

中国常见神话人物起源母题数据集

ISSN 2096-2223

CN 11-6035/N



文献 CSTR:

32001.14.11-6035.csd.2023.0039.zh



文献 DOI:

10.11922/11-6035.csd.2023.0039.zh

数据 DOI:

10.57760/sciencedb.07047

文献分类: 社会科学

收稿日期: 2023-01-16

开放同评: 2023-02-07

录用日期: 2023-03-20

发表日期: 2023-04-20

* 论文通信作者

王京: jwang@cnic.cn

王京^{1*}, 熊惠², 其乐格乐², 杜一¹, 周园春¹

1. 中国科学院计算机网络信息中心, 北京 100190

2. 中国社会科学院大学, 北京 102488

摘要: 神话是中华民族传统文化的重要载体。各民族神话数量众多, 内容丰富, 形式多样, 许多神话人物成为人类历史与中华文明进程的珍贵记忆。以作为神话主体的神话人物为切入点, 把相关神话文本与叙事作为人文社科领域的研究专题和重要数据来源, 在推动人文学科大数据建设、深化中华优秀传统文化数据分析、创新人文社科研究新方法等方面具有重要文化意义和学术价值。本数据集以中国常见的 50 个神话人物的起源母题为研究对象, 采集梳理了的 1620 条常见神话叙事有效数据, 包含母题、实例、民族、流传地和文献出处等标准字段, 同时对数据来源、数据采集方法、数据的分类和数据的结构等进行了系统的分析和阐释, 并对数据样本进行了详细描述, 在列举相关数据集样例的基础上, 对包括人物产生维度、民族维度和地理维度等进行了统计分析。最后, 提出数据集在中国各民族文化内涵研究、跨地区文化研究和中华优秀传统文化分析方法等方面的建议。

关键词: 人文数据; 神话人物; 神话元数据; 神话数据分析

数据库(集)基本信息简介

数据库(集)名称	中国常见神话人物起源母题数据集
数据通信作者	王京(jwang@cnic.cn)
数据作者	王京、熊惠、其乐格乐、杜一、周园春
数据时间范围	上限约为公元前 700 年, 下限为 2022 年
地理区域	中国
数据量	648.98 KB
数据格式	*.xlsx
数据服务系统网址	https://doi.org/10.57760/sciencedb.07047
基金项目	中国科学院网信专项(CAS-WX2021SF-0308); 国家社会科学基金青年项目(21CZW058); 中国博士后面基金项目(2020M680682)
数据库(集)组成	数据集包含中国 50 个常见神话人物在多民族、多区域内的多层次起源叙事母题, 涵盖人物产生的原因、方式、时间、地点等几大类型, 其中“人物的产生方式”类下又划分为自然产生、婚生、生育、感生等 8 个次级类型, 共计 1620 条数据, 每条数据包含 10 个字段信息。数据集同时包含一张基于上述数据的神话人物关系图谱, 该图谱基于 SPO 三元组构建, 包含 369 个节点, 1240 条边。

引言

中国神话起源于史前文明时期，是中华民族文明进程中的珍贵历史记忆和重要文化载体。对其进行全面的数据采集与数据研究，对继承和发展中华优秀传统文化具有重要意义。神话叙事的产生、发展与结局，都会围绕特定的神话人物展开，其中以“三皇五帝”为代表的众多神话人物叙事源远流长，内容丰富，记录了炎黄子孙性格塑成的历史渊源，传承着中华民族文化精神。特别是大量存世于古代文献、民间典籍和口头传承中与“神话人物起源”相关的神话叙事，深刻揭示了不同历史阶段、不同地区、不同民族对人类产生的积极思考、对中华文化祖先身份的高度认同、对人类智慧的全面发掘以及对人生观、价值观、世界观的深刻思考。从大规模文献中梳理出该主题的标准叙事文本数据，对进一步丰富民间文学视域下的中国话语权，以及深入挖掘中华优秀传统文化的内在精神，均具有重要的价值和意义。

目前，以特色文化资源为代表的人文社科数据在国际上同样受到学者的广泛关注，在相关数据支撑下，在诸多领域取得了杰出成果。如谷歌团队基于百万电子化书籍语言数据，完成了对 1800–2000 年间全球英文语言中反映的语言学和文化现象的量化分析和全多维探索^[1]。再如利用超过 15 万的名人的出生地点和死亡地点数据，建立跨度 2000 年的迁徙网络，量化分析欧洲文化中心变迁^[2]，等等。此外，国内外人文专题数据库的建设也实现了短时期内的爆发式增长，从数据对象的涉及广度方面分为以下几种形态：一是，以历史人物或史料典籍为核心的历史类数据库，如哈佛大学联合台湾中央研究院和北京大学研发的中国历代人物数据库^[3]（China Biographical Database, CBDB），其以关系型数据库为主要形式，收录了 52 万历史人物传记信息，这些数据在提供资料查阅的基础上，可广泛应用于统计、社交网络和空间分析；再如，ASP 世界历史人物索引数据库（Alexander Street Press, ASP）^[4]收录了超 70 万页世界范围内历史事件发生时所涉及人物的信件、日记、口述史及个人叙述等信息，可作为查找、探索与分析人类经验的起点，从而对相关历史事件进行溯源。二是，人文细分领域专题数据库，国外如芬兰土尔库大学口传文学档案馆（The TKU Archive at the University of Turku）对田野工作及研究环节中的资料数据进行整理并形成了口传文学数据集，为民间文学领域学者提供了相应的数据分析模块；再如，国内以国学经典、方志和文学作品为数据对象的细分领域数据库，代表性的有中国古典文献资源数据库^[5]、中国方志库^[6]、明清实录数据库^[7]，等等。此外，还有一些开源的数据资源项目如 Chinese-poetry，项目收录了唐宋两朝近 1.4 万名古诗人的近 5.5 万首唐诗和 26 万首宋诗，目前在 github 开放^[8]。三是，以特定研究目的为导向的更加细分领域数据库，如国研网“一带一路”战略支撑平台^[9]等，收录了与“一带一路”政策相关的项目、报告、信息等，有效支撑起特定领域方向的专门研究。

上述人文领域数据库特别是国内数据资源平台，在差异化目标和需求驱动下，大多处于分散自建的状态^[10]，在元数据完整度和涉及广度方面难以形成统一高标准，特别是针对文本知识维度方面的挖掘与关注远远落后于实际研究需要。事实上，特定人文叙事的母题、地点、时间、民族、文献出处等元数据对包括民俗学、文学、民间文学等在内的诸多具备明显交叉学科倾向的人文领域意义重大，特别是在文本溯源、跨领域对比研究和文化区域研究等方面具有重要作用。本数据集不仅共享了基本神话文本实例，还通过一定规模的人工抽取、校验等工作，完成了实例对应的民族、地点、出处、母题等信息的规范与补全，保证了数据的高价值与高可用。

中国神话人物起源母题神话叙事既是中华优秀传统文化专题大数据建设的案例性尝试，也是人文社

会科学大数据开发与利用的前沿探索。本文以 50 个常见神话人物起源叙事为研究对象,积极组织项目成员广泛查阅并收集散轶在经典文献、民间口头传统、文物、民俗中的相关神话,涉及各民族神话资料集成、少数民族文学史、特定民族田野调查资料集等资源中的有效神话资源,并通过叙事学、母题学和类型学研究方法完成了对神话人物起源母题信息的抽取与重构,同时完成了实例对应地点、民族、出处信息的人工补全,为相关研究领域提供了较为详尽的一手资料,为面向人文领域的知识开放共享珍贵素材和数据基础。

1 数据采集和处理方法

本文从《山海经》《楚辞》《淮南子》和《穆天子传》等经典古籍文献中梳理出出现较早、出现频率较高、关涉神话叙事较多的 50 个常见神话人物作为数据集的基本研究对象。以上述人物名称为检索关键词,进一步扩大文本检索范围,从相当规模的中国多民族神话文本中,依循叙事规则对目标神话人物“起源”相关叙事实例进行摘录,形成基本实例数据(字段)。然后,依照神话母题学^[1]理论方法对实例描述进行归纳演绎,确定一级母题、二级母题、三级母题、四级母题字段。最后,通过上下文和文本标注等信息,补全民族、流传地、文献出处等其他字段。其中,民族划定依据参照斯大林《马克思主义和民族问题》^[12]等标准规范确定,流传地信息根据国家标准《地名分类与类别代码编制规则》(GB/T 18521—2001)^[13]等确定。主要技术流程如图 1 所示。

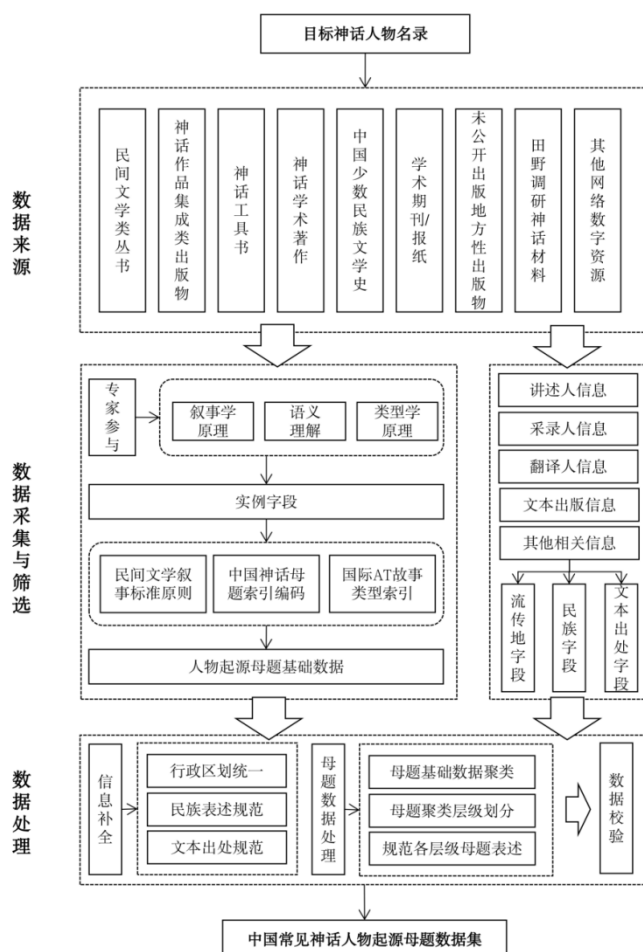


图 1 数据集生成技术流程图

Figure 1 Flowchart of the dataset generation technology

1.1 数据来源

本数据集的数据主要来源于 9 大类资料，详见表 1。

表 1 数据来源信息表

Table 1 Table of data sources

序号	数据源类型	举例
1	公开出版民间文学类丛书	中国民间文学集成全国编辑委员会编《中国民间故事集成》 ^[14] (中国 ISBN 中心出版, 各省卷本)
2	神话作品结集/集成类出版物	陶阳、钟秀编《中国神话》 ^[15] (商务印书馆 2008 年版)
3	与神话有关的工具书	吕大吉、何耀华主编《中国各民族原始宗教资料集成》 ^[16] (中国社会科学出版社, 分民族卷本)
4	与神话有关的学术著作	马昌仪编《中国神话学文论选萃》 ^[17] (中国广播电视出版社 1994 年版)
5	中国少数民族文学史	马学良、梁庭望、张公瑾主编《中国少数民族文学史》 ^[18] (中央民族大学出版社 2001 年版)
6	学术期刊、报纸	《民族文学研究》《民间文化论坛》等期刊中的论文
7	未公开出版地方性出版物	各省(市、州、县、区)三套集成办公室或领导小组收集整理的《中国民间故事集成》(县、市、区卷本)
8	项目组田野调研搜集的神话材料	项目组在民族地区采集的目标神话人物起源传说等
9	其他网络数字信息	百度百科、维基百科和中国民族文学网等

除此之外, 本数据集还收录了一些濒危珍稀的抄本, 大多为项目组成员深入偏远少数民族地区收集的一手资料。如, 胡崇峻据秭归县新滩镇《黑暗传》清末抄本而记载的《人祖创世》, 阳日镇龙溪村村民唐运青收集的祝天照抄本, 以及林区下谷土家族民族乡石磨村戊子垭、刘茂森处摘录转抄的《黑暗大盘头》《黑暗纲鉴》片段, 等等。上述资料作为民间文学领域珍贵的非物质文化遗产, 一定程度拓宽了数据来源和口径, 进一步提升了数据集的特色性、珍稀性和完备性。

1.2 数据采集与筛选

常见神话人物起源母题数据的采集与筛选主要依循“常见神话人物名称对齐—神话人物起源数据采集—其他相关数据采集”的基本路径展开。

鉴于叙事文学类数据的特殊性, 在数据采集过程中要充分考虑同一神话人物存在多种称谓、别称、异字和名号的现象, 同时, 在不同民族地区、不同年代印本和不同传播载体中对同一神话人物的名称载录也不尽相同。因此, 本数据集在具体数据采集之前, 就同一神话人物存在多种称谓、别称、异字和名号的问题进行过多次研讨, 通过专家经验和人工核验的方式完成了人物名称的规范和

对齐。数据集神话人物的名号、称谓、别称大多出自神话原文，可以通过同一文本上下文关联的方式取得映射。少数难以从同一文本得出结果的情况，项目组采取“多民族神话文本互照”的方式进行了印证，最终形成目标神话人物名称映射表，详见表 2。

表 2 常见神话人物名称映射表

Table 2 Mapping table of the names of 50 Chinese mythological characters

序号	人物名称	名号/称谓/别称
1	布洛陀	布洛朵；抱洛朵；保洛陀；黼洛陀；布碌陀；布罗托；陆陀公公
2	仓颉	先师
3	嫦娥	姮娥；药奶奶；宓妃；常仪；常羲；虚上夫人；素娥；常娥；嫦娥仙子
4	蚩尤	-
5	大禹	禹；夏禹；帝禹；神禹；夏后禹；文命；戎禹；大禹王；高密；夏伯
6	丁巴什罗	东巴教教主
7	二郎神	司门天；罗和二王；杨戩；二郎圣君；额勒勒大腾格里
8	风后	金川星
9	风神	云中君
10	伏羲	牺；必羲；宓羲；庖牺；包牺；炮牺；牺皇；皇羲；太昊；太皞；瓠系；虑戏；虑羲；伏羲；伏依；伏希；伏戏；伏牛羲；人祖爷
11	高辛	帝啻；姬俊；帝俊；帝啻王；高辛氏
12	格萨尔	觉如；嘎莎
13	观音	度母；观音菩萨；观音老母；千手观音；南海观音菩萨；南海观音；观音大士；自在观音
14	螭	大司空；崇伯；崇伯螭；白马；玄武大帝；荡魔天尊；真武大帝
15	洪钧	洪钧老祖；洪君；洪兴祖

序号	人物名称	名号/称谓/别称
16	后稷	弃；姬弃；稷神；司穡
17	后土	句龙；土正；地母娘娘
18	后羿	夷羿；仁羿
19	黄帝	轩辕；轩辕氏；公孙轩辕；姬轩辕；帝鸿；帝鸿氏；帝俊；帝喾；有熊氏；帝江；帝轩氏；玄元黄帝；元圃真人轩辕黄帝；仓枢纽；乔山熊；黄龙公；黄龙君
20	精卫	女娃
21	九隆	-
22	雷神	雷公；雷郎公公；雷神爷；雷公爷；雷婆；雷婆萨岂；玛由梭；雷神妈妈；雷神闻仲；雷将陆盟；雷王；雷兵契高；雷公爷梭；果索；阿路狄达力；阿克的恩都力；查和尔干腾格里；阿霏刹
23	嫫祖	祖神；嫫祖娘娘；蚕娘；先蚕
24	廌君	务相
25	妈祖	林默；灵惠夫人；护海女神；湄洲圣母；天妃娘娘；海神娘娘
26	密洛陀	神母密洛陀；万物之母
27	莫一大王	莫龙恭
28	牛郎	-
29	女娲	女娲氏；娲皇；娲皇圣母；女娲娘娘；娲皇奶奶；人祖奶；人祖婆婆；娲儿公主
30	盘古	盘古王；盘果王；盘古爷；盘古真人；盘古大王；扁古；盘神；盘古公公；盘老大；盘颇
31	盘瓠	盘王；槃瓠；盘瓠王；盘护；般瓠；盘皇
32	七仙女	玛尼；日吉纳；七姑星；七衣仙女；七仙姑

序号	人物名称	名号/称谓/别称
33	神农	赤帝；农皇；神农大帝；五穀神农大帝；炎帝；神农氏；神农爷；石耳；农皇；药王神；五谷先帝；魁隗氏；连山氏；列山氏；朱襄氏；稷
34	舜	虞舜；重华；虞氏；帝舜；湘君；有虞；舜王；息壤先人
35	燧人氏	遂皇；燧人
36	太白金星	太白；太白腾格里；太上老君；太玄；道德天尊；太一；天皇大帝；太乙天尊；金阙圣君；九玄帝君；重耳；郁华子；九灵老子；广寿子；广成子
37	檀君	坛君
38	天帝	白帝天王；天皇；天皇氏；太一；四大天王；护法四天王；四大金刚；四天王；护世四天王；天皇大帝；元始天王；天君；太帝；后帝；皇天
39	天女	墨美；木美姬；恩古伦；正古伦；佛库伦；恩库伦；正库伦；魑；
40	西王母	金母元君；德金巴乌；九光元女；金母；瑶池金母；金母元君；西姥；王母娘娘；云华夫人
41	炎帝	皇神农；石年；炎庆甲；神农氏；帝魁；祝融
42	阎王	五道阎罗；阎君；阎门坎；阎魔王；阎罗王；十殿阎罗
43	尧	帝啇；唐尧；陶唐氏；帝尧
44	瑶姬	巫山神女
45	有巢氏	巢父；大巢氏
46	玉皇大帝	玉皇；上帝；老天爷；玉帝；天神；天王
47	灶神	灶王爷；灶王；灶君；灶君菩萨；灶君老爷；穷蝉；灶君公；张灶王；火帝君；九司命真君；九天东厨烟主；护宅天尊；灶君娘娘；火塘神
48	织女	毛衣女；天孙娘娘；天女；东桥；天女娘；收阴；支机女
49	祝融	火神；赤帝；吴回；回禄；黎祖；南火
50	颛顼	高阳氏；黑帝；共工；高王爷；乾荒

依据上述名称映射表，项目组进行了面向上述 9 类资料数字化文本的广泛数据收集，通过“人物名称+关键词”模糊检索和专家泛读的方式，完成了常见神话人物起源实例数据（字段）的大规模采集。以盘古为例，其在三本书中的人物起源信息如图 2 所示。

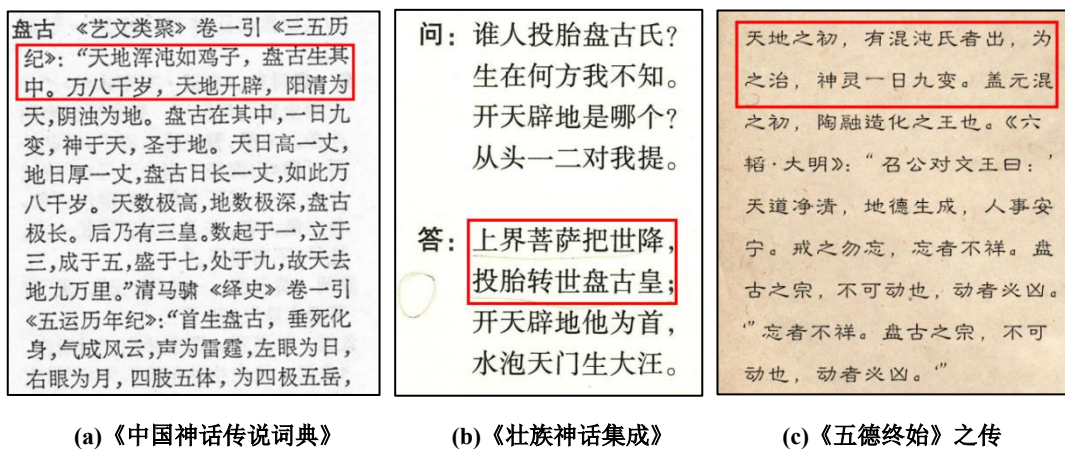


图 2 盘古起源数据来源

Figure 2 Data sources of the origin of Pangu

对于民族、流传地等信息的采集方面，在中国民间文艺家协会、国家文化部（现文旅部）、国家民族事务委员会及地方文化部门的主持和引导下，包括《中国民间故事集成》等在内的大部分专著均在文末或文注中标识出了规范、详尽的民族、讲述人、采集人、翻译人和采集地点等信息，为本数据集相关信息的科学、完整收录奠定了坚实基础。民族和流传地信息的采集技术路线如图 3 所示。

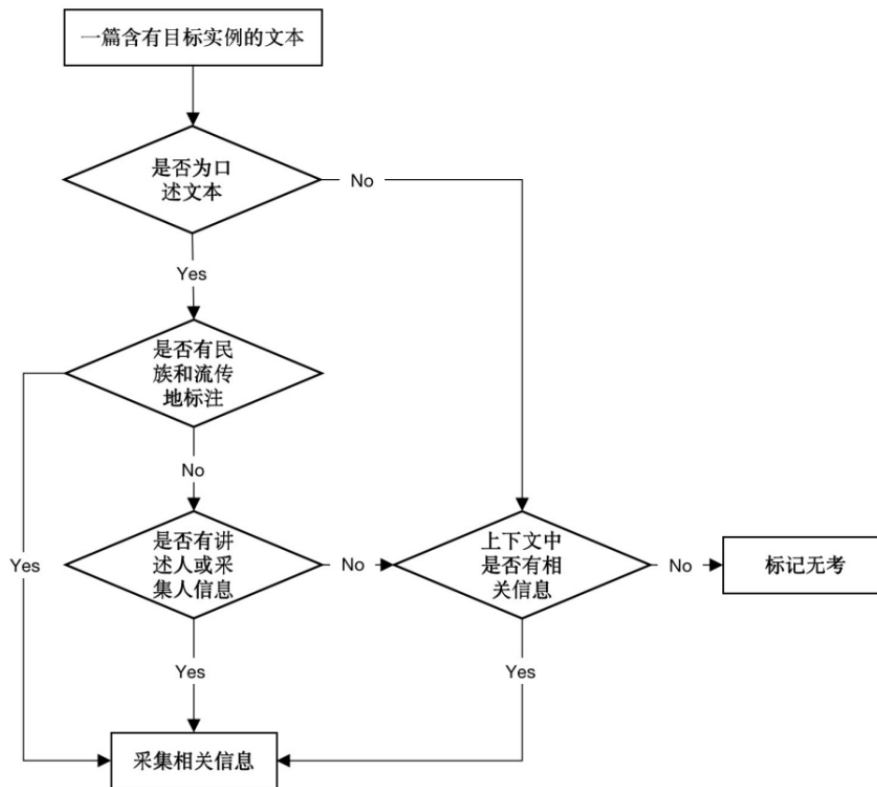


图 3 民族、流传地信息采集技术路线图

Figure 3 Technology roadmap of information collection for nationalities and regions

通过上述全字段信息的采集和筛选,完成对 50 位核心神话人物在多民族、多地区范围内的数据相对完整覆盖,为后续数据的进一步处理和分析提供了可能。

1.3 数据的处理

本数据集元数据共包括 9 (A-I 列) 字段信息, A 列为神话“人物名称”列, B-E 列为母题描述字段, F-I 列为母题实例相关信息列。

1.3.1 母题描述字段处理

在综合采集表 1 常见神话人物起源实例数据的基础上,参考《民间文学母题索引》^[19]、《都柏林核心元素集》^[20]、AT 分类法 (Aarne-Thompson Classification System)、《中国神话母题 W 编目》^[11]、《中国神话人物母题 W0 目录》^[21]、《中国神话人物母题数据与索引》^[22]等叙事类数据结构规范,将本数据集的采集和筛选标准按照“人物产生原因”“人物产生方式”和“与人物产生有关的其他母题”的树形结构划分为四个层级,详细母题层级结构标准表如表 3 所示。

表 3 神话人物起源母题数据结构表

Table 3 Data structure table of the origin motifs of mythical characters

一级母题	二级母题	三级母题	四级母题
A 的产生	A 产生的原因	(空)	A 产生的原因是……
	A 自然存在	(空)	A 自然存在于……
	A 来源于某个地方	(空)	A 来源于……
	A 是造出来的	神造 A	B 造 A
		神性人物造 A	B 造 A
	A 是生育产生的	神生 A	B 生 A
		神性人物生 A	B 生 A
		特定的人生 A	B 生 A
		动植物生 A	B 生 A
		无生命物生 A (自然物生 A)	B 生 A
		卵生 A	B 生 A
		其他特定物生 A	B 生 A
	A 是变化产生的	神变化产生 A	B 变成 A
		神性人物变化产生 A	B 变成 A
		动植物变化产生 A	B 变成 A
		其他特定物变化产生 A	B 变成 A
	A 是婚生产生的	神婚生 A	B 婚生 A
		神性人物婚生 A	B 婚生 A
		人婚生 A	B 婚生 A
		动植物婚生 A	B 婚生 A

一级母题	二级母题	三级母题	四级母题
		自然物婚生 A	B 婚生 A
		异类婚生 A	B 婚生 A
		怀孕生 A 的时间	怀孕生出 A 的时间是……
		怀孕生 A 的地点	怀孕生出 A 的地点是……
		怀孕生 A 的结果	怀孕生出 A 的结果是……
	A 是感生产生的	神感生 A	B 感生 A
		神性人物感生 A	B 感生 A
		动植物感生 A	B 感生 A
		其他特定人物感生 A	B 感生 A
		感生 A 的原因	感生 A 的原因是……
		感生 A 的时间	感生 A 的时间是……
		感生 A 的地点	感生 A 的地点是……
		特定人物感特定物生 A	B 感 C 生 A
	A 是转生产生的	特定对象转生 A	B 转生 A
		A 转生的原因	A 转生的原因是……
		A 转生的时间	A 转生的时间是……
		A 转生的地点	A 转生的地点是……
	与 A 产生有关的其他母题	A 产生的征兆	A 产生时的征兆是……
		A 产生的时间	A 产生的时间是……
		A 产生的地点	A 产生的地点是……

注：A、B、C 表示特定神话人物或事象，其中 A 为主体神话人物，B、C 为关联神话人物或特定神话事象。

其中，一级母题到三级母题为固定表述，标识了具体实例的所属类型，四级母题是采集实例信息的最终转译，是神话人物起源叙事的最小有意义单元。各层级母题语义表达含义层层递进，构建起围绕神话人物起源的完整叙事表述。如图 2(b) 中的实例为“上界菩萨把世降，投胎转世盘古王”，其对应的各级母题数据如表 4 所示。

表 4 图 2(b) 各层级母题数据示例

Table 4 Figure 2(b) Sample motif data of each level

一级母题	二级母题	三级母题	四级母题
盘古的产生	盘古是转生产生的	特定对象转生盘古	上界菩萨转生盘古

1.3.2 母题实例相关字段处理

本数据集的母题实例相关字段包括民族、实例描述、流传地和文献出处 4 项。其中：

1. 民族字段处理。本数据集的“民族”字段设置，旨在满足相应神话资源研究的需要，服务于跨民族对比研究等。民族的判断依据主要为原文本标注或神话讲述人（或采录人或翻译人）的籍贯身份。对于原文本中没有明确标注的情况，则根据上下文线索进行人工判断进行增补。同时，对原文本中少数民族支系进行了规范，统一归并为 56 个民族的标准表述，如将“撒尼”“阿细”“阿哲”“罗婺”统一为“彝族”，将“过山瑶”“盘古瑶”“红头瑶”“顶板瑶”统一为“瑶族”，等等。对没有明显标注少数民族信息的汉文字典籍实例，则统一标注为“汉族”。对于流传民族范围较广的、在多民族间得到共同认可的实例表述，则统一标注为“中华民族”。除此之外难以辨析者，均标注“（无考）”。

2. 实例描述字段处理。本数据集在从神话文本提取母题实例的过程中，研究团队旨在实事求是地使用神话材料，以关注神话文本的多样性为宗旨，兼顾了神话文本的异文。主要包括以下几种情况：（1）流传于不同地区的同一部作品；（2）不同讲述人讲述的同一部作品；（3）不同搜集者搜集的同一部作品；（4）不同出版物收集的表述上有差异的同一部作品。诸如此例，在实例描述数据采集过程中，选择全部保留，并详注属性数据（包括母题层级、文本、流传民族、采录人、讲述人、文本出处等），最大限度保证每一条数据的完整性，以备日后数据集使用者根据需要寻源讨本。

3. 流传地字段处理。标注流传地旨在使读者能够从神话流传的空间角度做出审视。信息录入依据为原文本标注、口传文本采集地点或神话讲述人（或采录人或翻译人）的籍贯信息。同时，根据国家标准《地名信息服务》第 1 部分：通则（GB/T 23295.1—2009）^[23]，本数据集中的地名字段包括地名专名、标准地名、地名简称、地名别名、现今地名、历史地名等地名基本属性信息，以及相关人物、事件、作品等历史地名文化相关信息，上述情形数据集采用“省/直辖市/自治区/特别行政区·地级市/地区/自治州/盟·市辖区/县/县级市/自治县/旗/自治旗/特区/林区”的格式进行规范表述，如“河南省·焦作市·武陟县”。有些古代典籍中的地名与今天的行政区划地名相比已发生很大变化，或已经消失，或考证困难，对这种情况直接使用原文表述，不再做对应性标注。对难以确定流传地的文本，标注“（无考）”。

4. 文献出处字段处理。本数据集采取实例原文本来源出版信息全面呈现的形式，针对原作品信息缺失的情况，数据集尽可能核实补充，旨在数据集成网络检索的精准定位与神话数据的规范化应用。同一个“实例描述”有时会有两个以上“文献出处”，这类情况或源于不同讲述人，或源于不同的采录者，或源于不同的出版物，具体情况根据出版物原文比对确定。出处信息的数据规范参照国家标准《信息与文献参考文献著录规则》（GB/T 7714—2015）^[24]以及《中国神话母题 W 编目》系列丛书规则。

2 数据样本描述

2.1 数据集样例

中国常见神话人物的起源母题数据集包含 1620 条数据，包含序号、人物名称、一级母题、二级母题、三级母题、四级母题、民族、实例描述、流传地和文献出处 10 个字段。表 5 为神话人物的起源母题的 4 条样例数据（伏羲、女娲、盘古、雷神）。

表 5 中国常见神话人物的起源母题数据样例

Table 5 Sample data of the origin motifs of core mythological characters in China

序号	人物名称	一级母题	二级母题	三级母题	四级母题	民族	实例描述	流传地	文献出处
1	伏羲	伏羲的产生	伏羲是生育产生的	神性人物生伏羲	华胥生伏羲	汉族	华胥在防风脚印里跑了一圈，怀孕生下伏羲。	浙江省·湖州市·德清县	《大禹找防风》（1988.12.20），见中国民间文学集成全国编辑委员会编《中国民间故事集成》（浙江卷），北京中国 ISBN 中心 1997 年版，第 68～69 页。
2	女娲	女娲的产生	女娲是变化产生的	动植物变化产生女娲	蛇族兄妹中的妹妹变成女娲	汉族	蛇族兄妹修炼自身，妹妹修炼成人头人身的女娲。	广东省·湛江市·坡头区	《女娲与海龟》，见中国民间文学集成全国编辑委员会编《中国民间故事集成》（广东卷），北京：中国 ISBN 中心 2006 年版，第 4 页。
3	盘古	盘古的产生	盘古是生育产生的	神生盘古	土地神生盘古	毛南族	盘古是土地神的子孙。	广西壮族自治区·河池市·环江毛南族自治县	《盘古的传说》，见袁凤辰等编《毛南族、京族民间故事选》，上海上海文艺出版社 1987 年版。（b）同，见姚宝琮主编《中国各民族神话》（土家族、毛南族、侗族、瑶族），太原山西出版传媒集团·书海出版社 2014 年版，第 45 页。
4	雷神	雷神的产生	雷神是婚生产生的	人婚生雷神	一对男女婚生雷公	苗族	引雄和妮仰婚后，妮仰吃了牛头和虎爪，生雷公。	广西壮族自治区·柳州市·融水苗族自治县	《创世记》，见谷德明编《中国少数民族神话》，北京：中国民间文艺出版社 1987 年版，第 545 页。

2.2 数据统计分析

2.2.1 中国常见神话人物的起源母题的数量统计分析

该数据集中伏羲、女娲、黄帝、盘古等 50 个神话人物的起源母题的数量统计结果如图 4 所示。本数据集中盘古、大禹、盘瓠起源母题数量较多，占比较大，特别是盘古的数据占据了总数据量的五分之一；黄帝和尧次之；其余神话人物的起源母题数量较少，占比小。

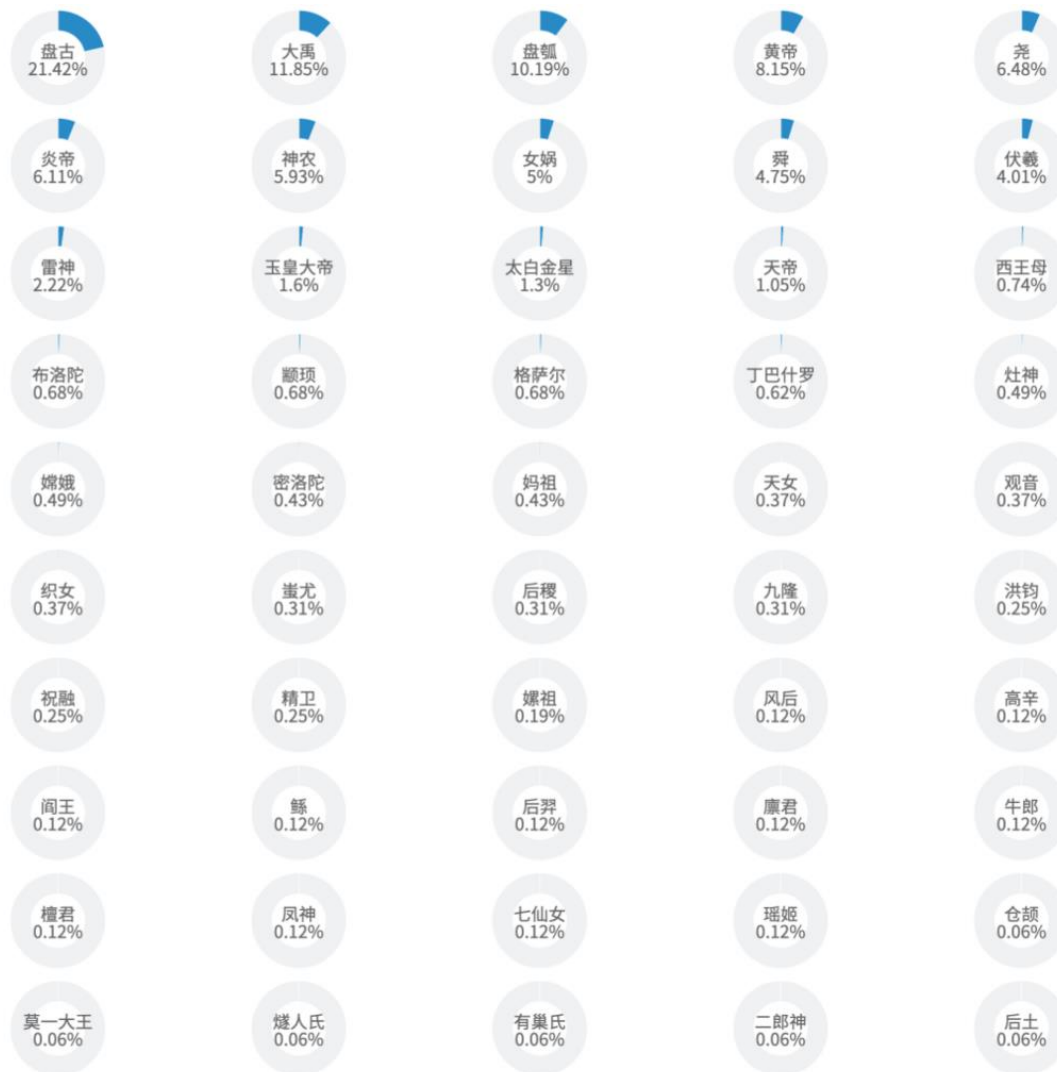


图 4 中国常见神话人物的起源母题的数量统计图

Figure 4 Statistics of the origin motifs of the core mythological characters in China

2.2.2 中国常见神话人物的起源母题的产生方式的统计分析

经统计，本数据集中的常见神话人物的起源母题中的产生方式主要以“生育产生”和“感生产生”为主，“变化产生”和“来源于某个地方”的方式次之，“自然存在”与“造出来”的方式产生的神话人物较少。如图 5 所示，常见神话人物的起源母题的产生方式中有 292 条数据是“生育产生”的，这也进一步印证了在神话作为大众精神意识象征在历史层累流传中的无意识取舍，通过对神“生育”、“生殖”和“繁衍”等人性行为的叠加，探索神的人性化演变。神话人物以“自然存在”方式产生的数据较少，只有 9 条。

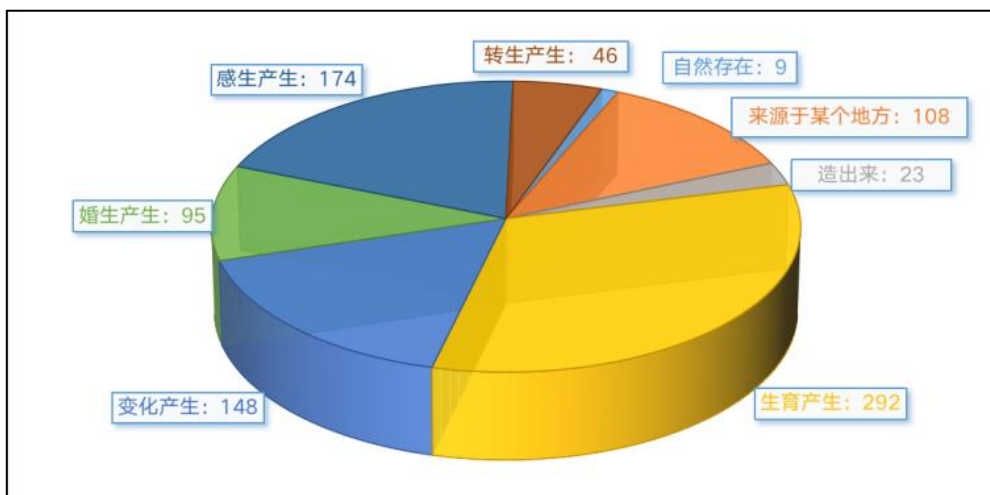


图 5 中国常见神话人物起源母题产生方式数量统计图

Figure 5 Statistical chart of the number of origin motifs of core mythological characters in China

经统计分析,可以发现“生育产生”“感生”“婚生”等人物产生方式具有较高的中介性,在全部人物关系中起到了重要的勾连作用。例如“颛顼生鲧、鲧生大禹、女娲生大禹、女娲生舜、女娲生太白金星、盘古婚生伏羲、盘古婚生西王母、雷神婚生伏羲”等,通过神话人物的产生方式,发现盘古、伏羲、女娲、雷神、颛顼、鲧、禹、舜等人物之间具有强烈互文关系,相关发现将辅助探索神话人物之间的社会关系网络和特定人物形象群关系谱系。

2.2.3 中国常见神话人物的起源母题的民族统计分析

经统计,本数据集中的 50 个常见神话人物的起源母题的流传民族有汉族、白族、苗族、瑶族、壮族、彝族等 32 个民族。如图 6 所示。

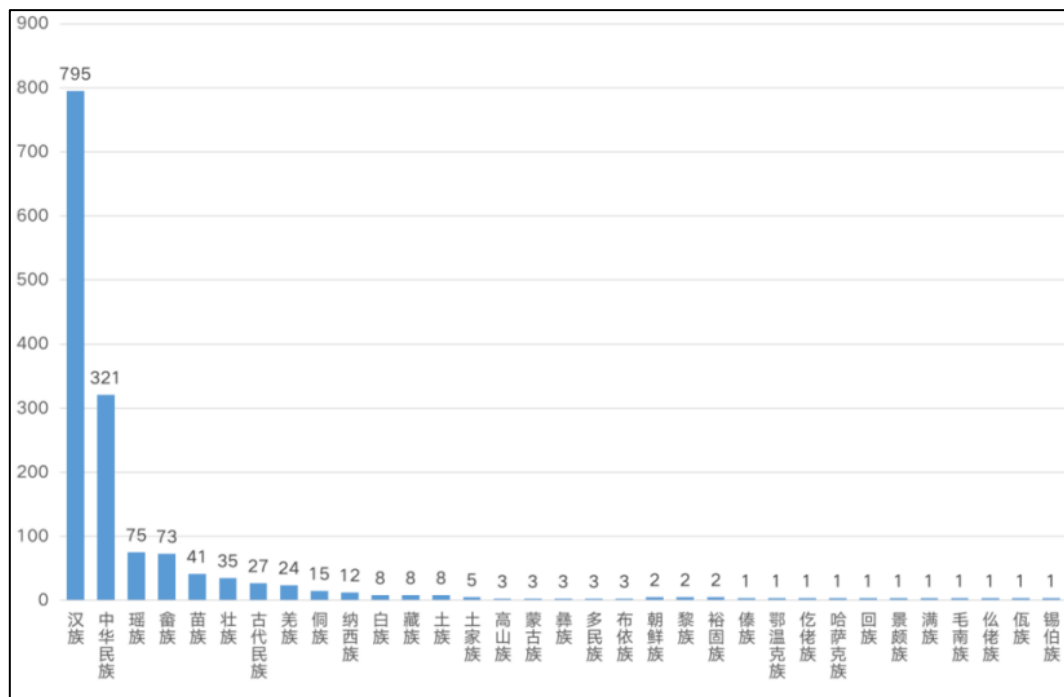


图 6 中国常见神话人物的起源母题的民族数量统计图

Figure 6 Statistical chart of the number of nationalities on the origin motifs of the core mythical characters in China

神话人物的起源母题的流传民族主要呈现以下三种特点。第一，常见神话人物的起源母题的流传民族以汉族和中华民族为主，瑶族和畲族次之，其余民族的数量较少，1116 条数据的流传民族都涉及到汉族和中华民族。第二，常见神话人物的起源母题的流传具有多民族传播的特点，以盘古的起源母题为例，其在汉族、壮族、瑶族、土族等 11 个民族都有流传。第三，部分神话人物起源母题具有单一民族间流传的倾向，如檀君的起源母题主要流传于朝鲜族，再如廪君的起源母题主要流传于土家族。因此通过对本数据集中的神话人物的起源母题的民族进行统计分析，可以发现神话人物的起源母题流传的主要民族，对增强民族文化认同、培育中华民族共同体意识意义重大。

2.2.4 中国常见神话人物的起源母题的流传省份统计分析

经统计，本数据集中的常见神话人物的起源母题的流传涵盖 29 个省份和 2 个地区，如图 7 所示，神话人物的起源母题的具有以中原地区为主，少数民族聚集地区流传为辅的特征。河南省、湖北省、山西省等为主的中原地区流传的神话人物的起源母题较多，例如河南省关于神话人物的起源数据高达 199 条；浙江省、四川省、广西壮族自治区等少数民族聚集省份流传的神话人物的起源母题次之；其余省份中的神话人物的起源母题流传较少，如吉林省、青海省、西藏自治区等。

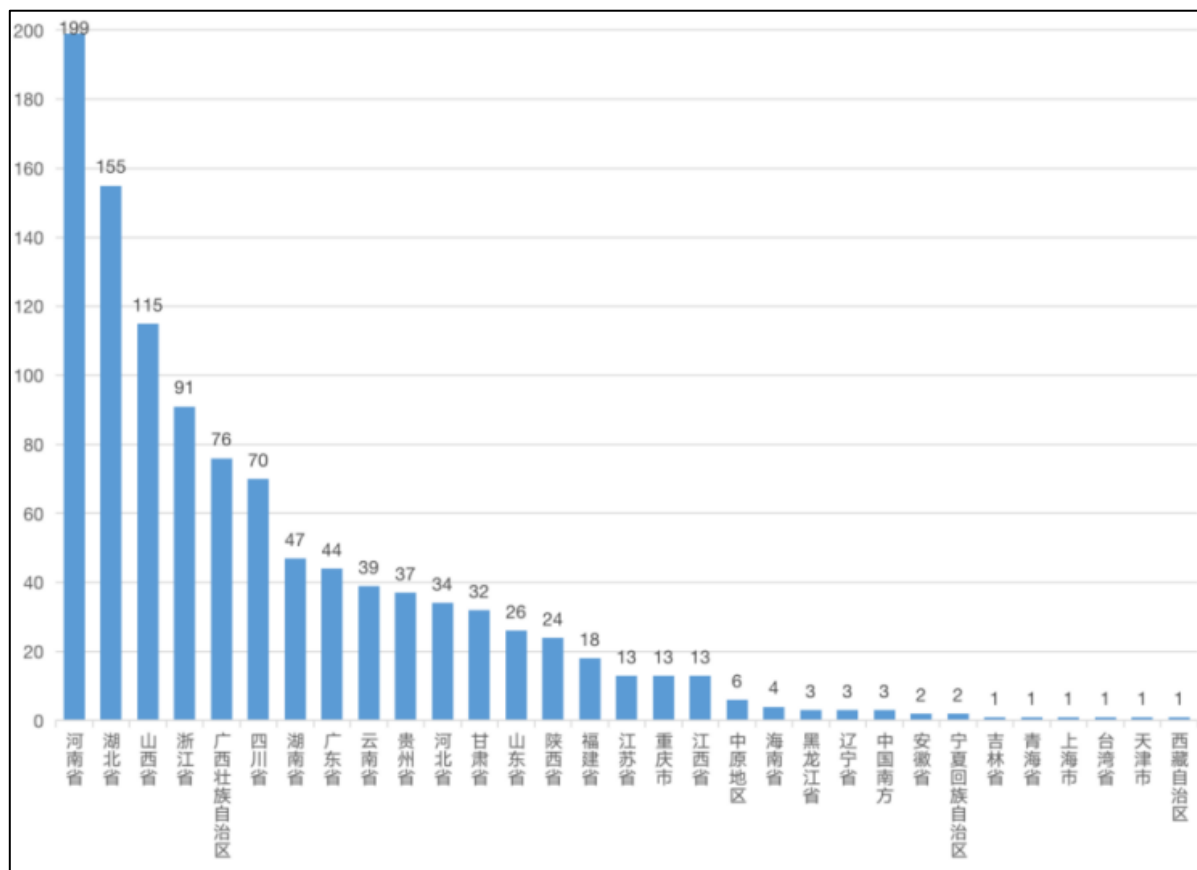


图 7 中国常见神话人物的起源母题的流传省份统计图

Figure 7 Statistical chart of the number of spreading provinces on the origin motifs of the core mythical characters in China

3 数据质量控制和评估

本数据集的数据质量由现有书籍资料记录的完整性和准确性保证，所涉及典籍资源来源广泛、

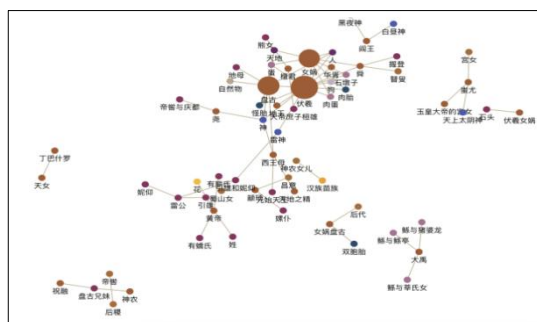
载体完备并具有较高的领域权威性，数据质量良好，可信度高。《中国民间故事集成》、《中国各民族原始宗教资料集成》等公开发行的民间文学类丛书中对神话人物产生叙事的讲述人、流传地、版本、民族等信息进行了相对统一规范的载录，数据完整、质量控制严格。同时，本研究根据《地名信息服务（GB/T23295—2009）》和《信息与文献参考文献著录规则（GB/T 7714—2015）》对文本采录标准字段和地名进行了规范统一。地名标注主要是依据神话原文本、口传文本采集地点或神话讲述人（或采录人或翻译人）的籍贯信息等信息进行标注。有些古代典籍中的地名与今天的行政区划地名相比已发生很大变化的部分通过人工修订和专家复合的方式完成了古今地名的转化，例如处理伏羲的产生地点“成纪”时，既保留了古地名的表述，并标注了对应的今天的地名，“成纪（甘肃省·天水市）”。而对于古籍中的地名现在已经消失，或考证困难的直接使用原文表述，不再做对应性古今地名的标注。例如，黄帝出生于“寿丘”，保留了原文表述“寿丘”。对于难以确定流传地的文本，标注为“（无考）”。

在叙事母题字段的确定方面，在沿用国际通用故事情节类型 AT 分类法的基础上，参照国内神话母题 W 编码进行进一步规范，确定了四级母题录入格式，保证了叙事内容的完全覆盖和叙事结构的合理完备，确保了核心数据的本土性、科学性和高可用。在数据录入和修正环节，经过多轮人工校验，确认无误，数据的存储与管理符合领域标准规范。

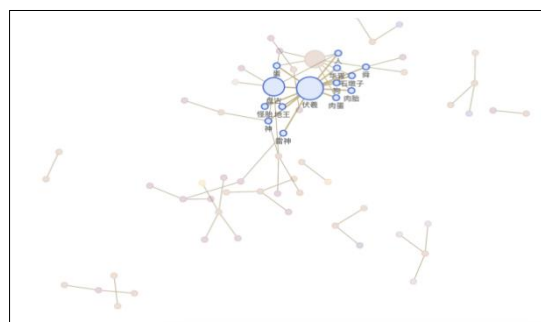
4 数据使用方法和建议

主流的编程语言均可调用相关函数库读写本数据集，实现数据的处理和分析。本数据集为研究中国各民族精神文化内涵、进行跨地区文化研究、创新中华优秀传统文化多维度分析以及助力国家和地方文化部门战略决策等提供了数据基础。同时也为传统文化创意产业、与神话有关的文化旅游、相关数据平台建设等提供直接数据服务，也将进一步服务于民众文化常识的普及和中华文化自信的提升。本数据集数据准确全面，可为多层次用户提供相应的内容参考、案例分析和数据定量定性研究。

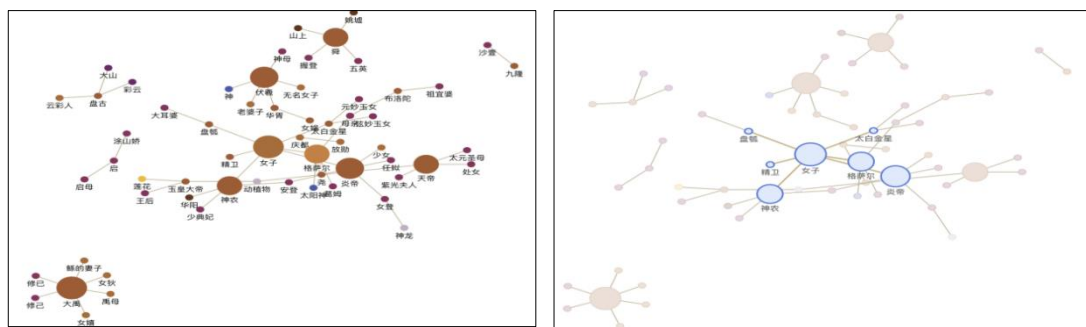
例如，利用本数据集可以进行相关领域知识图谱的开发，辅助研究者完成深层次知识挖掘与分析，完成从“数据服务”到“知识服务”的功能提升。基于绝大多数母题字段为“主—谓—宾”结构的数据基础，通过 SPO (subject-predicate-object) 方式完成三元组信息的抽取。同时，综合产生地点和民族信息等结构化数据，构建起常见神话人物起源叙事知识图谱，实现神话人物起源知识域的映射地图检索和完整关系呈现。同时，在此基础上搭建起相应的知识服务平台 (<http://myth.scidb.cn/>)，实现包括智能问答、关系检索、路径查询、社会网络分析等在内的多种功能，图 8 为其中与特定产生类型相关的人物知识图谱示意图。据此可以直观辅助跨民族、跨区域视域下的特定或多个人物的比较研究和社会网络分析。



(a) 婚生产关系图谱（部分）



(b) 伏羲婚生产关系谱系



(c)感生产生关系图谱（部分）

(d)女子感生产生关系谱系

图8 特定产生类型人物关系知识图谱

Figure 8 Map of relationship between specific generation types

数据作者分工职责

王京（1989—），女，山东聊城人，博士，研究方向为人文社科大数据，神话学。主要承担工作：数据处理、加工，数据集生成，论文撰写。

熊惠（1994—），女，贵州贵阳人，博士生，研究方向为神话学。主要承担工作：数据采集、处理，数据集生成，论文撰写。

其乐格乐（1996—），女，内蒙古兴安盟人，博士生，研究方向为神话学。主要承担工作：数据采集、处理，数据集生成，论文撰写。

杜一（1988—），男，山东聊城人，博士，研究员，硕士生导师，研究方向为大数据挖掘与分析技术。主要承担工作：研究思路设计与论文撰写指导。

周园春（1975—），男，江西鹰潭人，博士，研究员，博士生导师，研究方向为数据挖掘、大数据处理。主要承担工作：研究思路设计与论文撰写指导。

参考文献

- [1] MICHEL J B, SHEN Y K, AIDEN A P, et al. Quantitative analysis of culture using millions of digitized books[J]. Science, 2011, 331(6014): 176–182. DOI: 10.1126/science.1199644.
- [2] SCHICH M, SONG C M, AHN Y Y, et al. A network framework of cultural history[J]. Science, 2014, 345(6196): 558–562. DOI: 10.1126/science.1240064.
- [3] 中国历代人物数据库[DB/OL]. [2023-01-06]. <https://projects.iq.harvard.edu/cbdb/home>. [China Biographical Database[DB/OL]. [2023-01-06]. <https://projects.iq.harvard.edu/cbdb/home>.]
- [4] ASP 世界历史人物索引数据库 [DB/OL]. [2023-01-06]. <http://www.inthefirstperson.com/firp/index.shtml>. [Alexander Street Press World Historical People Index Database[DB/OL]. [2023-01-06]. <http://www.inthefirstperson.com/firp/index.shtml>.]
- [5] 中国古典文献资源数据库[DB/OL]. [2023-01-06]. <http://hanchi.ihp.sinica.edu.tw/mql/login.html>. [Chinese Classical Literature Resources Database[DB/OL]. [2023-01-06]. <http://hanchi.ihp.sinica.edu.tw/mql/login.html>.]
- [6] 中国方志库[DB/OL]. [2023-01-06]. <http://dh.ersjk.com/>. [Chinese Local Chronicle Database[DB/OL]. [2023-01-06]. <http://dh.ersjk.com/>.]

- [7] 明清实录数据库[DB/OL]. [2023-01-06]. <http://lib.nju.edu.cn/info/1013/2670.htm>. [Ming and Qing Dynasty Record Database[DB/OL]. [2023-01-06]. <http://lib.nju.edu.cn/info/1013/2670.htm>.]
- [8] 中文诗歌[DB/OL]. [2023-01-06]. <https://github.com/chinese-poetry/chinese-poetry>. [Chinese-poetry[DB/OL]. [2023-01-06]. <https://github.com/chinese-poetry/chinese-poetry>.]
- [9] “一带一路”战略支撑平台[DB/OL]. [2023-01-06]. <https://ydy1.drcnet.com.cn/#/>. [“The Belt and Road” Strategic Support Platform[DB/OL]. [2023-01-06]. <https://ydy1.drcnet.com.cn/#/>.]
- [10] 施艳萍, 李阳, 卞一洋. 基于区块链的人文社科专题数据库协同管理研究[J]. 信息资源管理学报, 2020, 10(5): 30–37. DOI: 10.13365/j.jirm.2020.05.030. [SHI Y P, LI Y, BIAN Y Y. Research on collaborative management of humanities and social sciences thematic database based on blockchain[J]. Journal of Information Resources Management, 2020, 10(5): 30–37. DOI: 10.13365/j.jirm.2020.05.030.]
- [11] 王宪昭. 中国神话母题 W 编目[M]. 北京: 中国社会科学出版社, 2013: 21. [WANG X Z. Wang’s catalogue: motif in China’s mythology[M]. Beijing: China Social Sciences Press, 2013.]
- [12] 斯大林. 斯大林全集: 第二卷[M]. 北京: 人民出版社, 1953. [Сталин, И. В. Complete works of Stalin: introduction to volume II[M]. Beijing: People's Publishing House, 1953.]
- [13] 国家质量监督检验检疫总局. 地名分类与类别代码编制规则: GB/T 18521—2001[S]. 北京: 中国标准出版社, 2002. [General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China. Rules for classification of geographical names and code representation: GB/T 18521—2001[S]. Beijing: Standards Press of China, 2002.]
- [14] 中国民间文学集成全国编辑委员会. 中国民间故事集成[M]. 北京: 中国 ISBN 中心, 1992–2008. [National Editorial Committee for the Integration of Chinese Folk Literature Integration of Chinese Folk Stories[M]. Beijing: China ISBN Center, 1992 – 2008.]
- [15] 陶阳, 钟秀. 中国神话[M]. 北京: 商务印书馆, 2008. [TAO Y, ZHONG X. Chinese mythology[M]. Beijing: The Commercial Press, 2008.]
- [16] 吕大吉, 何耀华. 中国各民族原始宗教资料集成[M]. 北京: 中国社会科学出版社, 1996–2013. [LV D J, HE Y H. A general introduction to a series of sourcebooks on the primitive religions of China[M]. Beijing: China Social Sciences Press, 1996–2013.]
- [17] 马昌仪. 中国神话学文论选萃[M]. 北京: 中国广播电视出版社, 1994. [MA C Y. Selected works of China’s mythology literary theory[M]. China Radio & Television Publishing House, 1994.]
- [18] 马学良, 梁庭望, 张公瑾. 中国少数民族文学史[M]. 修订本. 北京: 中央民族大学出版社, 2001. [MA X L. History of China minority literature[M]. Beijing: The Central University of Nationalities Press, 2001.]
- [19] THOMPSON S. Motif-Index of Folk-Literature[M]. Indiana: Indiana University Press, 1960.
- [20] 国家质量监督检验检疫总局, 中国国家标准化管理委员会. 信息与文献 都柏林核心元数据元素集: GB/T 25100—2010[S]. 北京: 中国标准出版社, 2010. [General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Information and documentation—the Dublin Core metadata element set: GB/T 25100—2010[S]. Beijing: Standards Press of China, 2010.]
- [21] 王宪昭. 中国神话人物母题(W0)数据目录[M]. 北京: 中国社会科学出版社, 2019. [WANG X Z. A motif catalogue of China’s mythical character[M]. Beijing: China Social Sciences Press, 2019.]

- [22] 王宪昭. 中国神话人物母题数据索引[M]. 北京: 中国社会科学出版社, 2020. [WANG X Z. Motif-index of China mythology characters database[M]. Beijing: China Social Sciences Press, 2020.]
- [23] 国家质量监督检验检疫总局, 中国国家标准化管理委员会. 地名信息服务.第 1 部分: 通则: GB/T 23295.1—2009[S]. 北京: 中国标准出版社, 2009. [General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. Geographical names information services—Part 1: general principle: GB/T 23295.1—2009[S]. Beijing: Standards Press of China, 2009.]
- [24] 国家质量监督检验检疫总局, 中国国家标准化管理委员会. 信息与文献参考文献著录规则: GB/T 7714—2015[S]. 北京: 中国标准出版社, 2015. [General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, National Standardization Administration of the People's Republic of China. Information and documentation—rules for bibliographic references and citations to information resources: GB/T 7714—2015[S]. Beijing: Standards Press of China, 2015.]

论文引用格式

王京, 熊惠, 其乐格乐, 等. 中国常见神话人物起源母题数据集[J/OL]. 中国科学数据, 2023, 8(2). (2023-04-20). DOI: 10.11922/11-6035.csd.2023.0039.zh.

数据引用格式

王京, 熊惠, 其乐格乐, 等. 中国常见神话人物起源母题数据集[DS/OL]. Science Data Bank, 2023. (2023-02-02). DOI: 10.57760/sciencedb.07047.

A dataset of origin motifs of major Chinese mythological characters

WANG Jing^{1*}, XIONG Hui², Qilegele², DU Yi¹, ZHOU Yuanchun¹

1. Computer Network Information Center, Beijing 100190, P. R. China

2. University of Chinese Academy of Social Sciences, Beijing 102488, P. R. China

*Email: jwang@cnic.cn

Abstract: Myths is an important carrier of traditional Chinese culture. There are numerous myths of different ethnic minorities, rich in content and diverse in form. Many mythological characters have become precious memories in the course of human history and Chinese civilization. With the analysis of mythical figures as the starting point, we take relevant mythological texts and narratives as research topics and important data sources in the field of humanities and social sciences, which is of great cultural significance and academic value in promoting the construction of big data in humanities, deepening the analysis of traditional Chinese cultural data, and innovating new methods of humanities and social science research. In this paper, we took the origin narrative motif of 50 core mythological characters in China as the research object, and collected 1,620 valid data from them, including the standard fields of motif, example, nationality, place of spread and

source of literature. Meanwhile, we systematically analyzed and explained the data sources, data collection methods, data classification and structure. Moreover, we described the data samples in detail; and on the basis of listing relevant dataset examples, we further carried out a multi-dimensional statistical analysis of the way of generation, ethnic groups and geographical location, etc. Finally, we provided some suggestions on the application of the dataset in exploring the connotation of China's multi-ethnic culture, cross regional cultural research and analysis of Chinese traditional culture.

Keywords: Humanities data; Mythical characters; Myth metadata; Myth data analysis

Dataset Profile

Title	A dataset of origin motifs of major Chinese mythological characters
Data corresponding author	WANG Jing (jwang@cnic.cn)
Data authors	WANG Jing, XIONG Hui, Qilegele, DU Yi, ZHOU Yuanchun
Time range	200 B.C – 2022 A.D
Geographical scope	China
Data volume	648.98 KB
Data format	*.xlsx
Data service system	https://doi.org/10.57760/sciencedb.07047
Sources of funding	CAS Network Security and Informatization Project (CAS-WX2021SF-0308); National Social Science Foundation Project (21CZW058); Chinese Post-doctorate Science Foundation Project (2020M680682)
Dataset composition	The dataset includes the origin narrative motifs of 50 core Chinese mythical characters in a multi-ethnic and multi-regional context, covering the causes, ways, times, places and other major types. Under the category of “generation mode of characters” which are subdivided into 8 types, i.e. natural generation, marriage, birth, and induction. There are a total of 1,620 pieces of data, each of which contains 10 fields of information. The dataset also contains a relationship map of mythical characters based on the above data. The map is based on SPO triplet extraction principle, including 369 nodes and 1,240 edges.