自然认识的日益深化,越来越多物产的药用和保健价值被发掘和利用,本草类物产为药用物产之首,动物类和矿物类物产也发挥着积极的药用价值,共同承担着防疫抗疫、防病治病的重要作用。本草类物产如“青蒿”有“治风、治疹、治疥、治瘰、治虚、治烦、治盗汗、开胃、明目、辟邪、杀虫”(乾隆四年《天镇县志》)之多重功效,现代科学提取的青蒿素已成为当今治疗疟疾最有效药物,而“小豆”可“辟瘟”(道光六年《霍州志》),可用于预防时疫;动物类物产如“蛇”可以“治风湿、治疥疮”(乾隆十八年《蒲县志》),“蚯蚓”(乾隆二十五年《赵城县志》)和“蜗牛”(民国二年《岳阳县志》)均可以“治蝎螫”;矿物类物产如“白石英”具有“润肺”(民国十八年《翼城县志》)的功效。一种物产具有一种或者多种药用价值,不同病症需要不同的物产进行防治,植物、动物、矿物等物产按照君臣佐使的原则协同作用,为健康保驾护航。

(三)经济贸易方面:交流媒介

物产作为商业活动的交易对象,很大程度上影响着一时一地的经济状况。除了珍贵的贡品以外,人类的日常活动也带来物产的交流、传播与利用。“棉花”作为大面积种植的经济作物,“河东蒲、解各县,本属产棉区域,而荣河尤以棉为特产,大约无村无种棉之户,有地百亩者,即种棉六七十亩,以是,商人货之四方,岁数以百万斤计。”尤其是“荣人生活专赖产棉,棉花丰收,则衣食俱足,否则,立呈艰困之象,县中风尚,大有宁可不种麦,而不肯不种棉者也。”(民国二十五年《荣河县志》),可见棉成为了当地人民的主要经济来源;“牛肚菘”作为“其植最广”的蔬菜,“恒贩往他邑”(光绪十六年《天镇县志》),通过地区间

贸易带动当地经济发展。除了因贸易产生的物产流动,人的其他活动也影响着物产的传播,主要表现在外交活动、政策引导等方面。《方志物产》山西分卷的记载中,域外作物的引种主要表现在两个方面,一是张骞出使西域带回了胡桃、胡麻、胡荽、苜蓿、葡萄、石榴、蒜等物产,二是胡峤征战回纥带回来了西瓜之种。而物产在本土的推广则记述较多,“开元末,裴士淹使幽冀,至汾州众香寺,得白牡丹,植于长安私第,为都下奇赏。”(雍正十二年《山西通志》),为白牡丹在山西到长安地区的推广传播提供了机会;“明万历年,知州胡楠,刊布蚕经,给散蚕种,教民以蚕”(康熙四十九年《保德州志》),为山西地区蚕桑推广提供了条件;“民国二年,邑人陈鼎铭,由他处买回三十尾,状略相似,而色之鲜艳、鱼之大小,远逊厥初矣。近日,生育繁衍,五色夺目,不减当年。”(民国二十一年《安泽县志》),为五色鱼在山西的养殖创造了机遇;“明万历中,知县新郑王良臣,悯其地瘠寡产,人少物利,乃自其乡携种至,使户植之。初以为烦苦,未几,柿长成林,取材落实,民皆受其利。”(同治十三年《阳城县志》),为柿子在山西省泽州府阳城县的普及奠定了基础;“乡民张士英,自闽中携种,种之。明季,兵燹踵至,民穷财尽,赖此颇有起色。今则邑民大食其利矣。”(乾隆六十年《曲沃县志》),推动了烟草在山西境内的种植。作为客观对象的物产,通过人建构的贸易活动和交流渠道,成为了特定时空下经济交流的使者。

(四) 文化传承方面:内涵载体

在《方志物产》山西分卷的记载中,人们赋予物产的名称来自对自然界的细致观察和深度了解,蕴含着显著的地区、季节、性能、外形等文化特征。一是外观特征,如“鸡冠花”是因为“花形似鸡冠”(道光十六年《繁峙县志》)等;二是物产习性,如“石鸡”是因为“常取碎石英食之”(乾隆四十八年《凤台县志》)等;三是时间特征,如“夏枯草”是因为“冬至生夏至枯”(康熙五十一年《定襄县志》)等;四是地域特征,如“蜀秫”是因为“种来自蜀”(民国六年《洪洞县志》)等;五是功能效用,如“透骨草”是因为“能透骨、除风湿疼痛”(康熙五十六年《解州志》)等;六是声音特征,如“百舌”是因为“能随百鸟之音”(乾隆二十五年《赵城县志》)等;七是人物影响,如“素馨花”是因为“昔刘王侍女,名素馨,冢上生此花”(乾隆二十五年《赵城县志》)而得名。另外,在《方志物产》的记述过程中,通过大量诗词歌赋的引用,一方面实现了对物产的客观描述和说明,另一方面也通过文学与艺术创作的方式,提升物产的文化内涵,使其从一个客观的物产升华为感情的寄托、情绪的发泄、意向的指代等。李义山久居山西,其诗句“池塘渐暖蒲堪结,两两鸳鸯护水纹。”(雍正十二年《山西通志》),运用蒲草可结、鸳鸯戏水的场景,映衬了鸳鸯代表的情感内涵;梅尧臣的诗句“下溉平田几百顷,稻苗杷杷稔曾不枯。”(乾隆四十八年《太原府志》),展现了山西种稻的盛况。还有,人们通过著书立说等方式为物产的传承提供渠道,如李时珍的《本草纲目》、罗愿的《尔雅翼》等,是后人认识历史、还原历史、重现历史的重要依据。

二、数字人文视角下物产与人物研究的技术路径

(一) 基本思路

由于《方志物产》横跨行政区域广、纵伸时间范围长、内容篇幅庞大,进行全部数据处理难度大,为了探讨社会网络分析技术在物产与人物关系研究中的适用性和可行性,本文选取山西一省语料进行实验性研究。山西东依太行、西南临黄河、北跨长城,拥有山地、丘陵、台地、平原等多种类型的复杂地形地貌,被誉为“表里山河”;作为中华民族文明发祥地暨农业开发最早的地区之一,山西物产种类相对丰富,被誉为“华夏文明摇篮”,有“中国古代文化博物馆”之称;《方志物产》山西卷时间上跨越了明至民国期间,长达455年,志书类型涵盖了通志、府志、州志、县志、乡土志、村镇志、山水志、祠庙志、边关志等。基于山西省的研究在区域上和语料上均具有一定的代表性,可以在更多地域、更大规模语料的研究中推广和借鉴。

以《方志物产》山西分卷为研究语料,通过分析文本中人物出现的特征以及物产与人物之间的关联

关系,构建人名的命名实体识别模型,实现人物信息的自动抽取,根据物产与人物之间的引用关系构建关联数据,并基于社会网络分析开展网络建构和结构剖析,归纳在物产的传承过程中,人物有可能发挥的作用。具体研究路径如图1所示。

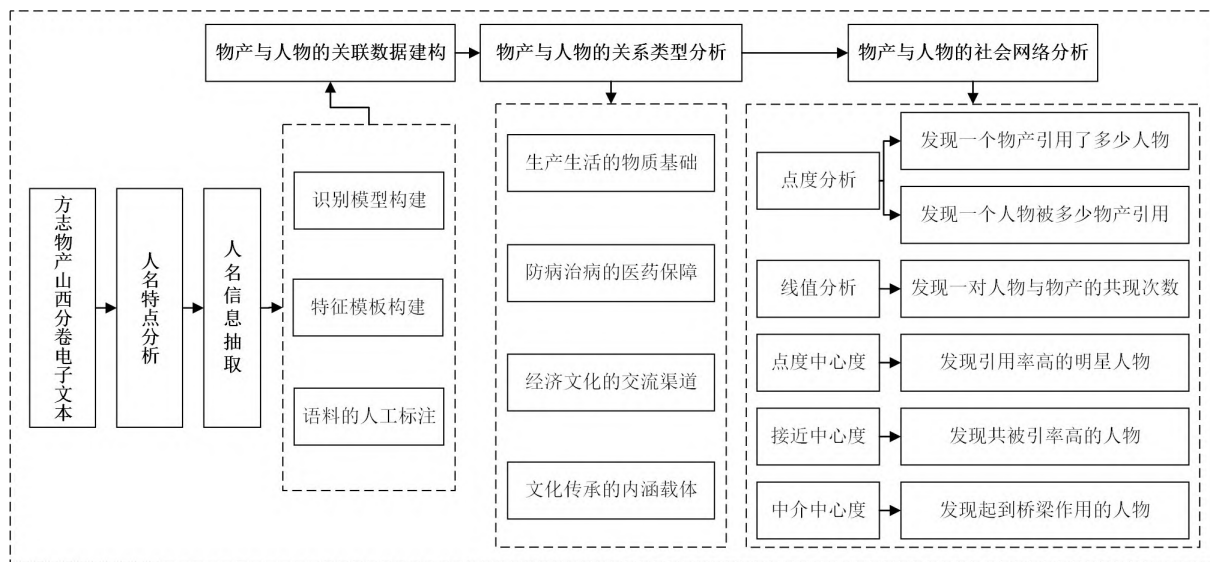


图1 研究技术路线图

《方志物产》山西分卷共13册,约43万字,时间范围从明成化二十一年(1485)至民国二十九年(1940),空间上涵盖了保德、大同、代州、汾州、霍州、绛州、解州、辽州、潞安、宁武、平定、蒲州、沁州、朔平、太原、隰州、忻州、泽州等山西全境的19个府州,志书类型包括通志、府志、州志、县志、乡土志、山川志、偏关志等多种,记载了51545条物产信息。在文本数字化和结构格式化^①的基础上,深入《方志物产》山西分卷文本内容,基于语言结构特征和条件随机场技术,构建人名自动识别模型,实现物产描述信息中人物的自动抽取,并根据物产与人物之间的引用关系,建立物与人之间二元关联数据集,为基于社会网络的可视化分析提供数据源。

(二)人名、物产名等实体抽取

在《方志物产》语料中,人名结构复杂、识别难度大。除了“杜甫”等“姓+名”的标准结构模式,还夹杂着多种类型的表述模式,如“宏景”等名,“太白”等字,“东坡居士”等号,“孔子”等尊称,“唐明皇”等帝王称号,“康节先生”等谥号,“汉高祖”等庙号,“康熙”等年号。另外,人名经常与其他实体混淆出现,如“唐渤海郡王高固”等人名与时间、官职的组合,“河南王孝瑜”等人名与地名的混合,“魏武卫将军奚康生”等人名官职名的混淆,“唐段成式”等人名与时间的混用。为了最大化提升人名识别的效率和精确率,本研究采用了命名实体识别技术中较为成熟的条件随机场模型,结合人工校对的双重方法进行人名信息的抽取^②。

首先,将语料平均随机分成十等份,每次选取其中九份进行人工标注,用“Person”的首字母“P”代表人名,用“【”“】”表示人名的左右边界,标注结果为“【P人名】”,如“脂麻即本草所谓胡,【P沈存中】笔谈云:胡麻即今油麻,古者中国止有大麻,汉【P张骞】出使大宛始得油麻以归,故名胡麻。”基于计算机对人工标注语料特征的学习,构建人名识别模型;剩下一份未标注语料作为训练语料,用于测试模型的识别性能。

① 李娜:《社会网络分析视角下方志古籍知识组织研究——以〈方志物产〉山西分卷为例》,南京农业大学博士学位论文,2017年。

② 李娜:《面向方志类古籍的多类型命名实体联合自动识别模型构建》,《图书馆论坛》2021年第12期。

其次,基于条件随机场模型构建人名自动抽取模型。为了最大化提升模型的识别性能,融入了人名的内、外部特征。其中,内部特征为人名的词长分布,即有多少个汉字组成。根据标注结果统计,最短词长为2,最长词长为10,词长为2和3的人名数量最多,占全部人数的90.35%。外部特征为人名的边界词,即与人名左右相邻两个字的分布。根据标注语料的统计,人名的常用左一元边界词主要有“也、唐、米、宋、明、令、种、汉、传、麻”,常用的右一元边界词主要有“曰、诗、云、使、谓、自、本、赋、征、纪”。

再次,将机器识别的结果与人工标注的结果进行对比,用识全率、识准率、调和平均数三个标志测评识别模型的性能,其中识全率=识别结果中正确的数量/(识别结果中正确的数量+没有识别出的数量),识准率=识别结果中正确的数量/(识别结果中正确的数量+识别结果中错误的数量),调和平均数=(2*识全率*识准率)/(识全率+识准率)。测试结果显示,人名识别的最好识全率为64.82%、最高识准率为75.00%,调和平均数最高为68.78%。虽然由于文本中人名特征复杂以及结构规则性弱等原因,导致面向《方志物产》山西分卷的人名自动抽取模型性能偏低,但后续研究中语料规模将不断扩大,模型识别效果也会随之提升。

最后,对模型识别的结果进行人工校对,保持正确的,矫正错误的,寻找漏掉的,汇总得到人名抽取的最终结果。统计结果显示,含有人物信息的物产语料共624条,共抽取人名846个。其中,有同一个名字重复出现,如“李时珍”出现了132次;也有同一个人的不同称谓,如“苏轼、苏东坡、东坡”。

(三)关系数据构建

如表1所示,为《方志物产》山西分卷中包含人物关系的部分语料,在人名实体识别的结果基础上,根据文本语义关系建立起物产与人物的关联关系,为社会网络模型构建与分析提供数据源。

表1 物产描述信息中的人-物关系语料样例

关系类型	物产名称	描述信息	人-物关系
物质基础	扁豆	释名沿篱豆,宏景曰:种之篱垣,其荚蒸食甚美。	扁豆-陶弘景
物质基础	白菜	一名菘,诸四时皆可常食。南史周顒曰:春初早韭,秋末晚菘,即白菜也。	白菜-周顒
康养保障	苍术	本草:术名山蓟,李时珍曰:异术言术者山之精也,服之令人长生。	苍术-李时珍
康养保障	艾	释名冰台,王安石字说云:艾可义疾,历久而弥善,故字从义。	艾-王安石
流通渠道	芫荽	香菜也,根软而白,绥绥多须,故谓之荽,张骞使西域得胡荽种之归。	芫荽-张骞
流通渠道	烟草	沃旧无此种,乡民张士英自闽中带来,明季兵燹踵至,民穷财尽,赖此颇有起色。	烟草-张士英
内涵载体	紫薇	花对紫薇郎之句,杜牧之诗曰:晓凝瑞露一枝新,不占园林最上春。桃李无言又何在,向风偏笑艳阳人。	紫薇-杜牧
内涵载体	鸳鸯	出安邑,古今注:鸳鸯,水鸟,似凫,毛有文采。本草:交颈而卧,其交不再,雌雄未尝相离。李义山诗:池塘渐暖蒲堪结,两两鸳鸯护水纹。义山久居晋,故引之。	鸳鸯-李商隐

识别结果中的824个人名包括了重复出现和同人异名的情况,对应的624个物产名中蕴含着同物异名和同名异物的情况,这将会导致物与人关系数据的冗余。为了保证人名和物产名的唯一性、完整性和准确性,本研究需要针对人名和物产名进行数据清洗工作。首先,合并同类项,即同人异名、同物异名的名称统一化处理,将同一个人或者物产的不同表述统一成规范名称,如“白居易、白文公、白香山”统一合并为“白居易”,“菠菜、菠薐、波菜、波菱”统一成“菠菜”;其次,拆分异类项,即同名异物的名称区别化处理,如物产名称“杜鹃”,根据分类信息和物产描述,可以区分为“杜鹃鸟”和“杜鹃花”两类不同的物产;再次,去除重复人名和物产名,筛选出不重复的人名和物产名,并统计每个人名和物产名重复出现的次数作为属性数据记录备用;继而,转换繁简字体,有利于阅读、理解和传播;最后,根据物产与人物之间的引用关系,构建物与人之间二元数据的关联数据集,如表1中第三列所示的人-物关系数据样例,并以此为

基础,继续开展方志物产中物与人之间的关系研究。

三、物产与人物的社会网络构建与分析

基于《方志物产》山西分卷,经过对人名和物产名的清洗处理,本文共得到了290个物产名和275个人名,以及二者之间的523对关联关系。其中,物与人的关联关系可以分为几种不同的类型,有一对一的关系,即有一个物产仅与一个人物相关;有一对多的关系,即一个物产与多个人物或者一个人物与多个物产相关;也有多对多的关系,即多个物产和多个人物交叉相关。由此,得到物产与人物之间的社会网络整体结构图。根据连通性计算,可以将整体结构图分成6种规模大小的模块,最大的模块包括343个顶点,其余均为较小规模,分别含2、3、4、5、6、7个顶点。为了增加图像显示效果的清晰度,本研究将不同大小的模块分别选择了一个作为案例,结果见图2。其中,方框代表人物实体,圆点代表物产实体,线代表物产和人物之间的关系。而图上部绿色圈内的绿色顶点“白菜-周颢”等一个圆点与一个方框相连的图形即为一对一的关系类型,图左上部蓝色圈内的蓝色顶点“王登汉-蚕-郑书忠”等一个圆点与多个方框或者图左下部紫色圈内的紫色顶点“天花-玄烨-娑罗树”等一个方框与多个圆点相连的图形即为一对多的关系类型,图中黄色顶点组成的最大模块凸显的则是多对多的关系类型。

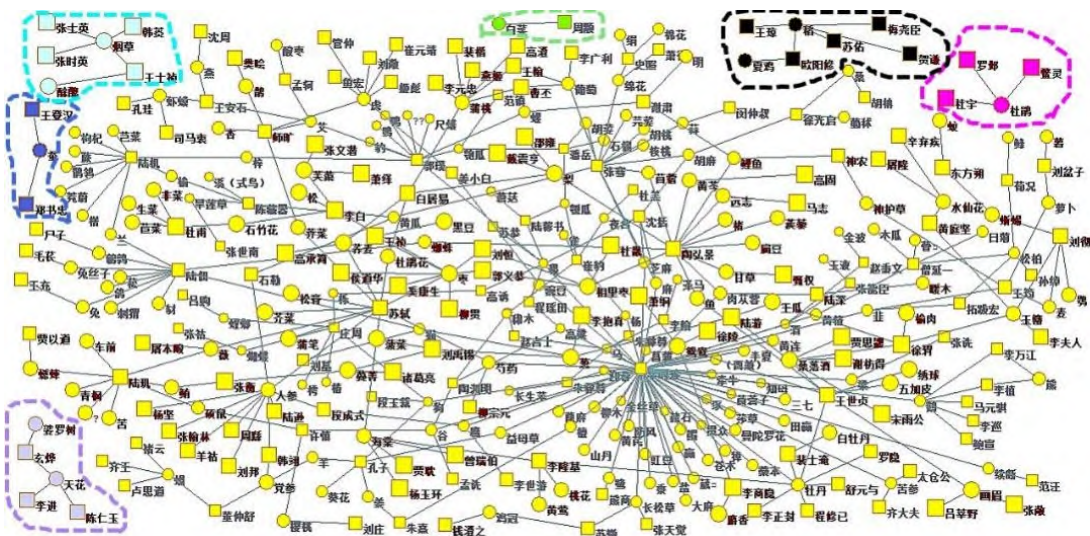


图2 物与人之间的整体社会关系网络图(局部)

基于物产-人物关系网络,从整体和局部等视角展开结构性分析,通过点度中心度厘清物产的受关注度,人物的被引率,物产与人物之间的关系紧密程度以及人物在物产传承中的作用强弱等,为相关领域研究提供资料查找路径以及量化数据支撑。

(一)物产与人物社会网络的整体分析

1.点度分析。点度是指定顶点与其他顶点之间的所有连线数量。物产顶点的点度,表示一个物产的描述中引用了多少位人物,反映出该物产对人物的引用广度;而人物顶点的点度,表示一个人物被多少个物产引用,反映出该人物被物产引用的频率。经过计算,物产和人物顶点的点度分布如表2所示。

表2 物产、人物顶点的点度分布统计表

物产点度大小	同点度的物产数量	人物顶点大小	同点度的人数数量
1	169	1	194
2	64	2	47

续表2

3	34	3	10
4	11	4	10
5	4	5	4
6	2	6	1
7	3	7	1
8	2	8	1
10	1	9	1
		10	2
		12	1
		13	2
		56	1

物产顶点的点度共有九个类别,其中最小点度为1,共有169个物产在物产描述信息中仅引用了1位相关人物;最大点度为10,仅有1个物产在物产描述信息中引用了10个相关人物,发现排名前三的明星物产分别为人参、枣和牡丹。为了进一步了解每个物产具体引用了哪些人物,本文以点度为10的物产“人参”为例,抽取了其个人中心网络,即该物产和与其直接相连的人物组成的子网络,见图3(左)。可以发现,其引用的相关人物为苏轼、韩翃、杨坚、羊祜、陆逊、段成式、石勒、刘邦、张翰林、周繇。在人物的引用形式和内容上,一是引用人物的诗词作品,如苏东坡的《紫团参》、韩翃的《送客之潞府》、周繇的《以人参遗段成式》;二是引用诗词作品中出现的人物,如段成式、羊祜、陆逊就是周繇的《以人参遗段成式》中出现的人物;三是引用时间定位人物,如杨坚、石勒、刘邦等均为皇帝,主要是作为时间背景,记录其在位期间,关于人参的种植和利用信息;四是引用人物的言论,如张翰林是明嘉靖十一年(1532年)翰林官员名为张春,此处引用的是他所谓“人参数量少不易得”的表述。

而人物顶点的点度共分13个类别,最大点度为56,仅有1个人物同时被56个物产引用;最小点度为1,有194个人物仅被一个物产引用,被引频次大于10的明星人物依次是李时珍、陶弘景、张騫和苏轼,而李时珍因其名著《本草纲目》而高居被引榜首。本文以点度为12的人物顶点“苏轼”为例,抽取其个人中心网络,见图3(右)。可以发现,引用其的物产有“菠菜、杜鹃花、海棠、芥菜、蒲笔、荠菜、荞麦、人参、芍药、松膏、薇、枣”,多是在物产的描述信息中引用了苏轼的诗词作品,如在物产“芥菜”的描述信息中就引用了苏东坡《雨后行菜圃》中“芥蓝如菌蕈,脆美牙颊响”的诗句,用于描述芥菜的外形和口感。

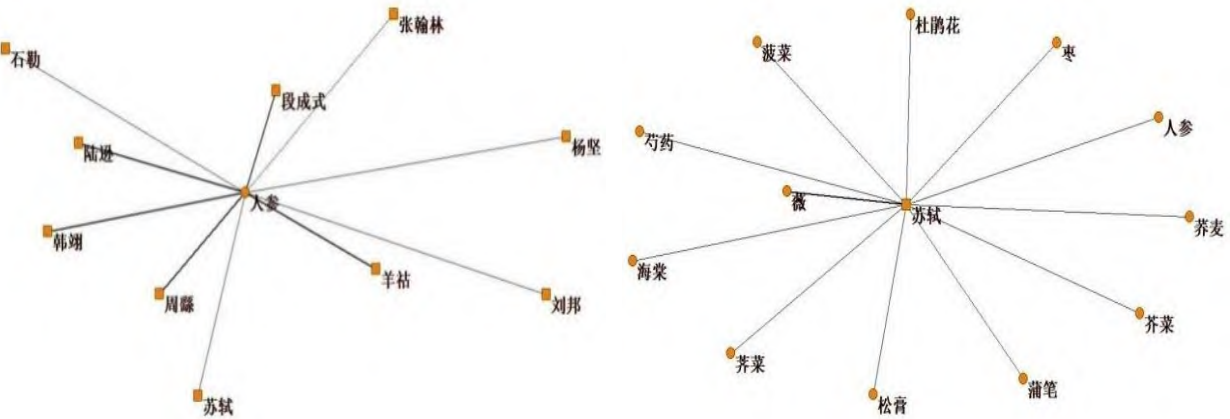


图3 物产顶点“人参”和人物顶点“苏轼”的个人中心网络图

2. 线值分析。线值表示两个顶点共同出现的次数,反映了顶点之间关系的强弱,线值越大,关系越

强,线值越小,关系越弱。在物产与人物的社会网络中,线值表示在《方志物产》山西分卷中某物产引用某人物的次数。统计结果显示,物产与人物网络中,线值共有13个类别,最大值为31,仅有一对物产和人物之间的引用次数达到了31;最小值为1,有373对物产与人物之间的引用次数为1,约占所有关系数据的71.32%。其中,最大线值位于物产“黍”与人物“李时珍”的连线上,主要引用了李时珍在《本草纲目》中关于“黍、稷”二者关系的记载,即“苗穗与稷同,有黄、白、赤、黑四种,米皆黄,俗呼为黄米。李时珍曰:稷黍一类二种,粘者为黍,不粘者为稷,今俗通呼为黍子。”其次是线值为21和20的物产“穀”与人物“孟诜”以及“李时珍”之间的关系,主要涉及物产名称变化以及人们认识不清的内容,李时珍记载的谷名称变化“今之粟,在古但呼为粱,自汉以后,始以穗大而毛长粒粗者为粱,穗小而毛短粒细者为粟。”与孟诜记载的人们对谷的认知情况“人不识粟,而近世皆不识粱也”相互呼应。

(二)人物共现网络分析

物产与人物的社会关系网络中包含两个物产和人物两类顶点集合,属于2-模网络。人物共现是指若两个人物被同一个物产引用过,那么这两个人物之间则具有共现关系。为了深度分析人物在物产传承中的作用,本文对物产与人物构成的2-模社会网络进行降维处理,从整体网络中抽取了因物产引用而产生的人物共现网络,即275个人物及其之间连线的集合。经过聚类计算,图形可以分为6种不同规模的子网络,其中最大子网包含151个顶点,其余子网含1、2、3、4、5个顶点不等。每种类型的子网络选取一个显示(见图4)。

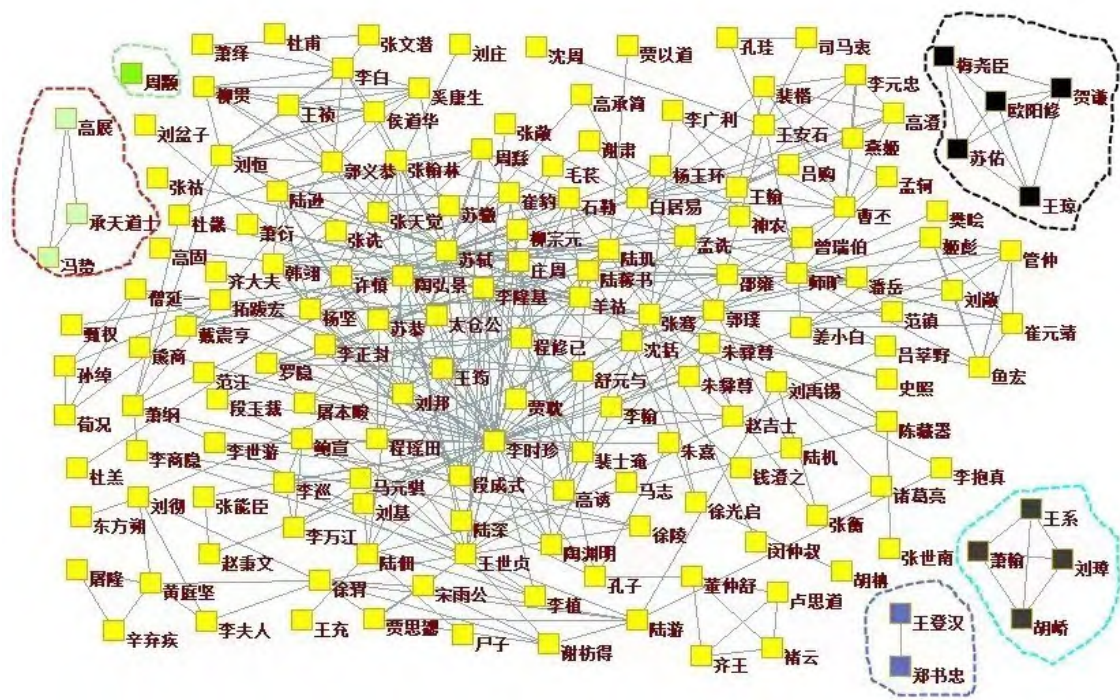


图4 人物共现关系网络图(局部)

图4中的聚类大致可以分为三种,一是左上角的一个较大规模且密集相连的子网,即最大连通子网络,即由黄色顶点组成的部分,包含了151个人物和778条连线,蕴含着丰富的人物关系;二是周边部分小规模连通子网络,即黑、红、深蓝、浅蓝色圈内的顶点组成的部分,包含的人物为2个至5个不等,这是由于物产与人物之间一对多的关系形成的;三是上部绿色圈内的孤点代表的部分,说明在物产与人物的社会关系网络中,这些人物与物产之间存在一对一的关系,未与其他人物产生共现现象。这些孤点和零散子网可以在图中直接观察到结果,没有必要单独进行结构分析;但左上角的最大联通子网络包含

的人物数量多,无法直观分析,需要进行深度结构剖析,以挖掘一些关键人物。

1.通过加权点度中心度挖掘网络中的核心人物。加权点度中心度同时考虑了节点的点度和与其相关的线值,综合反映了人物之间关联的强度。其中,通过加权点度中心度大小发现,最高值为815,节点为“李时珍”,为明代著名医药学家,著有《本草纲目》等,是整个网络中影响力最大的节点人物;次高值为567,节点为“孟诜”,为唐朝医学家和食疗学家,著有《食疗本草》等;再次高值为217,节点为“许慎”,为东汉时期经学家和文学家,著有《说文解字》。而通过人物之间连线的线值发现,与李时珍联系最密切的五个人物为孟诜(420)、许慎(100)、张骞(89)、沈括(75)、陶渊明(43);他们之间也存在者较强的共现关系,如孟诜-许慎(105)、孟诜-陶渊明(42)、许慎-陶渊明(10)、张骞-沈括(51),括号中的数字为两个人物的共被引次数。

2.通过中介中心度探索网络中的中介人物。中介中心度反映的是人物在网络中起到的桥梁或者中介作用,能够控制或者影响别人之间的联系。中介人物是一个网络中的关键节点,承担着连通网络的重任。往往是因为这个中介顶点的存在,整个网络或者两个凝聚团体才得以连通,如果这个中介顶点不存在,那么两个凝聚的团体或者整个网络都无法连通。在物产与人物的关系网络中,可以通过中介人物查找更多可能相关的物产资料线索,如诗词歌赋、名人笔记等。结果显示,人物的中介中心度较高的人物有李时珍(0.631633)、苏轼(0.263126)、王世贞(0.141344)、陶弘景(0.134880)、张骞(0.119726),括号中的数字为人物的中介中心度,说明他们与其他人物共同被引用的频率高。

3.通过接近中心度寻找网络中的最佳信息传播者。接近中心度代表顶点拥有较短的路径,可以快速联系上他人,往往能够成为一个网络中信息传播的关键人物。在物产与人物的关系网络中,通过接近中心度可以找到物产相关人物中的核心团体,有可能成为资料查找的辐射源。结果显示,151个人物顶点中,接近中心度排序前十的人物有李时珍(0.545455)、苏轼(0.454545)、陶弘景(0.429799)、李隆基(0.412088)、王世贞(0.412088)、张骞(0.406504)、段成式(0.398936)、郭璞(0.387597)、柳宗元(0.387597)、沈括(0.384615),括号中的数字为人物的接近中心度。

可见,在明至民国时期的《方志物产》山西分卷的语料中,经常引用的是相关人物的话语、事迹、著作、诗词歌赋等,以丰富物产描述信息。而无论从被引数据还是共被引数据分析,李时珍都是被引网络中当之无愧的核心人物,其代表作《本草纲目》为本草学之集大成者,被誉为“古代中国百科全书”,其关于物产的记载内容在《方志物产》的相关物产描述中被引频率遥遥领先。

(三)社会网络分析在《方志物产》整理中的应用价值

在物产与人物关系构成的社会网络当中,不同时空下的物产与历史人物之间存在着错综复杂的引用关系,单纯的人工阅读或者整理难以厘清。社会网络分析技术提供了一种分析复杂网络的方法,可以用量化的方式实现从整体、局部等不同的视角进行网络剖析,实现挖掘网络中蕴含的更多隐性知识,从而为物产或者人物相关研究提供研究支持。主要表现在以下三个方面:

1.运用整体网视角从宏观上把握物产与人物关系的整体概况。本研究利用点度和线值两个结构指标对物产-人物网络进行整体量化分析,通过物产引用多少个人物反映物产记载的丰富程度,通过人物被多少物产引用反映人物对物产的关注广度,通过物产与人物之间的线值大小反映物产与人物的强弱关系,让读者能够快速了解物产-人物关系网络的概况,迅速找到网络中的重要节点以及联系密切的人-物组合等。

2.运用局部网视角从细节上展现物产与人物关系的具体情况。本文通过网络降模处理,运用联通子网和个人中心网络两个结构指标进行局部网分析。其中,降模处理将复杂的2-模网络转换成简单的1-模网络,有利于进行单集合关系数据的分析,可以实现人物的共被引网络分析。通过发现哪些人物

被同一个物产引用了多少次,反映出物产的受关注程度;通过提取联通子网,尤其是拥有最多人物、关系最为密切的最大联通子网络,直观展现了人物因被物产引用而形成间接关系网络;通过抽取个人中心网络能够清晰地展示任何一个指定物产或者人物的个体信息,如一个具体物产在记述中引用了哪些人物以及每个人物的引用次数等,可以为相关研究提供较为明确的信息检索指引。

3. 基于网络结构开展隐性知识的挖掘。本文主要是通过点度中心度、中介中心度、接近中心度进行核心人物、中介人物、传播人物的挖掘,通过分析不同的人物在网络承担的角色及其重要程度,为研究者提供了信息关联检索的参考,如研究物产的传播或传承过程中需要找寻关键人物线索等。这些关键人物的发现,是在记载频次、关系强弱等基础上综合计算的结果,在一定程度上突破了人工整理资料方式的限制。

结 语

本文立足中国特色地方志文献,引入命名实体识别和社会网络分析等数字人文技术,以《方志物产》山西分卷为例挖掘了其文本中蕴含的物产与人物之间的关系。研究显示,第一,人与物产之间产生了密不可分的关系,物产为人提供饮食、医药、经济等物质基础,人为物产提供传播、利用、文化等渠道支撑。第二,《方志物产》山西分卷中物产引用的人物类型主要分为以下几个方面:一是以医药学家为代表,引用了《本草纲目》《食疗本草》《尔雅》等著作;二是以文学家为代表,引用了苏轼、白居易、罗隐等的诗词作品;三是以政治家为代表,引用了唐明皇为黄鹂命名、汉武帝推广种宿麦等事件;四是以外交家为代表,引用了张骞和萧翰出使西域、胡峤征回纥等活动;五是以普通人为代表,引用了陈鼎铭从外地购回五色鱼、张士英自关中引种烟草等行为。第三,基于数字人文研究方法与传统研究获得的结论基本吻合,但是跨时空的、可视化的、量化的表达与呈现,彰显了数字史学的独特魅力,验证了数字人文方法在《方志物产》内容整理中的适应性和良好前景。

数字人文技术的应用,在研究语料规模、隐含知识挖掘、结果可视化呈现等方面实现了有别于传统研究的突破。山西一省的数据规模虽然偏小,但本文从技术路线和实验开展方面验证了社会网络分析技术在《方志物产》内容整理中的可行性。《方志物产》全文所载资料数据规模庞大,时间跨越了900余年,内容多达3000多万字,物产信息逾150万条,包含了复杂的显性和隐性关系。随着语料规模的不断扩大、数据挖掘与可视化技术的逐步优化,未来研究范围将从单一省份推广到一个地区乃至全国,拓展物产、人物、时间、地点、事件、功效等多种实体之间的关系类型,智能揭示蕴含在传统地方志文献中的错综复杂关系,充分发挥中国特色典籍的时代价值。在大数据发展趋势下,人工辅助依赖的减弱,研究宽度和深度的拓展,可视化效果和分析精度的有效提升,均有效凸显了数字人文研究的优越性。除了探索技术方法的适用性外,后续研究还需充分重视语料内容的考证,确认人、物、地、时、事等元素的匹配,以增加研究的厚度和信度。

(责任编辑:徐定懿)