

# 中国省际信息“流空间”的结构与特征： 地域文化力的交互与内卷\*

陈云松 贺光烨 张柏杨 潘 雨

**提 要：**基于“流空间”理论和广义文化视角，本文提出省际信息“流空间”的文化力特征指标以及测量方法，并就其空间等级结构和宏观影响机制进行计量分析。我们利用2010—2016年各省公众在“百度”中搜索他省频率的“数据对”，生成省际信息流的强度积和强度比，以此构建信息“流空间”的省域文化力概念指标：文化“交互度”和“内卷度”。可视化分析显示，省域文化互动呈明显的多中心结构，而省际相互关注具有高度不对称性。聚类分析表明，在中国省际信息“流空间”内，文化“交互度”和“内卷度”均存在清晰的四层等级体系，但两者和“胡焕庸线”的关联则存在截然相反的地理分野。基于动态面板广义矩模型的计量分析进一步表明，两大指标分别受到人口、经济、社会、文化等一系列宏观发展因素的影响，而影响的途径和方向则大相径庭。

**关键词：**流空间 文化交互 文化内卷 大数据 互联网

如同书籍是印刷术发明以来直至20世纪人类文化最正式和最重要的载体一样，互联网是21世纪承载和展示人类文化更为新颖和丰富多彩的巨大信息母舰。从尼罗河的楔形文字到16世纪古登堡印刷术出现，从甲骨文到竹简时代再到宋代毕昇发明印刷术，降至20世纪，数千年来人类通过书籍实现了知识和精神的传播。而自20世纪晚期互联网横空出世以来，以二元编码为基因的网络信息流，已把连接对象从单纯的书写者和阅读者，迅速扩展到政治、经济、文化和社会生活全域的多元参与个体：观察者、引导者、投票者、购买者、娱乐者、社会动员者……如果说书籍所

---

\* 本文系国家社会科学基金重大项目“大数据驱动的网络社会心态发展规律与引导策略研究”（项目编号：19ZDA149；主持人：陈云松）。感谢匿名评审专家的评审意见。文责自负。

承载的，更多是知识、价值观、意识形态等狭义的文化，那么互联网的信息空间里所流通着的内容，则是更为广义的“大文化”，映射了人类社会历史实践过程中所创造的物质财富和精神财富的总和。恰因如此，无论是以文化还是以空间研究为志业的学者，都应有强烈的使命感，把嵌入和流通在互联网中的海量信息，作为审视剖析的重要对象。

在“流空间”（Space of Flows）理论的创始者卡斯特尔看来，信息技术的发展，使得人类社会不同区域的空间联系越来越有可能。早在90年代，“流数据”就已成为区域和空间研究的富矿，只是主要研究均集中在城市空间体系分析。例如，学者们先后采用航空客流、互联网域名、IP地址、企业布局、人才流动和地域经济联系作为区域空间连接的测度（Alderson & Beckfield, 2004; Beaverstock, 2007; Taylor, 2001; Taylor & Derudder, 2004; 李敬等, 2014）。

尽管这些研究利用组织或基础设施数据对“流空间”理论进行了实证分析，但有关互联网的“流数据”却因难以获得而少有学者问津（Alderson & Beckfield, 2004; Derudder, 2006）。随着大数据在学界登堂入室，近年来国内学者对互联网数据的“流空间”已经着手进行了初步探索（陈映雪等, 2013; 熊丽芳等, 2013, 2014）。不过，这些研究多为局部、小样本的城市分析。更重要的是，学界尚未从大文化的角度来进行信息“流空间”结构的分析，也未能进一步深入探讨影响“流空间”宏观结构特征的政治、经济和社会因素。

本文聚焦的是基于中国大陆互联网“流数据”的省际信息空间结构体系。这一体系，是互联网中各省域主体经由信息互动所创造出的具有节点、连接和层级结构的典型“流空间”。对该体系进行社会学角度的空间分析之所以重要，是因为省级行政主体（省、少数民族自治区和直辖市）是中国国家治理的第一层级区域主体，而省份在互联网中的互动势能与动能，特别是基于搜索引擎门户的信息流的相互辐射和穿透，实际上是省域广义“文化力”（或软实力）这一宏观、抽象概念的科学测量。依赖传统的问卷方法，我们几乎无法实现这一目标。在本文中，借助中文互联网中最具代表性的搜索引擎（百度）所记录的海量信息，对省际信息“流空间”体系进行观察，对该空间体系所映射出的文化力指标进行理论构建、测量描述和机制分析。具体而言，我们基于2010—2016年中国各省公众在“百度”中搜索他省名称的两两“大数据对”，构建出省际信息的两两“互动”和“穿透”强度，进而加总构建各省的总体文化“交互度”和“内卷度”两大文化力指标。基于这两个指

标,我们用聚类方法对“流空间”进行等级划分和描述,并用动态面板模型(dynamic panel model)来分析省域文化力的两大指标,受到哪些经济社会文化要素的影响。本文首次提出的地域文化“交互”和“内卷”两大概念,为从社会科学的视角理解“流空间”,增加了新的理论维度和分析对象。

## 一、文献回顾与分析框架

“文化软实力”在当今世界已经成为国家或地区综合实力竞争的重要因素,以及衡量社会文明程度和人民生活质量的重要标尺。作为一种文化的凝摄自固力、竞争力、传创力和感召力的整合表达,文化力的表现形式往往依赖于“场景”,一个社会成员共同的生活环境和地理空间,并在这一空间下,“场景”对其成员的价值观、生活方式产生影响。从而,“场景”不仅可以表现为一定的区域空间结构,更可以为社会成员提供一种归属感和文化需求,令身处其中的成员从生活方式及情感表达上获得社会身份。也正是“场景”下的这种文化内蕴不断形塑着地域间的互动模式,与地域长期持续的发展紧密相联。

国际文献中,对地域互动空间体系的研究主要围绕城市网络来进行。总体来看,城市体系的研究范式和方法主要经历了三个重要转变:由中心论到网络化,由属性到关系,由经济联系到社会文化多元联系。

### (一) 由地理中心论到区域网络化转向

德国城市地理学家克里斯泰勒 20 世纪 60 年代最早提出了“中心地理论”,也即中心城市作为区域人口、职能、行政等的集中地域,担负起向周边辐射的功能,周边城市作为其腹地而存在。在等级关系下,城市间的联系通常是单向的,由中心流向边缘城市,形成一个类六边形的向心式区域系统。进入 20 世纪八九十年代,相关研究由国家城市体系扩展为世界城市、全球城市,并逐渐形成世界城市体系的思想。随着全球化、信息化进一步发展,地域空间结构呈现从等级化向网络化发展的趋势。城市空间的组织形式由静止、封闭、等级性的中心地模式逐渐向多中心、扁平化、流动性的网络模式转变(Castells, 2010; Hall & Pain, 2006),相关的研究范式也开始出现了相应变化(Taylor & Derudder, 2004)。

## （二）由区域属性数据到区域联系数据的转向

20世纪90年代著名学者卡斯特尔提出“流空间”理论（Castells, 1996）。其主要理论构成有三：第一，“流空间”的物质基础是交通、通信等基础设施网络，技术发展使得不同区域的联系越来越成为可能。第二，“流空间”具有节点和中心，不同属性的城市承担不同功能，影响力呈等级化状态。第三，“流空间”的支配者实际是管理精英，也即“流空间”是人的空间组织。基于上述理论，实证研究从以航班货运量、港口吞吐量、公路铁路航空客运流等数据来量化城市网络（Mahutga et al., 2010; Matsumoto, 2004），逐步发展到以通信包裹、互联网域名、容量及IP地址数据进行分析（Choi et al., 2006; Malecki, 2002; Townsend, 2001），并开始聚焦企业“流空间”层级体系和精英流动（Alderson & Beckfield, 2004; Beaverstock, 2007）。

## （三）由经济联系到社会文化多元联系的转向

经济联系只是区域体系互动的一个部分，政治、社会、文化等多元联系也应纳入“流空间”的研究之中（Robinson, 2005）。在多元联系的视野下，理论学者或区分了与人相关的“软网络”和以基础设施为主的“硬网络”（Malecki, 2002），或基于流的形态（人文、信息和物质）和功能（文化、经济、政治和社会）建构起区域总体联系拓扑结构（Smith & Timberlake, 1995）。实证学者则进一步拓展空间关联的分析对象，包括非政府组织办公网络等一系列新的空间互动结构也开始被纳入学术分析的视野之中。

随着区域空间研究的积累，特别是“流空间”理论的进一步传播，国内学者也在21世纪初逐步开始关注相关的区域体系研究。在理论研究方面，中国学者重点关注了互联网、企业空间组织等对区域空间结构的表征和影响（汪明峰、宁越敏，2004；宁越敏，2011；顾朝林，2011）。在实证研究方面，学者一方面利用经典的城市间基础设施硬件指标（主要是铁路、航空、互联网通信节点）等来分析中国城市体系（汪明峰，2006；吴威等，2009；梁辉，2008），另一方面也开始使用连锁企业网络、母子企业、同行业网络等来分析区域互动和城市体系（张闯、孟韬，2007；金钟范，2010；谭一洛等，2011；尹俊等，2011；武前波、宁越敏，2012；路旭等，2012；曹有挥、陈国伟，2014）。然而，利用互联网虚拟空间的数字信息进行区域“信息流空间”分析，目前研究还相对有限。

为数不多的“信息流空间”研究，主要集中在城市地理学界。甄峰等（2012）

和陈映雪等（2013）分析了2011年1020名新浪微博用户的互粉关系，以此来研究城际信息互动。熊丽芳等学者（2013）获取2009年和2012年长三角城市两两间的搜索关注度数据，以此来计算城际信息流并分析城市网络时空演变。使用同样的方法，熊丽芳等（2014）分析了长三角、珠三角、京津冀三大经济区之间城市跨区搜索热度（2009—2012）。蒋大亮等（2015）和于欣洺、吴相利（2015）则分别获取长江中游三大城市群之间和东北地区城市之间的搜索热度，探讨城市网络格局在2011年到2014年之间的变化。上述研究，从互联网“信息流空间”的角度为我国区域体系分析注入了新的思路，但也存在以下几个方面的不足：

第一，在研究对象上，受制于搜索引擎数据的提取难度，现有研究仅针对我国局部地区（如长三角、珠三角等）进行分析，缺乏对全国性地域信息流的总览；而利用新浪微博数据进行的全国性研究，则存在代表性严重不足的问题（50个城市，每个仅取样20人）。第二，在研究层级上，相关研究全部集中于城市层面，对更为宏观的省域互动的研究仍旧是空白。且时间跨度多为1—3年数据，无法对近年来我国信息流空间的动态特征进行稳健的纵贯分析。第三，在研究深度上，对信息流所表征的广义文化概念挖掘不深。一方面，现有研究大多停留在描述分类阶段，未能通过严格的计量方法对其经济社会影响因素进行深入的机制分析；另一方面，现有研究也未能从大文化的视角对地域之间信息流的互动和差异化格局进行具有社会科学意蕴的探讨。本文试图通过大规模数据分析与计量模型相结合，来弥补这些不足。

## 二、“流空间”文化力指标的构建

### （一）大文化的量化思路

据中国互联网络信息中心（CNNIC）2019年8月发布的第44次《中国互联网络发展状况统计报告》，截至2019年6月，我国网民规模达8.54亿，普及率达到61.2%，<sup>①</sup>超过全球平均水平约4个百分点，超过亚洲平均水平约9.1个百分点。而其中搜索引擎用户规模达到约6.4亿，占总体网民的83.8%。<sup>②</sup>在网民规模继续增长的同时，

① 网址为 [http://www.cac.gov.cn/2019-08/30/c\\_1124938750.htm](http://www.cac.gov.cn/2019-08/30/c_1124938750.htm)。

② 全文PDF地址为 <http://www.cnnic.net.cn/hlwfzyj/hlwzbg/hlwtjbg/201803/P020180305409870339136.pdf>。

我国互联网模式不断创新、线上线下服务融合进一步加速，“万物互联”的特征进一步凸显。也因此，网络信息流作为网民行为、观念、需求的集中体现，蕴含了重要的社会、经济、文化的全方位信息。而网络搜索引擎作为使用最为频繁、覆盖面最广的互联网应用工具，可以广泛捕捉地区之间的网络信息交互流。这种信息流，从数据角度仅是互联网搜索行为的记录，但恰恰因为其内容的广谱性，使得其可以成为地域之间“大文化”互动强度的代理变量（proxy），从而可以被用来展现中国省域层次的总体文化联系，进而借助其来探究省域信息流空间结构的特征和规律。需要注意的是，本文从广义文化的角度来理解“文化力”，将其理解为一种地域综合性的软实力，是一种结合了政治、经济、文化和社会诸方面的总体知名度、关注度。

在大数据分析中，一个基于逻辑直觉的假设就是，无论是一个单词、一位名人、一座城市、一个省份甚至一个国家，如果它在文化大数据载体（海量书籍、新闻媒体或互联网搜索引擎）中出现得越多、被检索得越多，越被关注，就在相应的范围内具有越高的知名度、影响力、被关注度。国际国内最新的大数据实证研究，已经借助这一思路在文化史、城市史、社会分层研究、自杀研究和互联网文化领域进行了大量探索，积累了丰富的经验和测量模式（Chen et al., 2016；陈云松等，2017）。因此，我们使用百度搜索引擎作为本文的数据分析来源：省份 A 的公众利用百度对省份 B 的省名搜索量高低，被用来作为省份 A 对省份 B 的政治、经济、社会、文化等全方位关注程度的指标。<sup>①</sup> 百度搜索作为全球最大的中文搜索服务提供商，在中文搜索中占据绝对主导地位。中国所有提供搜索引擎的门户网站中，八成以上是由百度提供搜索引擎技术支持。因此，使用百度搜索引擎数据不会受到用户选择性问题的困扰。

具体而言，我们以中文省份名（含自治区、直辖市，包括全称、简称）作为检索关键词，获取 2010 年 1 月 1 日到 2016 年 12 月 31 日（7 年间）的中国大陆 31 个省份两两相互搜索的热度指标——百度指数。百度指数是以“网民在百度的搜索量为数据基础，以关键词为统计对象，科学分析并计算出各个关键词在百度网页搜索中搜索频次的加权和”（百度指数，2015）。它并非简单的每日搜索数量的加总，而

<sup>①</sup> 在实际搜索时，省 A 对省 B 的关注，往往并不一定是以省 B 为直接搜索关键词，而是省 B 内的市、县、乡为搜索关键词。因此，我们的测量很大程度上采用的是自上而下的测量，也即，对于某省内的市、县、乡的搜索，我们会将其匹配到相应的省份。因此，对某省的关注度实际同时包含了对该省内的市、县、乡的关注度。

是和谷歌趋势指数一样,由其技术团队处理计算后生成。尽管这类指数算法未全部公开,但由谷歌、百度等大型平台提供的科学数据,本身经由专业团队处理,在计算过程中考虑了搜索总量和分布等特征并进行后台加权计算,较之研究者通过各种非官方渠道获得部分文本资料再自行检索计算更为可信(百度指数,2015)。在大数据分析文献中,使用这类大型自媒体平台 and 搜索引擎提供的官方指标(如谷歌趋势指数、推特第三方词频工具等)已经成为主流的方法(Curme et al., 2014; Bryden et al., 2013; 陈涛、林杰, 2013; Scheitle, 2011; Mellon, 2014; Nuti et al., 2014)。<sup>①</sup>

## (二) 流空间的文化力指标:交互度和内卷度

在两两省份的信息互动层面,考虑到各省份人口数量的差异,我们将一定时段内某一省份*i*在百度中被另一省份*j*网民检索的次数与省份*j*的人口数量之比,定义为省份*i*对于省份*j*的“被搜索热度”。而中国省域之间存在巨大的地理距离差异,省份*i*被省份*j*真正关注的程度,应该同时考虑真实空间的关注度和虚拟空间的关注度,前者与地理空间距离相关,省*i*与省*j*距离越近,省*j*对省*i*的在真实空间的关注也越多<sup>②</sup>,而后者则与虚拟空间的“被搜索热度”相关,被搜索得频率越高,省*j*对省*i*在虚拟空间的关注越多,因此,省份*i*被省份*j*真正关注的程度实际是“被搜索热度”与两省距离的一个函数。假定上海在江苏的被搜索热度为100,在北京也为100,考虑到北京的地理位置较之江苏更为远离上海,则北京因距离遥远、直接信息沟通缺少而带来的地缘性关注相对江苏就更少,即便虚拟空间文化作用力相同,真实的上海从江苏获得的关注度应该大于其从北京获得的关注度。<sup>③</sup> 考虑这种距离

① 百度指数(2015)则明确指出:“百度指数在防作弊(反作弊)方面作了大量工作,研究并建立了一整套相应的防作弊(反作弊)技术体系,力求将作弊现象减少到最低限度。用户发现任何作弊情况(嫌疑)可随时点击百度指数意见反馈进行举报。为最大限度保证百度指数整体信息公平、公正,对于已经有明确证据证明使用作弊手段来提高排名的关键词,百度指数会对其数据进行很大幅度的降权处理。”

② 分析假定省份关注度与省份间的距离相关,一个重要的前提在于,对于省份A,对其的关注度主要由与省份A邻近的省份所驱动。其中最重要的驱动力之一便是人口迁移。不少研究显示,从90年代至今,我国省级人口迁移流动性不断增强,尤其在东南沿海地区,但是总体上仍旧呈现规律:省际人口净迁移流的规模随距离增加的衰减速度较快,距离越近分布越多,数量越大(王桂新,1993;闫庆武、卞正富,2015;于文丽等,2012)。

③ 我们用“被所搜热度”体现了这两个方面的关注,因为从个体角度出发,关键词的搜索,既存在满足真实空间的互动需求而进行的搜索,即基于真实空间互动产生的相关互联网关注,也存在对其经济、政治、历史、文化感兴趣而进行的搜索,即超越地理边界的互联网关注。这两方面尽管侧重点存在差异,但搜索逻辑相似,搜索的大多是经济、政治、历史、文化、环境等一方面或者多方面在某时段处于高位,这种“高位”并不是绝对意义上的。

效应的非线性和地理衰减，我们将省份  $i$  对于省份  $j$  的“被关注度”定义为其“被搜索热度”除以两省份距离的平方根（Taylor & Derudder, 2015）<sup>①</sup>：

$$R_{jt} = \frac{C_{ijt}}{\sqrt{r_{ij}}} = \frac{S_{i_{-j}t}}{P_{jt} \sqrt{r_{ij}}} \quad (1)$$

在上式中， $R_{jt}$  表示省份  $i$  在第  $t$  年被省份  $j$  的关注程度，也即“被关注度”， $C_{ijt}$  为省份  $i$  在第  $t$  年在省份  $j$  的“被搜索热度”， $S_{i_{-j}t}$  表示省份  $i$  在第  $t$  年被省份  $j$  居民检索的百度指数， $P_{jt}$  为省份  $j$  在第  $t$  年的人口数量（稳健性检验中，我们会用网民人口而非居民人口进行人口校正）， $r_{ij}$  表示省份  $i$  与省份  $j$  之间的地理距离（用省会之间距离来代替）。<sup>②</sup> 表 1 中展示了 2010—2016 年各省份在互联网信息流空间的“被关注度”。

表 1 中国信息省级流空间的被关注度、平均省级文化交互度、内卷度

代码	省份	流空间的“被关注度” ( $\times 10^{10}$ )	总体交互度 ( $\times 10^{12}$ )	总体内卷度 ( $\times 10^8$ )
11	北京	10.61	27.89	10.43
12	天津	5.00	5.39	8.45
13	河北	2.64	3.80	12.93
14	山西	2.91	3.37	7.66
15	内蒙古	4.68	3.83	6.41
21	辽宁	5.33	6.87	24.62
22	吉林	6.35	6.09	26.10
23	黑龙江	8.63	9.19	24.26
31	上海	9.40	17.56	8.72
32	江苏	3.25	7.19	17.11
33	浙江	3.84	10.12	24.84
34	安徽	2.95	3.69	8.35
35	福建	4.07	5.98	19.79
36	江西	3.05	3.25	9.35
37	山东	5.69	11.51	11.31
41	河南	2.86	5.35	9.03
42	湖北	2.35	3.54	10.46

① 有关距离的衰减效应，除了距离平方根之外，或许还存在线性衰减、对数衰减、逆对数衰减等形式。为检验结果的稳健性，基于不同的形式的衰减，我们均构建了相应的被关注度指标，受篇幅所限，具体分析结果没有在文中显示，感兴趣的读者可以向作者索取。

② 需提及的是，由于“被搜索热度”既体现了由于真实地理空间互动而产生的出于文化需求的互联网搜索，又体现了超越了地理空间的出于文化兴趣的互联网关注，我们所计算的省份  $i$  对于省份  $j$  的“被关注度”实际仅反映了前者，是“真实被关注度”的一个低估。



续表

代码	省份	流空间的“被关注度” ( $\times 10^{10}$ )	总体交互度 ( $\times 10^{12}$ )	总体内卷度 ( $\times 10^8$ )
43	湖南	3.88	5.12	7.14
44	广东	5.10	16.23	34.55
45	广西	6.12	6.49	10.32
46	海南	9.85	6.61	7.15
50	重庆	8.37	8.58	2.89
51	四川	6.11	8.50	5.82
52	贵州	5.58	4.69	4.93
53	云南	10.95	9.80	5.43
54	西藏	18.29	4.78	2.59
61	陕西	2.73	3.62	6.46
62	甘肃	4.27	3.22	4.43
63	青海	5.11	2.03	2.81
64	宁夏	4.08	2.09	3.55
65	新疆	26.62	18.74	10.42

基于省份“被关注度”这个重要的测量概念，我们可以构建出信息流空间的两大文化指标：交互和内卷。而这两个文化指标，在以往文献中未得到应有的关注，缺乏相应的量化指标构建。

### 1. 文化交互度（互引力）

不同于单向的关注，文化的相互吸引，也即互引力是省份两两间双向吸引关系的结果，体现的是一对省份之间信息双向流动的强度。因此，省份*i*和省份*j*之间的文化互引力计算方法为：

$$A_{ij} = R_{ij}R_{ji} = \frac{C_{ij}C_{ji}}{r_{ij}} = \frac{S_{ij}S_{ji}}{P_j P_i r_{ij}} \quad (2)$$

其中， $A_{ij}$ 表示省份*i*与省份*j*之间的文化互引力。如果我们用*N*代表“他省”的总数（例如，分析中国大陆省级行政区域，则*N*为31-1=30），将其他各省和省份*i*之间的互引力加总，便得到省份*i*在第*t*年的总体对外文化交互度 $A_{it}$ ，计算公式可表达为：

$$A_{it} = \sum_{j=1}^N \frac{C_{ijt}C_{jit}}{r_{ij}} = \sum_{j=1}^N \frac{S_{ijt}S_{jit}}{P_{jt}P_{it}r_{ij}} \quad (3)$$

我们在图1的左侧展示了基于2010—2016年平均指标的各省份之间的文化互引力。从地图可见，省份之间信息流互动强度最大是京冀线。其他比较强的还有闽赣、闽鲁等流空间。经过加总计算，各省份的总体文化交互度见表1。

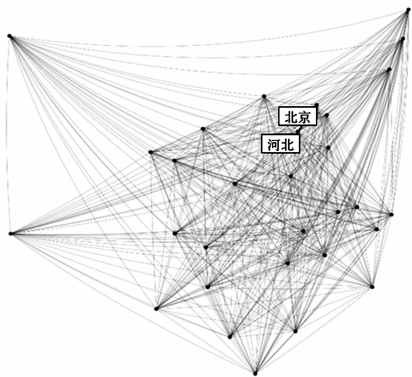


图1 中国大陆各省份信息流空间的文化交互（2010—2016年）

如何解读这种文化交互度的省域差异？在人类历史上，任何一个地域都会被赋予某种文化意涵，经济的、政治的、民族的等（白欲晓，2011）。我国幅员广阔，各省份通过相互竞争、碰撞、融合，建立、发展和延续了它们在历史上经济、政治、社会 and 狭义文化的多维度关联。从图1分析，中国省级信息流空间的互动总体格局呈如下几个特点：

第一，多中心互动模式。我国省级信息流空间的互动并非向心型的中心地模式，而是明显的多中心互动，这与流空间理论十分契合（Castells, 2010；Hall & Pain, 2006）：起码，各省和首都北京、国际化大都市上海之间，并没有明显超过和其他省份的互动强度。<sup>①</sup>

第二，空间距离影响弱化。尽管互动最强的连接仍然是河北—北京、江西—福建这样的邻接省份，但地理跨度巨大的两省之间，并非互动最小。这实际上也充分体现了互联网是对受制于地理空间的人物流互的重要补充：在互联网的世界，地理空间距离不再重要。此外，在市场经济的推动下，多中心、扁平化、流动性的地域互动网络模式，也日渐模糊了地域间的界限，弱化了以距离为单位测量的绝对空间上的邻近（孙中伟、路紫，2005）。

第三，东西梯次现象明显。我国的总体文化交互度大致呈现从东到西的递减。而这一变化梯度可能与省域的社会经济发展水平存有一定联系。值得注意的是，这

<sup>①</sup> 需提及的是，省信息流的互动与流动人口流动的格局存在相似之处，由于流动人口的第一选择通常是邻近省份和发达地区。因此，以省份（省会）为质点，做出的流动网络或者流动矩阵所体现出的多中心性可能是一种外显表象，而不一定说明信息流动是多中心的。

一递减趋势以著名的“胡焕庸线”（黑色虚线）也即黑河—腾冲线为界。胡焕庸线以东，文化交互频繁，这种更频繁的交互可以通过地图上更多深色的色块来体现，而胡焕庸线以西，文化交互程度较弱，因此从地图上看，浅色色块居多。这一现象也毫不令人吃惊：我国人口和经济社会活动主要密集分布在“胡焕庸线”以东。互联网信息的流空间特征，和真实的物流人流互动存在映射性。

## 2. 文化内卷度（穿透力）

两个省份彼此的关注度之比，则可以从另外一个角度反映某一省份对其他省份文化的相对不关心，我们将其定义为省省之间的文化穿透力：对于省份  $i$  而言，外省  $j$  对其搜索得越多，而自身对省份  $j$  搜索得越少，则省份  $i$  对省份  $j$  的文化穿透力越强。这样，我们把省份  $i$  对省份  $j$  的文化穿透力定义计算为：

$$P_{ij} = \frac{R_{ij}}{R_{ji}} = \frac{C_{ij}}{C_{ji}} = \frac{S_{i,j}P_i}{S_{j,i}P_j} \quad (4)$$

把省份  $i$  对所有省份的文化穿透力加总，我们可以得到省份  $i$  在信息流空间的总体文化内卷度  $P_{ii}$ ，表达为：

$$P_{ii} = \sum_{j=1}^N \frac{R_{ij}}{R_{ji}} = \sum_{j=1}^N \frac{S_{i,j}P_i}{S_{j,i}P_j} \quad (5)$$

这个总体的内卷度，体现的是一个省份在信息和大文化角度的相对自我眷顾和独立性。

各省份之间的文化穿透力见图 2，总体文化内卷度见表 1。在两两成对的文化穿透线中，北京对天津的穿透最强。在文化内卷度指标上，西藏和云南总体较高。考虑视觉效果，我们进一步把 31 省份分为六大片区，将各片区省份和外界省份的文化穿透展示在图 3 的下半部分。不难发现，省际穿透存在不对称性。也即，省份之间的两两相互文化关注程度有很大的不对称性或不等性；有的省份不关注其他省份，而有的省份则被外界广泛关注。前文提及的北京对天津、西藏和云南对外界诸省份均为如此：北京人对天津的互联网搜索较之天津人对北京的搜索，是相差最大的。西藏和云南则不关心外界省份，而几乎各个省份都对这两省有较高频率的搜索。

但是，这种关系的不对称性并不能用单一的理论来进行解释。就省际关注两两不对称性的来源，我们总结有三大模式：

第一，地位优势的虹吸模式：北京对天津的穿透，逻辑的直觉就是：北京在政

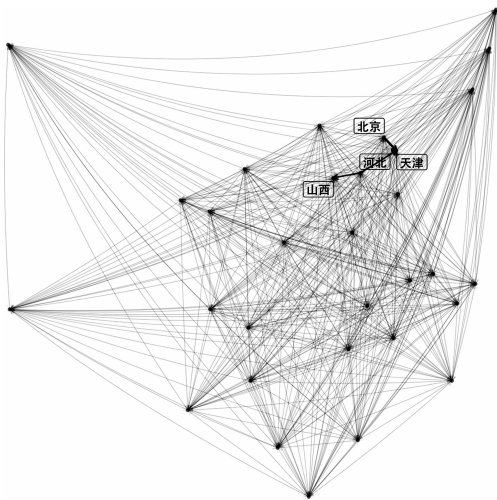


图2 省份与省份之间的文化穿透力

治经济地位上对比天津这一近在咫尺的直辖市，具有压倒性优势，形成对天津巨大的虹吸作用，导致北京人不关心天津，而天津人高度关注北京。这种所谓的“不对等关注”，在国际政治领域也有存在。学者们通常从相对权力角度阐释国与国之间不对称的双边关系。如果两国存在权力上的差异，且权力较大的国家的政策变动、发展方向等关悉权力较小的国家的切身利益时，这种差异便会导致两国在心理层面的不对称认知，从而使得权力较小的国家会对权力较大的国家报以更多关注，反之则不然（Womack，2001）。

第二，文脉渊源的吸引模式：四川对重庆的穿透，很可能是因为重庆作为直辖市，从四川独立而出，而四川的历史文脉、旅游资源等对于重庆人具有巨大的吸引力。类似地，广东对海南的穿透，也可能是类似的机制在起作用：海南本是广东的下辖行政区域。这种文脉渊源，实际上体现了文化母体对脱胎于此的文化子体的独特吸引力。

第三，文化独立性的内卷模式。福建对周边诸多省份的穿透，无法用政治经济地位优势或者文脉渊源来解释。而可能的解释是，福建地区具有多元文化特征，对部分历史传统保存较好（如家谱、宗祠、客家文化等）。这种文化特征形成了高度的独立性和内卷倾向，使得福建人对省外信息的检索较少，在指标上呈现福建对周边诸省份穿透力较强的表象。

这三种模式混合作用，在加总的层面，呈现独特的“逆胡焕庸线”现象：黑河—腾冲线以西地区的文化内卷度，总体明显高于该线以东地区。其中，西藏和云

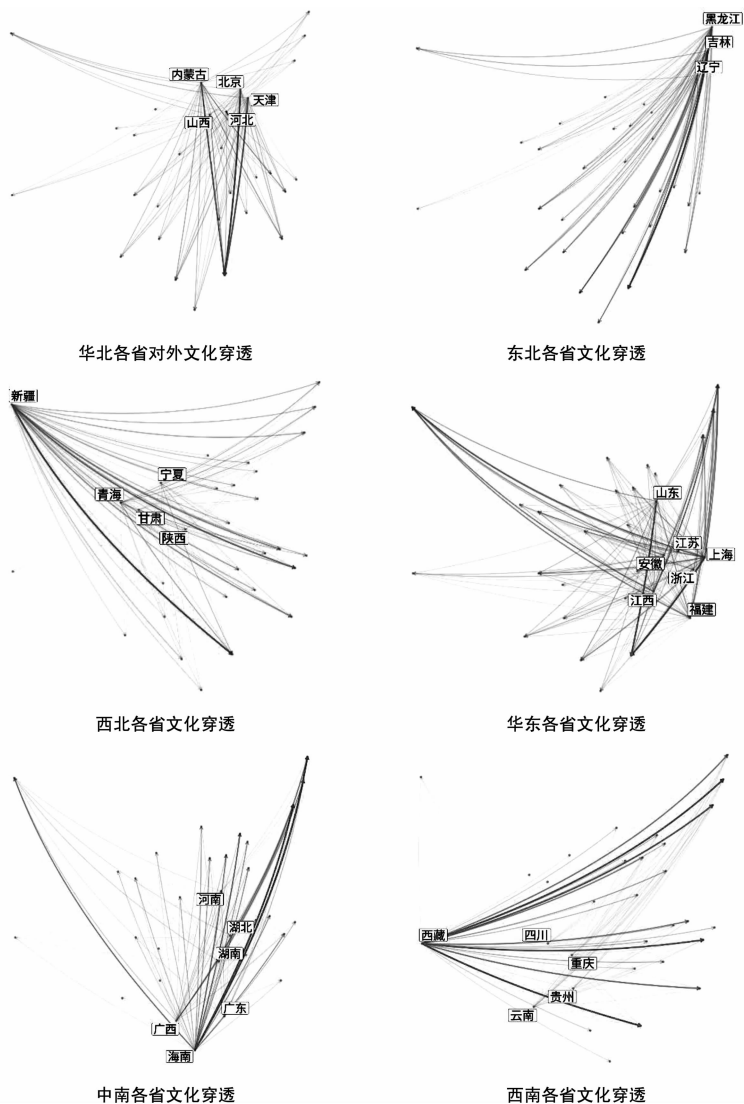


图3 中国大陆各省份信息流空间的文化内卷（2010—2016年）

南的经济地位在全国范围内并不高，但由于民族传统，具有较高的文化独立性，且两者具有的丰富文化旅游资源，使得它们对外省的关注不高，而外省对它们则高度关注。其余如新疆、青海、宁夏，也皆如此。而京津冀地区、广东、浙江等地，则显出更低的内卷水平。

### 3. 流空间文化力指标的等级结构

为进一步厘清省域信息流空间的等级结构，我们采用“聚类分析”（cluster

analysis) 方法, 分别对两大文化指标, 生成各自的流空间地位等级。其基本思路是: 首先将 31 个省份看作 31 个类, 建立相关指标的省际相似程度距离或相关系数, 然后按接近程度的大小, 将接近程度最大的两省份合并为一个新类, 得到 30 个类, 并计算新类与其他类的接近程度; 再从中找出接近程度最大的两类加以合并变成 29 个类, 如此重复进行, 每次减少一类, 直至所有样本归并成一类为止, 形成一个表示亲疏关系的谱系图, 为实际的分类提供定量的依据。具体而言, 我们采取欧几里德距离的最长距离法 (complete linkage) 进行聚类。

其中, 交互度的省级信息流空间结构等级, 具体为 (由高到低排序):

第一等级: 江苏、山东、广东、浙江、北京、上海、云南、湖北、河南

第二等级: 山西、黑龙江、河北、陕西、天津、湖南、福建、甘肃、江西、贵州、四川、辽宁

第三等级: 安徽、广西、重庆、吉林、内蒙古、宁夏、青海

第四等级: 海南、新疆

内卷度的省级信息流空间结构等级, 具体为 (由高到低排序):

第一等级: 西藏、云南

第二等级: 青海、甘肃、黑龙江、贵州、湖北、宁夏、新疆、海南

第三等级: 广西、江西、吉林、内蒙古、湖南、山西、重庆、福建、上海、陕西、天津、辽宁、山东、河南、四川、安徽、江苏

第四等级: 广东、河北、浙江、北京

我们还测试了其他聚类方法, 都获得了和以上一致或非常接近的结果。

### 三、信息流空间文化力指标的影响因素

前文提及, 省级信息流空间的两大文化力指标, 其模式特征的形成具有多种机

制和来源,受到各类省级经济社会政治和文化因素的影响。为厘清文化力的形成机制,我们将省级统计年鉴等相关数据与通过百度搜索数据生成的文化力指标进行匹配,获得了2010—2016年7期31个省级行政区的面板数据。利用这个面板数据,我们为两种文化指标分别做动态面板回归,分析影响流空间文化力两大指标的省域宏观因素。使用动态面板分析,是因为我们认为,文化力指标具有高度的路径依赖,也即上一年的文化力指标,会对下一年的文化力指标具有重要的影响。

### (一) 分析变量

在回归分析中,我们的变量如下<sup>①</sup>:

因变量:省份文化交互度;省份文化内卷度。也即前文使用百度搜索引擎计算生成的各省份年度指标。

自变量:①省份人均GDP,代表省份的经济发展水平和市场规模(Blomstrom & Lipsey, 1991);②省份基础设施水平,用每10万人拥有高速公路长度来计算;③省份城镇居民收入,以省份城镇居民的实际家庭年收入情况来衡量;④省份人均教育水平,每10万人口中的大学生人数占全省教育的百分比;⑤省份社会发展水平,以每10万人的非政府组织(NGO)数量来衡量;⑥省份旅游业的对数年收入;⑦省份人口规模;⑧省份国际知名度,该变量为五年前该省份在800万种谷歌英语数字化书籍数据中的提及次数,我们通过Python语言编程从谷歌相关网站获得。相关语料库的背景知识和词频数据的提取方法参见陈云松(2015)。

除了省份国际知名度之外,所有指标均来自不同年份的中国统计年鉴。在分析时我们均取了自然对数以期所有变量呈近似正态的分布。对于GDP数值我们以2008年为CPI基线年进行了调整。主要变量的相关描述性统计请见表2。

表2 中国大陆31省份主要变量时间序列(2010—2016年)

变量	操作化定义	缩写	均值	标准误
因变量				
文化交互度	与各省互联网关注度乘积的总和	<i>Inter</i>	8.938	1.703
文化内卷度	与各省互联网关注度之比的总和	<i>Invol</i>	8.894	1.810

<sup>①</sup> 在最初的分析中,我们考虑了农村城市移民(中国的主要移民模式)对于省份被关注的影响,我们在模型中纳入了城镇化水平。然而由于该指标与经济发展水存在多重共线性。故此,城镇化水平最终没有纳入模型。

续表

变量	操作化定义	缩写	均值	标准误
滞后因变量				
文化交互度	上年的文化交互度		.057	.661
文化内卷度	上年的文化内卷度		.054	.662
自变量				
经济发展水平	人均 GDP (可比价)	<i>GDP</i>	9.475	.485
基础设施水平	人均高速公路里程	<i>Infra</i>	4.138	.900
城镇居民收入	城镇居民家庭年收入	<i>Inco</i>	9.631	.273
人口教育水平	大学生人口占比	<i>Edu</i>	7.706	.355
社会发展水平	人均非政府组织个数	<i>Soc</i>	1.249	.358
旅游业发展水平	旅游业年收入	<i>Tour</i>	5.535	1.721
人口规模	省份年度人口数量	<i>Pop</i>	8.095	.853
国际知名度	谷歌数字化书籍词频	<i>Vis</i>	-11.277	1.010

## (二) 动态面板模型

文化交互度和内卷度的具体模型可以分别写为：

$$Inter_{i,t} = \alpha Inter_{i,t-1} + \beta_1 GDP_{i,t-1} + \beta_2 Infra_{i,t-1} + \beta_3 Inco_{i,t-1} + \beta_4 Edu_{i,t-1} + \beta_5 Soc_{i,t-1} + \beta_6 Tour_{i,t-1} + \beta_7 Pop_{i,t-1} + \beta_7 Vis_{i,t-1} + c_i + \mu_{it} \quad (5)$$

$$Invol_{i,t} = \alpha Invol_{i,t-1} + \beta_1 GDP_{i,t-1} + \beta_2 Infra_{i,t-1} + \beta_3 Inco_{i,t-1} + \beta_4 Edu_{i,t-1} + \beta_5 Soc_{i,t-1} + \beta_6 Tour_{i,t-1} + \beta_7 Pop_{i,t-1} + \beta_7 Vis_{i,t-1} + c_i + \mu_{it} \quad (6)$$

在上式 (5) (6) 中,  $c_i$  为不随时间变化的省效应,  $\mu_{it}$  为方程的随机误差项。如果我们采用固定效应模型来估计以上方程, 虽然可排除不随时间变化的省级层面干扰项, 但是由于滞后项的纳入, 固定效应模型的估计量必然有偏误。因此, 我们进而采用广义矩 (GMM) 方法进行估计 (Arellano & Bond, 1991; Holtz-Eakin, et al., 1988)。广义矩估计主要包含两种方法, 差分 GMM (简称“DGMM”) 和系统 GMM (简称“SGMM”)。具体而言, 差分 GMM, 即, 对上式进行一阶差分。为了避免滞后项与误差存在相关, 模型使用自变量和因变量的滞后阶作为“内部”工具变量来解决模型内生性问题。当然, 差分 GMM 估计可能存在弱工具变量问题 (Griliches & Hausman, 1986), 并加重测量误差对因变量的影响。而系统 GMM 正好可以克服差分 GMM 的这些缺陷。该方法相当于联立了差分方程和原水平方程, 用变量水平滞后项作为一阶差分项的工具变量, 同时一阶差分又作为水平变量的工具变量以提高



估计精度 (Arellano & Bover, 1995; Blundell & Bond, 1998)。分析中, 我们同时采用了这两种方法, 并进行了一系列标准检验, 包括 (1) 单位根检验, 确保所有变量在去除时间趋势后都是平稳时间序列; (2) 二阶自回归检验 [AR (2)], 用以观察  $\mu_{it}$  是否存在自相关; 以及 (3) Henson/Sargan 过度识别检验, 用以验证工具变量的有效性。其中, AR (2) 和 Henson/Sargan 过度识别检验的结果将与模型回归结果在下一节中一起展示。

表 3 中国信息流空间的省级文化交互度影响因素 (2010—2016 年)

	模型 1 混合 OLS	模型 2 固定效应	模型 3 差分 GMM	模型 4 系统 GMM
省份上年文化交互度	.212 *** (.028)	.121 *** (.023)	.084 *** (.005)	.129 *** (.009)
省份人均 GDP	.301 (.195)	1.554 * (.598)	2.386 *** (.288)	.713 *** (.214)
省份人口教育水平	.272 (.182)	-1.663 ** (.595)	-1.798 *** (.347)	.990 ** (.336)
省份居民收入水平	-1.367 *** (.348)	-1.340 ** (.513)	-2.011 *** (.610)	-1.322 *** (.246)
省份基础设施水平	.306 *** (.067)	-1.953 * (.779)	-1.998 + (1.182)	-.281 (.219)
省份社会发展水平	.162 (.142)	-.584 * (.251)	-.925 *** (.180)	-1.565 *** (.337)
省份旅游业收入	-.025 (.042)	.437 *** (.108)	.575 *** (.049)	.268 *** (.037)
省份人口规模	.246 ** (.082)	-.865 (1.108)	.436 (.868)	.429 + (.244)
省份国际知名度	.219 *** (.061)	-.122 (.164)	-.187 ** (.057)	-.153 ** (.050)
截距	15.075 *** (3.491)	31.488 ** (9.742)	19.837 ** (6.391)	3.189 (4.640)
AR (2) 检验 P 值			.164	.179
Hansen Sargon 检验 P 值			.125	.599
N	186	186	155	186

注: (1) 双括号内为经过省份稳健调整的标准误; (2) +  $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$  (双尾检验); (3) AR (2) 为 Arellano-Bond 二阶误差自相关检验 ( $H_0$ : 无自相关), Hansen 检验为工具变量过度识别检验 ( $H_0$ : 过度识别合法)。

采用动态面板估算含因变量滞后项的模型，对模型设置是否合理有一个重要的经验性指标：滞后项的系数应该低于混合 OLS 回归模型，高于静态固定效应模型（Bond, 2002）。因此，我们分别对“交互度”和“内卷度”进行了混合 OLS 模型、静态面板模型、差分 GMM 模型和系统 GMM 模型的回归。经过比选，我们发现系统 GMM 模型的估计值符合这一要求，相对于混合 OLS 和静态面板模型，其既考虑了文化力的路径依赖效应，矫正了省份内随时间变化的不可观测异质性，相比差分 GMM，其改善了可能存在的弱工具变量问题，因此，系统 GMM 估计量更为可靠。我们在表 3、表 4 展示了全部 8 个模型的回归结果。

表 4 中国信息流空间的省级文化内卷度影响因素（2010—2016 年）

	模型 1 混合 OLS	模型 2 固定效应	模型 3 差分 GMM	模型 4 系统 GMM
省份上年文化内卷度	.920 *** (.033)	.104 (.071)	-.034 * (.015)	.510 *** (.026)
省份人均 GDP	-.123 (.094)	.099 (.296)	.251 + (.139)	-.268 + (.137)
省份人口教育水平	.034 (.078)	.524 (.343)	.289 (.186)	-.075 (.177)
省份居民收入水平	.095 (.125)	-.373 (.299)	-.500 *** (.146)	.563 *** (.128)
省份基础设施水平	-.025 (.029)	.827 + (.461)	.158 (.308)	.359 ** (.123)
省份社会发展水平	.005 (.060)	-.148 (.144)	.215 (.162)	-.303 *** (.078)
省份旅游业收入	.008 (.018)	-.124 + (.064)	-.098 ** (.037)	-.129 *** (.022)
省份人口规模	-.032 (.036)	-.677 (.645)	.159 (.261)	-.478 *** (.118)
省份国际知名度	-.035 (.024)	-.063 (.093)	.064 * (.031)	-.150 *** (.031)
截距	-.091 (1.225)	.769 (5.552)	-.845 (1.969)	-.517 (1.331)
AR (2) 检验 P 值			.101	.241
Hansen Sargon 检验 P 值			.267	.636
N	186	186	155	186

注：(1) 双括弧内为经过省份稳健调整的标准误；(2) + p < 0.1；\* p < 0.05；\*\* p < 0.01；\*\*\* p < 0.001（双尾检验）；(3) AR (2) 为 Arellano-Bond 二阶误差自相关检验（H0：无自相关），Hansen 检验为工具变量过度识别检验（H0：过度识别合法）。

## 四、分析结果

表5为文化交互度和文化内卷度的系统GMM估计结果。首先对模型残差采用AR(2)和Sargan检验,以考察系统GMM估计的适用性,结果显示,不论是文化交互度还是文化内卷度均通过了AR(2)和Sargan检验,这表明文化交互度和文化内卷度均不存在二阶序列相关,且两模型所用的内部工具变量均是有效的。进一步观察模型发现,文化交互度和文化内卷度均具有动态持续性,即,当年的文化交互度和文化内卷度会受上年相应测量的影响。

尽管如此,影响文化交互度和文化内卷度的因素存在很大差异。这很大程度上与省份交互度、内卷度的分布有关。由图2、图3所示,交互度较高的省份(如江苏、山东、广东、浙江、北京、上海)主要集中在东部地区,属于我国经济政治文化中心;而内卷度较高的省份(如西藏、云南)身处中西部内陆地区,多为少数民族聚居区域,经济较为落后、基础设施薄弱、互联网普及程度相对低,信息闭塞,但属于著名的旅游胜地,也是重要的政治要塞。表5模型结果显示,在控制了其他变量的情况下,省份人均GDP每增加1%,省份文化交互度便会提高0.713%;类似地,省份人口大学占比每增加1%,该省文化交互度就会增加0.99%,而这两个变量对文化内卷度却不存在显著影响。可见,省份文化互动总体依托于经济发展和高质量人口,而对于文化内卷度则不然。相比之下,省份基础设施的水平对文化内卷度的影响更加重要。模型结果显示,在其他因素一致的情况下,人均高速公路里程越长,该省文化内卷度越高。原因可能在于交通网络可达性使得人流物流的直接联系可以作为互联网信息需求的替代。

进一步观察省城镇居民家庭平均收入,发现居民家庭平均收入对文化交互度有显著负影响,而对文化内卷度有显著正影响。可能的解释是:省份家庭平均收入代表着省内人民生活水平。而较高生活水平带来的对外界信息获取渠道相对更为多元和便利:既形成了主观阶层定位的优势,又形成了通过互联网检索对外界信息进行主动吸取的积极性,从而降低了文化交互度,增加了独立性的倾向,提高了内卷度。与此相映成趣地,我们发现省份的旅游业发达程度会增加文化交互度而减少文化内卷度。旅游业是外界了解旅游目的地省份的重要途径之一,旅游收入越高意味着前来该省旅游的人次越多,文化交互度也越高。而旅游收入对文化内卷度的负向关联

则可能是因为，旅游业发达意味着与外省人流直接进行接触的机会大大增加，从而带动了对外省信息的主动性关注。

最后我们还发现，省份社会组织发展水平和国际知名度对文化交互度和内卷度具有显著的负影响。社会组织的增多，意味着信息流动的渠道进一步多元化，会使得主动通过互联网来获取外界省份信息的重要性下降，因此会降低文化交互度和内卷度。而国际知名度往往来源于媒体报道以及和国际人流的直接交流。知名度越高，意味着借助其他非互联网渠道的信息流互通性更高，这也会降低对省份居民通过互联网进行检索外省信息的需求。

表 5 系统 GMM 模型：流空间省级文化力的影响因素（2010—2016 年）

	Y = 文化交互度 系统 GMM	Y = 文化内卷度 系统 GMM
省份上年 Y 值	.129 *** (.009)	.510 *** (.026)
省份人均 GDP	.713 *** (.214)	-.268 (.137)
省份人口教育水平	.990 ** (.336)	-.075 (.177)
省份居民收入水平	-1.322 *** (.246)	.563 *** (.128)
省份基础设施水平	-.281 (.219)	.359 ** (.123)
省份社会发展水平	-1.565 *** (.337)	-.303 *** (.078)
省份旅游业收入	.268 *** (.037)	-.129 *** (.022)
省份人口规模	.429 (.244)	-.478 *** (.118)
省份国际知名度	-.153 ** (.050)	-.150 *** (.031)
截距	3.189 (4.640)	-.517 (1.331)
AR (2) 检验 P 值	.179	.241
Hansen Sargon 检验 P 值	.599	.636
N	186	186

注：(1) 双括弧内为经过省份簇稳健调整的标准误；(2) \*  $p < 0.05$ ；\*\*  $p < 0.01$ ；\*\*\*  $p < 0.001$ （双尾检验）；(3) AR (2) 为 Arellano-Bond 二阶误差自相关检验（ $H_0$ ：无自相关），Hansen 检验为工具变量过度识别检验（ $H_0$ ：过度识别合法）。

## 五、讨论和结论

广义文化力，本质上可以看作一种综合性的软实力。观察和解读我国区域文化互动的空间结构、等级和特征机制，既具有学术领域的开拓意义，也具有相当的政策意涵。本文认为，在省域层面，互联网信息的“流数据”反映了省域主体在互联网平台的互动行为，尤其是搜索引擎门户的信息流辐射与穿透，可以被用来作为区域文化力指标的基础测量工具。聚焦互联网中的中国省际信息“流空间”体系，我们借助百度搜索的十年数据，构建起省域广义“文化力”的两大指标——文化“交互度”（互动力）和“内卷度”（穿透力），并对其宏观影响因素进行了分析。

研究发现，在空间结构上，无论是文化“交互度”和“内卷度”，聚类分析的结果都表明，两大指标存在清晰的等级体系。从省际互引和穿透的两两关系角度，前者主要呈现多中心结构，而后的不对称特征则十分明显。在空间分野上，著名的胡焕庸线和两大文化力指标均具有重要的关联，但地理分野的指向却截然相反。在形成机制上，通过可视化和描述，我们可以判断导致文化互动和内卷差异化格局的因素具有多元混合的特征。我们利用前沿的动态面板广义矩分析发现，两大指标分别受到各省人口、经济、社会、文化等一系列发展水平因素的影响，而影响的大小和方向大相径庭。

本文的研究贡献主要有几个方面：在理论上，我们首次提出“流空间”的文化交互、文化内卷等地域文化指标概念，并发现一系列流空间的结构特征。在方法上，我们利用在互联网中的海量信息，对文化力指标进行了新颖的概念操作化，构建了通过传统方法难以测量的“文化力”指标。在学科上，通过文化力概念的介入和其影响因素的严格计量分析，为传统地理学科“流空间”分析注入了社会学视野和定量分析的科学检验，为大数据助力社会科学和空间分析提供了一定的经验。从政策意涵上讲，本文有助于理解、提升区域“文化软实力”作为增强其核心竞争力、提升发展质量、形成特色发展模式以及扩大影响力的重要路径。

本文仍旧存在一些局限。首先，网民是否上网，或者说成为网民仍然有选择性。但是由于缺乏相网民个体层面的信息，我们无法直接对该选择性进行校正。然而近年来我国互联网发展迅速，普及率已近60%，网民规模越发庞大，也因此，用户选择性的问题不会对我们的结果构成严重威胁。其次，有关“被关注度”指标构建的

前提在于人口迁移是影响省份关注度的主要驱动力。然而该假设在信息社会的大背景下可能存在“替代”的可能，即当人们试图了解某个省份时，会根据自己的检索习惯进行检索，从而使地理空间距离的重要性较之从前有所下降。也因此，我们的指标仍旧是一个保守估计。最后，“文化力”指标的构建是基于2010—2016年中国各省公众在“百度”中省份名称的搜索。中国地理空间广袤而复杂，不论是对于交互度还是内卷度在多大程度上是由距离相近带来的仍有待检验。比如西部地区，即便是省份相邻，省会之间的距离仍旧十分遥远，基于当前的指标，无形中降低了西部省份的相关指标。此外，省域内也存在巨大的差异，且通常除直辖市外，大众对省域的文化关注往往体现在更低的地理层级，比如市、县，甚至与城市相关的风景名胜等，我们的做法在一定程度上忽略了对省内低层级地理区域的关注。尽管如此，我们的研究仍具有高度启发性。比如，从省域内部经济文化发展的差异来考虑，江苏省内苏州的经济社会文化最为发达，泰州较之苏州，被外省搜索关注的程度可能略低。可以说，在对江苏的搜索中，苏州占据的权重较大。但这恰恰体现了江苏省的文化力也即被外界关注的主观特征：受到首位度或者权重较大的城市影响。为进一步推进文化力的研究，我们将进一步补充市级层面的分析。再如，省级面板回归中，模型解释变量难免遗漏，例如少数民族人口比例、宗教人口比例、省级文化设施等。尽管这些测量随年度变化不大，理论上通过固定效应模型可以排除遗漏变量的干扰，但仍然难免挂一漏万。

此外，还有一点值得注意的是，有关省份的搜索可能是基于搜索时点的事件，其可能是政治经济文化方面的，也可能是自然环境方面的等。常常正是这些事件触发了人们对某些区域的搜索。那么，基于事件的区域搜索是否能够反映“文化力”这一概念，并作为合适的文化力测量呢？有关这点，我们仍需谨慎。但可以肯定的是，事件的发生往往具有一定的文化基础，这里我们提及的文化基础，是一种广义上的文化基础，其囊括了人类历史实践过程中产生和创造出来的各种物质和精神财富的总和。但是这些搜索最终可以转化成多少的“有益”文化力尚待检验。不论如何，总体上我们的分析不仅提供了基于信息“流空间”的文化力省域总览，展现了中国省域层次的社会文化多元互动和联系，更从经济社会等多元视角考察了省域间信息流互动的长期动态变化趋势和影响因素。这为后续的空间结构研究和文化研究提供了总览信息和重要基础。

## 参考文献:

- 白欲晓, 2011, 《“地域文化”内涵及划分标准探析》, 《江苏社会科学》第1期。
- 曹有挥、陈国伟, 2014, 《基于生产性服务业的长江三角洲城市网络》, 《地理研究》第2期。
- 陈涛、林杰, 2013, 《基于搜索引擎关注度的网络舆情时空演化比较分析——以谷歌趋势和百度指数比较为例》, 《情报杂志》第3期。
- 陈云松、吴青熹、黄超, 2015, 《大数据何以重构社会科学》, 《新疆师范大学学报》(哲学社会科学版)第3期。
- 陈云松、朱灿然、张亮亮, 2017, 《代内“文化反授”: 概念、理论和大数据实证》, 《社会学研究》第1期。
- 陈映雪、甄峰、王波、邹伟, 2013, 《基于社会网络分析的中国城市网络信息空间结构》, 《经济地理》第4期。
- 顾朝林, 2011, 《城市群研究进展与展望》, 《地理研究》第5期。
- 蒋大亮、孙焯、任航、陈影影、张振克, 2015, 《基于百度指数的长江中游城市群城市网络特征研究》, 《长江流域资源与环境》第10期。
- 金钟范, 2010, 《基于企业母子联系的中国跨国城市网络结构——以中韩城市之间联系为例》, 《地理研究》第9期。
- 梁辉, 2008, 《我国省际信息流动规模测算与空间结构分析》, 《经济问题》第12期。
- 李敬、陈澍、万广华、付陈梅, 2014, 《中国区域经济增长的空间关联及其解释——基于网络分析方法》, 《经济研究》第11期。
- 路旭、马学广、李贵才, 2012, 《基于国际高级生产者服务业布局的珠三角城市网络空间格局研究》, 《经济地理》第4期。
- 孙中伟、路紫, 2005, 《流空间基本性质的地理学透视》, 《地理与地理信息科学》第1期。
- 谭一滔、杨永春、冷炳荣等, 2011, 《基于高级生产者服务业视角的成渝地区城市网络体》, 《地理科学进展》第6期。
- 王桂新, 1993, 《我国省际人口迁移与距离关系之探讨》, 《人口与经济》第2期。
- 汪明峰、宁越敏, 2004, 《互联网与中国信息网络城市的崛起》, 《地理学报》第3期。
- 汪明峰, 2006, 《城市的网络优势——中国互联网骨干网络结构与节点可达性分析》, 《地理研究》第2期。
- 武前波、宁越敏, 2012, 《中国城市空间网络分析——基于电子信息企业生产网络视角》, 《地理研究》第2期。
- 吴威、曹有挥、梁双波等, 2009, 《中国铁路客运网络可达性空间格局》, 《地理研究》第5期。
- 熊丽芳、甄峰、王波、席广亮, 2013, 《基于百度指数的长三角核心区城市网络特征研究》, 《经济地理》第7期。
- 熊丽芳、甄峰、席广亮、朱晓清、王波等, 2014, 《我国三大经济区城市网络变化特征——基于百度信息流的实证研究》, 《热带地理》第1期。
- 闫庆武、卞正富, 2015, 《基于GIS的中国省际人口迁移流的空间特征分析》, 《人文地理》第3期。

- 尹俊、甄峰、王春慧, 2011, 《基于金融企业布局的中国城市网络格局研究》, 《经济地理》第5期。
- 于文丽, 蒲英霞, 陈刚、王洁臣民, 2012, 《基于空间自相关的中国省际人口迁移模式与机制分析》, 《地理与地理信息科学》第2期。
- 于欣洺、吴相利, 2015, 《基于百度指数的东北地区城市网络特征研究》, 《哈尔滨师范大学自然科学学报》第6期。
- 甄峰、王波、陈映雪, 2012, 《基于网络社会空间的中国城市网络特征——以新浪微博为例》, 《地理学报》第8期。
- 张闯、孟韬, 2007, 《中国城市间流通网络及其层级结构——基于中国连锁企业百强店铺分布的网络分析》, 《财经问题研究》第5期。
- Alderson, A. S. & J. Beckfield 2004, “Power and Position in the World City System.” *American Journal of Sociology* 109 (4).
- Arellano, M. & S. Bond 1991, “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations.” *Review of Economic Studies* 58 (2).
- Arellano, M. & O. Bover 1995, “Another Look at the Instrumental Variable Estimation of Error-components Models.” *Journal of Econometrics* 68 (1).
- Beaverstock, J. V. 2007, “World City Networks from Below: International Mobility and Inter-city Relations in the Global Investment Banking Industry.” In *Cities in Globalization: Practices, Policies, Theories*, edited by Peter Taylor, Ben Derudder, Peter Saey, Frank Witlox. London and New York: Routledge.
- Blundell, R. & S. Bond 1998, “GMM Estimation with Persistent Panel Data: An Application to Production Functions.” *Journal of Econometrics* 19 (3).
- Bond, S. 2002, “Dynamic Panel Models: A Guide to Micro Data Methods and Practice.” *Portuguese Economic Journal* 1 (2).
- Bryden, J., S. Funk & V. A. Jansen 2013, “Word Usage Mirrors Community Structure in the Online Social Network Twitter.” *Epj Data Science* 2 (1).
- Castells, M. 1996, *The Rise of the Network Society* (Volum 1). Malden, MA: Blackwell.
- Castells, M. 2010, “Globalisation, Networking, Urbanisation: Reflections on the Spatial Dynamics of the Information Age.” *Urban Studies* 47 (13).
- Chen, Y., F. Yan & Y. Zhang 2016, “Local Name, Global Fame: The International Visibility of Chinese Cities in Modern Time.” *Urban Studies* 54 (11).
- Choi, J. H., G. A. Barnett & B. S. Chon 2006, “Comparing World City Networks: A Network Analysis of Internet Backbone and Air Transport Intercity Linkages.” *Global Networks* 6 (1).
- Curme, C., T. Preis, H. E. Stanley & H. S. Moat 2014, “Quantifying the Semantics of Search Behavior before Stock Market Moves.” *Proceedings of National Academy of Science* 111 (32).
- Derudder, B. 2006, “On Conceptual Confusion in Empirical Analyses of a Transnational Urban Network.” *Urban Studies*



43 (11).

- Griliches, Z. & J. A. Hausman 1984, "Errors in Variables in Panel Data." *Journal of Econometrics* 31 (1).
- Geoffrey, H. P. & P. Kathy 2006, *The Polycentric Metropolis: Learning from Mega-city Regions in Europe*. London and New York; Routledge.
- Hall, Peter & K. Pain 2006, *The Polycentric Metropolis: Learning from Mega-City Regions in Europe*. London, Earthscan Publications Ltd.
- Holtz-Eakin, D., W. Newey & H. S. Rosen 1988, "Estimating Vector Autoregressions with Panel Data." *Econometrica* 56 (6).
- Mahutga, M. C., X. Ma, D. A. Smith & M. Timberlak 2010, "Economic Globalisation and the Structure of the World City System: the Case of Airline Passenger Data." *Urban Studies* 47 (9).
- Malecki, E. J. 2002, "The Economic Geography of the Internet's Infrastructure." *Economic Geography* 78 (4).
- Matsumoto, H. 2004, "International Urban Systems and Air Passenger and Cargo Flows: Some Calculations." *Journal of Air Transport Management* 10 (4).
- Mellon, J. 2014, "Internet Search Data and Issue Salience: The Properties of Google Trends as a Measure of Issue Salience." *Journal of Elections Public Opinion & Parties* 24 (1).
- Nuti, S. V., B. Wayda, I. Ranasinghe et al. 2014, "The Use of Google Trends in Health Care Research: A Systematic Review." *Plos One* 9 (10).
- Robinson, Jennifer 2005, "Urban Geography: World Cities, or A World of Cities." *Progress in Human Geography* 29 (6).
- Scheitle, C. P. 2011, "Google's Insights for Search: A Note Evaluating the Use of Search Engine Data in Social Research." *Social Science Quarterly* 92 (1).
- Smith, D. A., Timberlake Michael 1995, "Cities in Global Matrices: Toward Mapping the World-system's City System." *World Cities in a World-system*.
- Taylor, P. J. 2001, "Specification of the World City Network." *Geographical Analysis* 33 (2).
- Taylor, P. J. & B. Derudder, 2004, *World City Network: A Global Urban Analysis*. London and New York; Routledge.
- Townsend, A. M. 2001, "Network Cities and the Global Structure of the Internet." *American Behavioral Scientist* 44 (10).
- Womack, B. 2001, "China's Border Trade and its Relationship to the National Political Economy." *American Asian Review* 19 (2).

作者单位：南京大学中美文化交流中心、南京大学社会学院  
责任编辑：郭 冉