

基于优势流约束下聚类方法计算的 中国都市圈数据集研发

李娟, 张扬*, 唐松, 刘旭宏

成都理工大学地理与规划学院, 成都 610059

摘要: 都市圈是推进新型城镇化建设和高质量发展的重要空间载体, 科学识别都市圈空间范围是开展相关研究和规划实践的重要基础。本研究在厘清都市圈概念的基础上, 首先选取都市圈中心城市, 然后利用百度人口迁徙数据构建城市人流联系网络, 最后利用优势流约束下的DBSCAN聚类进行中国都市圈空间范围识别。数据集内容包括: (1) 都市圈中心城市列表及空间分布; (2) 城市间人流联系强度矩阵和人流联系网络; (3) 都市圈识别结果列表及空间分布。数据集存储为.shp和.xlsx格式, 数据集由29个数据文件组成, 数据量为67.9 MB(压缩为1个文件, 31.3 MB)。

关键词: 都市圈识别; 人流网络; 中心城市; 优势流约束; DBSCAN聚类

DOI: <https://doi.org/10.3974/geodp.2025.03.04>

CSTR: <https://cstr.escience.org.cn/CSTR:20146.14.2025.03.04>

数据可用性声明:

本文关联实体数据集已在《全球变化数据仓储电子杂志(中英文)》出版, 可获取:

<https://doi.org/10.3974/geodb.2025.08.01.V1> 或 <https://cstr.escience.org.cn/CSTR:20146.11.2025.08.01.V1>.

1 前言

都市圈本质上是一个城市功能地域概念^[1]。都市圈通常依托综合性特大城市的扩散与辐射效应, 带动周边大中小城市协同发展, 逐步形成空间紧密联系、功能深度融合的城市密集区域^[2]。其地域范围往往比城市群小, 是城市群的核心区域^[3,4]。作为推动新型城镇化战略实施和实现高质量发展的关键载体, 都市圈的重要作用日益凸显^[5-7]。都市圈建设近年来持续得到国家战略层面的政策聚焦, 如2014年《国家新型城镇化规划(2014-2020)》提出构建通勤高效、协同发展的一体化都市圈。此后, 国家发展改革委于2019年印发《关于培育发展现代化都市圈的指导意见》, 进一步推动都市圈建设迈向现代化与高质量发展。此后, 国家再次强调要以城市群、都市圈为依托构建大中小城市协调发展格局。作为规划实

收稿日期: 2025-06-03; 修订日期: 2025-08-29; 出版日期: 2025-09-25

基金项目: 国家自然科学基金(42501258, 52478045)

*通讯作者: 张扬, 成都理工大学地理与规划学院, zhangyang2020@cdut.edu.cn

数据引用方式: [1] 李娟, 张扬, 唐松等. 基于优势流约束下聚类方法计算的中国都市圈数据集研发[J]. 全球变化数据学报, 2025, 9(3): 290-298. <https://doi.org/10.3974/geodp.2025.03.04>. <https://cstr.escience.org.cn/CSTR:20146.14.2025.03.04>.

[2] 李娟, 张扬, 唐松等. 基于优势流约束下聚类方法(DBSCAN)计算的中国都市圈数据集[J/DB/OL]. 全球变化数据仓储电子杂志, 2025. <https://doi.org/10.3974/geodb.2025.08.01.V1>. <https://cstr.escience.org.cn/CSTR:20146.11.2025.08.01.V1>.

践和相关研究的基础，科学界定都市圈空间范围，可为优化要素跨区域配置、促进城乡融合发展、推动新型城镇化高质量发展提供必要参考。

全球化和信息化进程加速了城市间要素（人流、物流、信息流、资金流、技术流等）的流动，深刻塑造了城市网络关联并影响着都市圈的形成与发展^[8]。然而，传统都市圈识别方法多依赖静态地理空间分析^[9,10]，主要采用非农人口占比^[1]、人口规模^[11]和二三产业产值等属性数据结合地域邻近因素判别^[12]，忽视动态要素流的空间交互作用。近年来，随着大数据技术的蓬勃发展，类型丰富且获取成本较低的地理流大数据可为都市圈识别提供助力^[13]。因此，本文在选取都市圈中心城市的基础上，基于百度人口迁徙数据构建全国人流网络，运用优势流约束下的 DBSCAN（Density-Based Spatial Clustering of Applications with Noise）聚类进行中国都市圈的空间范围识别。

2 数据集元数据简介

《基于优势流约束下聚类方法（DBSCAN）计算的中国都市圈数据集》^[14]的名称、作者信息、地理区域、数据集组成、数据出版与共享服务平台、数据共享政策等信息见表1。

表1 《基于优势流约束下聚类方法（DBSCAN）计算的中国都市圈数据集》元数据简表

条目	描述
数据集名称	基于优势流约束下聚类方法（DBSCAN）计算的中国都市圈数据集
数据集短名	MetropolitanAreaDelineation
作者信息	李娟，成都理工大学地理与规划学院，lijuan@stu.cdut.edu.cn 张杨，成都理工大学地理与规划学院，zhangyang2020@cdut.edu.cn 唐松，成都理工大学地理与规划学院，2910356995@qq.com 刘旭宏，成都理工大学地理与规划学院，1303940151@qq.com
地理区域	中国 365 个市域单元（包括 4 个直辖市、333 个地级行政区、28 个省级直管单位，港澳台地区数据暂缺）
数据年代	2020 年
数据格式	.shp、.xlsx
数据量	31.3 MB（压缩后）
数据集组成	都市圈中心城市列表及空间分布、城市间人流联系强度矩阵和人流联系网络、都市圈识别结果列表及空间分布
基金项目	国家自然科学基金（52478045）
出版与共享服务平台	全球变化科学研究数据出版系统 http://www.geodoi.ac.cn
地址	北京市朝阳区大屯路甲 11 号 100101，中国科学院地理科学与资源研究所
数据共享政策	（1）“数据”以最便利的方式通过互联网系统免费向全社会开放，用户免费浏览、免费下载；（2）最终用户使用“数据”需要按照引用格式在参考文献或适当的位置标注数据来源；（3）增值服务用户或以任何形式散发和传播（包括通过计算机服务器）“数据”的用户需要与《全球变化数据学报（中英文）》编辑部签署书面协议，获得许可；（4）摘取“数据”中的部分记录创作新数据的作者需要遵循 10% 引用原则，即从本数据集中摘取的数据记录少于新数据集总记录量的 10%，同时需要对摘取的数据记录标注数据来源 ^[15]
数据和论文检索系统	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS, GEOSS, PubScholar, CKRSC

3 数据研发方法

3.1 数据来源与处理

本研究所用数据主要包括社会经济统计数据 and 百度人口迁徙数据。社会经济统计数据主要为常住人口数量、GDP 等,来自于《中国城市统计年鉴 2020》^[16]。百度人口迁徙数据依托大规模地理位置服务技术,可动态记录不同时段各城市人口流向、流量及迁徙强度,多用于城市网络的相关研究^[17]。本文选取 2020 年 1 月 1 日至 1 月 14 日中国 365 个城市(港澳台地区数据暂缺)两两之间的人口迁入迁出数据,涵盖了工作日、周末和法定节假日¹。之后,本文对获取到的百度人口迁徙数据进行了严格的清洗和预处理,通过数据预处理达到去重、处理缺失值、检测异常值的目标,并进行统一数据格式和数据校准,以确保数据的可靠性和安全性。最后,基于预处理完的人口迁入迁出数据构建 365 个城市间的人口迁移规模矩阵。

3.2 主要算法

3.2.1 人流联系强度测算

由于城市间人口的流入流出在方向上有差异,计算人流联系强度时,其数值通过城市 i 和 j 之间的人口流入流出强度之和得到,以消除人流联系的有向性。两个城市之间的人流联系强度计算公式如下^[18]:

$$P_{ij}=P_{i,j}+P_{j,i} \quad (1)$$

式中,城市 i 流向城市 j 和由城市 j 流向城市 i 的人流联系规模分别由 $P_{i,j}$ 和 $P_{j,i}$ 表示。

3.2.2 优势流约束下的 DBSCAN 聚类

DBSCAN 是一种基于密度的空间聚类算法,通过定义邻域半径 Eps 和最小样本数 Minpts,将邻域内样本数达到阈值的核心点及其密度相连对象划分为簇。由于其能将高密度区域划分为簇、有效处理噪声数据并形成任意形状聚类的特点,在空间数据聚类、复杂网络社区识别等领域广泛应用^[19]。在面向城市关联网等流数据场景时,张扬等^[20]创新性地 将优势流分析法与 DBSCAN 聚类算法相结合,提出了一种改进的空间聚类方法。该方法在传统密度聚类的基础上,引入节点间主导要素流向(优势流)作为约束条件^[21]。同时考虑地理邻近性和网络关联强度,以实现 对都市圈空间范围的更为科学的识别。

3.3 技术路线

本数据集研发流程主要包括 3 个步骤:选取都市圈中心城市,全国人流联系网络构建及优势流分析,都市圈空间范围识别(图 1)。

3.3.1 都市圈中心城市确定

综合《都市圈国土空间规划编制规程》^[22]以及《关于培育发展现代化都市圈的指导意见(发改规划〔2019〕328号)》^[23]中对都市圈中心城市的界定,本文优先选取中心城区人口在 300 万以上的 7 个超大城市、14 个特大城市以及 14 个 I 型大城市作为都市圈的中心城市。同时,依据申明锐等^[24]的研究,在西部地区可适当降低都市圈中心城市的常住人口规模门槛。因此将中心城区常住人口在 100–300 万的 12 个西部 II 型大城市也选为都市圈的

¹ 百度地图慧眼. <https://huiyan.baidu.com>.

中心城市。这些中心城市人口虽在 300 万以下，但其对周边区域产生较大的辐射带动作用，且多为省会城市或城市群核心城市。最终选取得到 47 个都市圈中心城市。

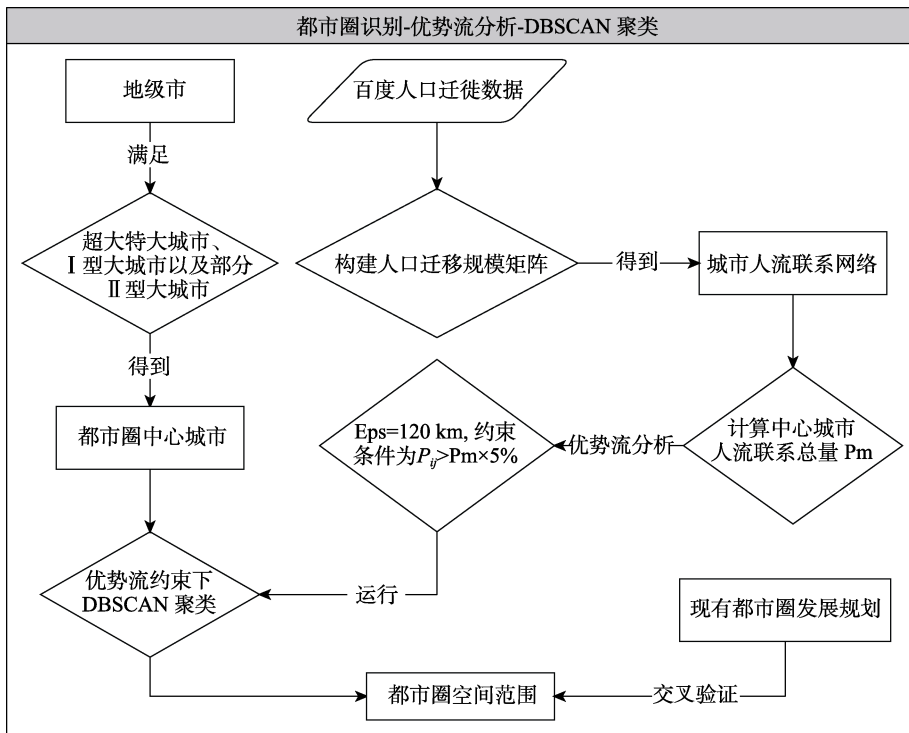


图1 数据集研发技术路线图

3.3.2 城市人流联系网络构建

首先对所获取百度人口迁徙数据进行预处理后，由逐日人口迁入迁出数据计算得到人口迁移规模矩阵。进一步利用公式 1 得到人流联系强度矩阵，并对其进行可视化表达，利用自然断裂点法将其划分为不同的等级，得到城市间人流联系网络空间分布格局并进行优势流分析。

3.3.3 都市圈空间范围识别

本数据集结合现有规划都市圈中最大城际平均距离，以及 1h 通勤圈的空间覆盖范围，将都市圈的识别半径设置为 120 km；将城际人流联系强度大于中心城市人流联系总量的 5% 作为优势流约束条件。在 VScode 平台的 node.js 运行环境中，使用 JavaScript 运行了改进的 DBSCAN 聚类算法。即以中心城市点集 D 中的点 p 为起点，若 p 的 Eps 邻域内城市数量超过 $Minpts$ ，则 p 为核心对象，创建以 p 为核心的簇并将其 Eps 邻域内的城市加入，同时引入优势流阈值作为簇扩展的约束条件，实现中心城市都市圈空间范围的识别。

4 数据结果与验证

4.1 数据集组成

本数据集共包括 3 个方面的数据：（1）都市圈中心城市列表及分布；（2）365 个城市

间人流联系强度矩阵及人流联系网络；(3) 中国都市圈识别结果列表及分布。数据集存储格式为.xlsx 和.shp。

4.2 数据结果

经过筛选共得到 47 个都市圈中心城市(图 2), 其中包括 7 个超大城市、14 个特大城市、14 个 I 型大城市以及 12 个 II 型大城市。可以看出, 中心城市多位于胡焕庸线东南侧, 西部地区除一些省会城市外, 多数城市的辐射带动能力弱。

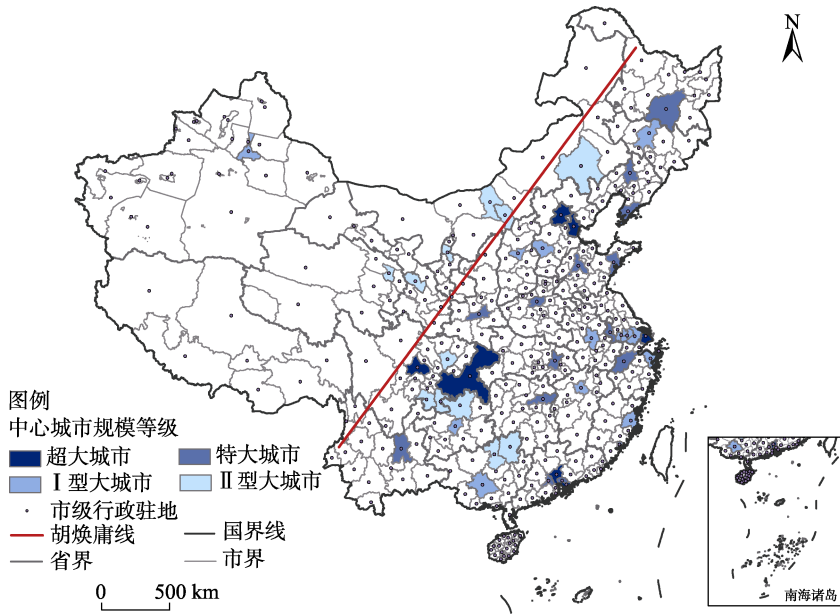


图 2 都市圈中心城市分布图

(参照审图号 GS(2024)0650 的标准地图制作, 底图无修改)

由图 3 可知, 全国人流联系网络整体呈现东密西疏和东强西弱的分布格局, 重要城市群(长三角城市群、京津冀城市群、珠三角城市群、成渝城市群等)为高等级联系流的主要集中区, 并以城市群核心城市(北京、上海、深圳、成都、重庆等)为中心向外辐射。部分省会城市同样为高等级联系流的发出地, 如西安、昆明、长沙、武汉等。各等级联系流的规模分布不均衡, 高等级的联系流占比较少且多为短距离联系流, 低等级的联系流分布范围广泛且多为长距离联系流。

通过设定都市圈识别半径及优势流约束条件, 都市圈的识别结果如图 4 和表 2 所示, 共得到 37 个都市圈。因上海与全国多数城市间人流联系紧密, 且与周边的城市未形成明显的优势流, 所以并未将上海纳入都市圈空间范围的识别。总体来看, 中国都市圈空间分布呈东南密、西北疏的总体格局, 除乌鲁木齐都市圈、西宁都市圈、兰白都市圈、银川都市圈和呼和浩特都市圈外全部位于胡焕庸线东南侧。同时京津冀城市群、长三角城市群、珠三角城市群以及成渝城市群等重要城市群是都市圈分布密集区。都市圈呈现单中心、双中心、多中心并存现象, 其中广深都市圈和苏锡常都市圈为多中心都市圈, 首都都市圈、川南都市圈、贵阳都市圈和呼和浩特都市圈为双中心都市圈。这些都市圈的中心城市规模等

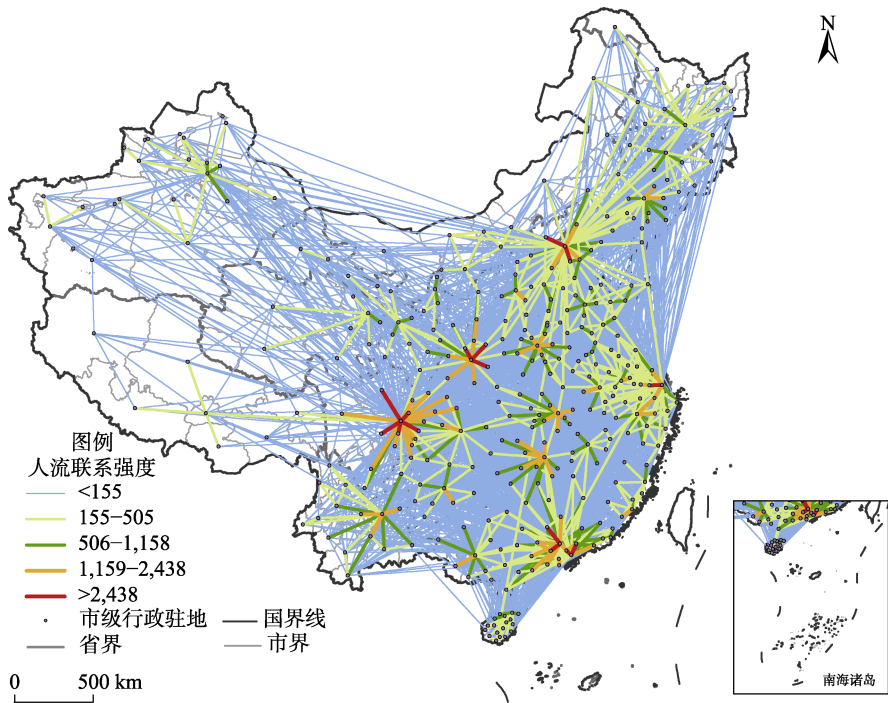


图 3 城市人流联系网络图

(参照审图号 GS (2024) 0650 的标准地图制作, 底图无修改)

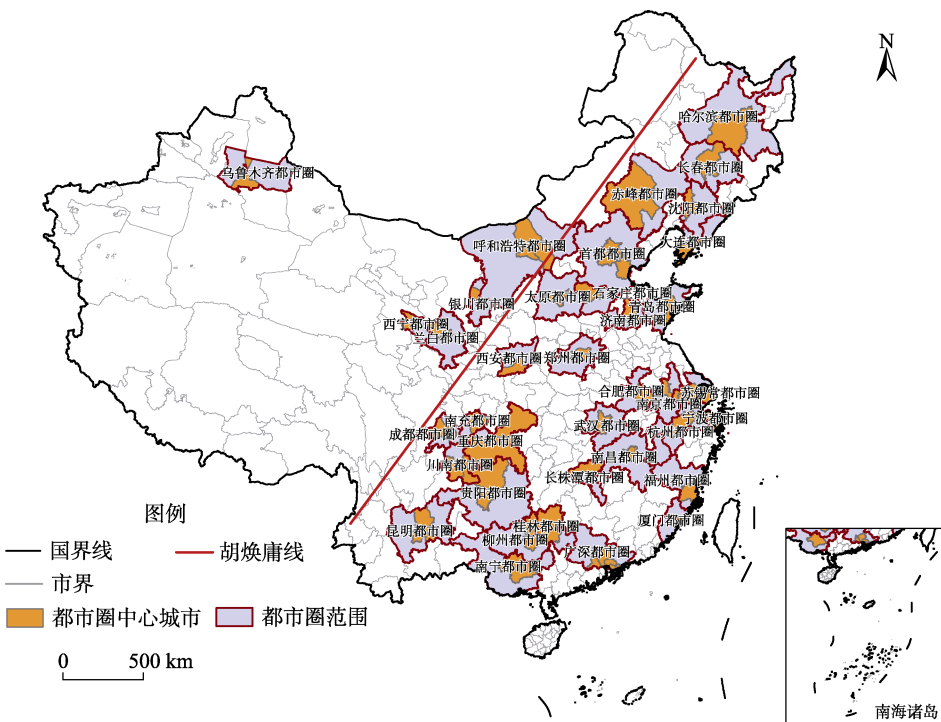


图 4 都市圈识别结果分布图

(参照审图号 GS (2024) 0650 的标准地图制作, 底图无修改)

表2 都市圈识别结果汇总表

都市圈名称	中心城市	组成城市	人口(万人)	面积(km ²)	GDP(亿元)
苏锡常都市圈	苏州、无锡、常州	苏州、无锡、常州、南通	3,362.80	25,670.78	66,225.79
南京都市圈	南京	南京、镇江、扬州、芜湖、马鞍山、宣城	2,587.98	37,913.25	41,809.10
杭州都市圈	杭州	杭州、湖州、嘉兴、绍兴、衢州、黄山	3,074.20	53,837.00	45,408.26
合肥都市圈	合肥	合肥、淮南、六安、蚌埠	2,058.70	38,335.00	19,844.19
宁波都市圈	宁波	宁波、舟山	1,095.30	11,216.00	20,373.90
南昌都市圈	南昌	南昌、九江、抚州、宜春、上饶	2,595.40	86,516.59	21,427.15
长株潭都市圈	长沙	长沙、株洲、湘潭	1,693.31	28,069.70	22,128.24
武汉都市圈	武汉	武汉、鄂州、黄石、黄冈、孝感、咸宁、仙桃	3,166.11	53,657.05	33,398.22
福州都市圈	福州	福州、莆田、南平、宁德	1,748.30	55,828.24	23,671.79
厦门都市圈	厦门	厦门、漳州、泉州	1,934.00	25,315.61	27,747.59
广深都市圈	广州、佛山、东莞、深圳	广州、佛山、东莞、深圳、惠州、清远、肇庆	7,147.68	60,942.97	104,785.73
南宁都市圈	南宁	南宁、钦州、贵港、防城港、崇左、百色	2,408.89	103,366.00	15,096.76
柳州都市圈	柳州	柳州、河池、来宾	951.99	65,500.00	5,385.24
桂林都市圈	桂林	桂林、贺州	696.12	39,552.64	3,481.81
贵阳都市圈	贵阳、遵义	贵阳、遵义、铜仁、毕节、安顺、黔南、黔东南	3,263.93	149,418.00	19,477.66
重庆都市圈	重庆	重庆、广安	3,517.17	88,744.00	33,808.65
川南都市圈	宜宾、泸州	宜宾、泸州、自贡、内江	1,438.4	35,273.18	10,661.07
成都都市圈	成都	成都、资阳、德阳、眉山	3,009.10	33,104.00	29,756.75
南充都市圈	南充	南充、遂宁	823.70	17,822.25	4,731.96
昆明都市圈	昆明	昆明、曲靖、玉溪、楚雄彝族自治州	1,891.40	93,671.95	16,553.44
西安都市圈	西安	西安、咸阳、铜川、渭南	2,253.62	37,304.51	19,065.60
郑州都市圈	郑州	郑州、新乡、焦作、开封、许昌、平顶山、洛阳	4,373.10	54,276.00	35,323.77
济南都市圈	济南	济南、淄博、泰安、聊城	2,533.98	32,599.45	25,201.74
青岛都市圈	青岛	青岛、潍坊、日照、烟台	2,974.58	46,750.90	38,262.38
大连都市圈	大连	大连、营口、丹东	1,187.80	33,027.81	12,085.00
沈阳都市圈	沈阳	沈阳、铁岭、抚顺、本溪、辽阳、鞍山	1,925.10	59,535.77	14,937.50
哈尔滨都市圈	哈尔滨	哈尔滨、牡丹江、大庆、绥化、伊春、佳木斯	2,003.63	213,467.00	12,505.70
长春都市圈	长春	长春、吉林、四平、辽源、松原	1,728.63	92,989.00	11,375.87
赤峰都市圈	赤峰	赤峰、通辽、朝阳	946.38	169,255.14	5,161.50
呼和浩特都市圈	呼和浩特、包头	呼和浩特、包头、鄂尔多斯、巴彦淖尔、乌兰察布	1,172.17	251,500.00	17,401.71
银川都市圈	银川	银川、吴忠、石嘴山	504.61	35,735.38	4,438.78
首都都市圈	北京、天津	天津、北京、张家口、承德、保定、廊坊、沧州、唐山	7,233.97	160,101.00	94,878.72
石家庄都市圈	石家庄	石家庄、衡水、邢台	2,225.43	35,701.00	12,941.10
太原都市圈	太原	太原、忻州、阳泉、晋中、吕梁	1,605.84	74,238.63	12,720.51
兰白都市圈	兰州	白银、兰州、临夏、定西	1,049.00	60,969.00	5,728.28
西宁都市圈	西宁	西宁、海东	380.55	20,860.00	2,467.56
乌鲁木齐都市圈	乌鲁木齐	乌鲁木齐、昌吉回族自治区	576.15	87,300.00	7,011.39

级相当,呈现出强强联合或一主多副的格局。37个都市圈的GDP总量所占全国比例为64.76%,其中特别突出的为广深都市圈和首都都市圈,两者均为发展较为成熟的都市圈,GDP总量更是在全国都市圈中处于领先地位,常住人口规模亦居前列,为人流网络的高等级联系流集聚区,具有高度的经济活跃度与人口吸引力。

4.3 数据结果验证

将都市圈识别结果与国家已批复的都市圈发展规划进行对比验证,可以发现大部分都市圈与所识别都市圈的空间范围是高度重合的,如成都都市圈、重庆都市圈、长株潭都市圈、西安都市圈等;也存在一些有所区别的都市圈,如广深都市圈,因广州都市圈和深圳都市圈城市之间人流联系紧密,因此识别到同一都市圈空间范围内,这也印证了利用优势流约束下的DBSCAN聚类进行都市圈空间范围识别具有科学性。此外,本研究识别出20个尚未获得国家批复的都市圈,这些都市圈同样具有较大的人口经济规模,内部人流联系较为紧密,具备培育现代化都市圈的潜力,如川南都市圈、呼和浩特都市圈、大连都市圈、南宁都市圈等。

5 讨论和总结

本数据集基于百度人口迁徙数据构建城市人流联系网络,运用优势流约束下的DBSCAN聚类算法,通过构建“中心城市选取-优势流分析-动态聚类”的综合识别框架对中国都市圈进行科学识别。由此识别的都市圈与现行规划的核心发展区域相契合,所覆盖空间范围更为完整,体现出较强的科学性和适用性,可为我国都市圈范围划定及规划政策制定提供方法参考和科学依据。未来可进一步纳入技术流、信息流等多维要素流以完善研究方法。此外,还可基于都市圈识别结果开展竞争力评价与发展能级划分,针对不同发展阶段制定差异化规划建设方案及相关政策。

作者分工:李娟负责数据可视化和数据论文的撰写;张扬负责数据集开发的总体设计,并对论文进行整体把关;唐松负责采集和处理人流联系数据及其他相关基础数据;刘旭宏完成了都市圈空间范围识别模型与算法的设计与实现。

利益冲突声明:本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献

- [1] 周一星,史育龙. 建立中国城市的实体地域概念[J]. 地理学报, 1995(4): 289-301.
- [2] 钮心毅,李凯克. 跨城功能联系视角下的都市圈国土空间规划实施监测[J]. 资源科学, 2021, 43(2): 380-389.
- [3] 张京祥,胡航军. 新发展环境下的都市圈发展、规划与治理创新[J]. 经济地理, 2023, 43(1): 17-25.
- [4] 袁家冬,周筠,黄伟. 我国都市圈理论研究与规划实践中的若干误区[J]. 地理研究, 2006, 25(1): 112-120.
- [5] 张艺帅,黄建中,王启轩等. “尺度重组”与“元治理”视角下我国都市圈治理模式的建构思路研究[J]. 规划师, 2023, 39(4): 19-27.
- [6] 张京祥,邹军,吴启焰等. 论都市圈地域空间的组织[J]. 城市规划, 2001(5): 19-23.
- [7] 汪光焘,李芬,刘翔等. 新发展阶段的城镇化新格局研究——现代化都市圈概念与识别界定标准[J].

- 城市规划学刊, 2021(2): 15–24.
- [8] 马丽亚, 修春亮, 冯兴华. 多元流视角下东北城市网络特征分析[J]. 经济地理, 2019, 39(8): 51–58.
- [9] Stone, L. O. On the correlation between metropolitan area in-and out-migration by occupation [J]. *Journal of the American Statistical Association*, 1971, 66(336): 693–701.
- [10] Feria, J. M., Casado-Diaz, J. M., Martinez-Bernabeu, L. Inside the metropolis: the articulation of Spanish metropolitan areas into local labor markets [J]. *Urban Geography*, 2015, 36(7): 1018–1041.
- [11] 杨永春, 张从果, 吴文鑫. 中国西部地区大都市圈发展规划研究——以兰州大都市圈规划为例[J]. 城市规划, 2005(4): 23–29.
- [12] 邹军, 陈小卉. 城镇体系空间规划再认识——以江苏为例[J]. 城市规划, 2001(1): 30–33.
- [13] 杨延, 尹丹, 刘紫玫等. 基于大数据的流空间研究进展[J]. 地理科学进展, 2020, 39(8): 1397–1411.
- [14] 李娟, 张扬, 唐松等. 基于优势流约束下聚类方法(DBSCAN)计算的中国都市圈数据集[J/DB/OL]. 全球变化数据仓储电子杂志, 2025. <https://doi.org/10.3974/geodb.2025.08.01.V1>. <https://cstr.escience.org.cn/CSTR:20146.14.2025.08.01>.
- [15] 全球变化科学研究数据出版系统. 全球变化科学研究数据共享政策[OL]. [https://doi.org/10.3974/dp.policy.2014.05\(2017年更新\)](https://doi.org/10.3974/dp.policy.2014.05(2017年更新)).
- [16] 国家统计局城市社会经济调查司. 中国城市统计年鉴(2020)[M]. 北京: 中国统计出版社, 2021.
- [17] Zhan, C., Tse, C. K., Fu, Y. X., et al. Modeling and prediction of the 2019 coronavirus disease spreading in China incorporating human migration data [J]. *Plos One*, 2020, 15(10): e0241171.
- [18] 张扬, 赵银兵, 赵晶晶等. 基于人流-物流-信息流的成渝城市群网络结构特征研究[J]. 地域研究与开发, 2023, 42(6): 40–45.
- [19] 李新延, 李德仁. DBSCAN 空间聚类算法及其在城市规划中的应用[J]. 测绘科学, 2005(3): 51–53+5.
- [20] 张扬, 李娟, 王兴平. 多维要素流视角下成渝城市群空间结构特征研究[J]. 西部人居环境学刊, 2024, 39(3): 14–20.
- [21] Zheng, L., Long, F., Zhang, S. Comparison of the spaces of call and traffic flows: an empirical study of Qianzhong urban region, China [J]. *Cities*, 2020, 107, 102927.
- [22] 中华人民共和国自然资源部. 《都市圈国土空间规划编制规程》正式实施[EB/OL]. (2024-04-02) [2025-07-30]. https://www.mnr.gov.cn/dt/ywbb/202404/t20240402_2841150.html.
- [23] 国家发展改革委. 关于培育发展现代化都市圈的指导意见[EB/OL]. (2019-02-19) [2025-07-30]. <https://zfxxgk.ndrc.gov.cn/web/iteminfo.jsp?id=16110>.
- [24] 申明锐, 王紫晴, 崔功豪. 都市圈在中国: 理论源流与规划实践[J]. 城市规划学刊, 2023(2): 57–66.