

# 中国地级以上城市人口流动网络研究<sup>\*</sup>

## ——基于百度迁徙大数据的分析

蒋小荣 汪胜兰

**【摘 要】**文章基于百度迁徙大数据,运用复杂网络分析方法,构建中国334个城市间人口日常流动网络模型,测度并分析其复杂性网络特征。结果显示,城市行政级别的高低与人口流动影响力存在一定的正相关关系,大部分城市在人口流动上处于“相对平衡”状态;城市间人口流动网络显现出明显的等级层次性,东部沿海三大城市群依旧是人口流动的热点区域,但西部地区中心城市网络影响力大幅提升。整个网络在空间上呈现出“两大四小”多中心的格局,其中北京与上海为“两大”中心,重庆、西安、广州和武汉为“四小”中心,6个中心节点在整个网络中处于绝对主导地位,但各自的流场分布又有所差异。网络中城市等级与流动距离存在一定的正相关关系,空间距离依然是人口流动的重要约束力。整个人口流动网络符合无标度网络特性,并呈现出典型的“小世界”网络特征,说明中国城市人口流动网络具有较高的连通性和可达性。

**【关键词】**百度迁徙 大数据 人口流动网络 出度 入度 城市

**【作 者】**蒋小荣 兰州大学资源环境学院,工程师;汪胜兰 兰州财经大学陇桥学院,讲师。

人口流动与城市发展是一个互动过程,人口流动既是城市发展的必然产物,又反过来影响城市经济社会发展进程。人口与城市的关系,如同网络中流要素与节点的关系(张永庆,2006)。人口流动是城市间物质流、信息流、资本流、技术流的重要载体,人口在城市之间高速自由地流动迁移,尤其是在创新经济背景下,各种形式的面对面交流和知识的传播学习已经成为城市创新联系的关键角色,推动区域内物质、资本、信息等生产要素的加速流动、优化,人口流动网络化正逐步形成。而基于城市视角的人口流动网络研究是人口地理学与城市化研究的重要补充。因此,加强对中国城市间人口流动的分析

<sup>\*</sup> 本文为国家自然科学基金项目“中国西部的企业迁移、园区发展与城市重构的互动模式研究”(编号:41571155)的阶段性成果。

与研究,具有重要的意义。

近年来,国内学者的研究视角从人口流动的空间分布开始转向人口迁移流模式和特征研究。如丁金宏等(2005)和孙峰华等(2006)分别采用2000年人口普查数据和2005年1%人口抽样调查数据,阐述了中国省级人口迁移的区域分异性。严善平(2007)、刘望保等(2012)和王桂新等(2012)采用两个时期的人口普查数据,通过对比分析,揭示了人口迁移的区域模式及其变化,并对其影响因素进行了计量分析。这些研究大多在省级空间单元下进行。城市层面近几年才有所涉及,如鲍常勇(2007)通过城市统计年鉴数据,对中国主要城市的人口流动规模进行了简要的比较分析。段成荣、杨舸(2009)则从区域、省级和城市3个层面分析了流动人口流入地分布的变动趋势。王国霞等(2012)和于涛方(2012)主要侧重于地级市层面。劳昕、沈体雁(2015)加入了对人口流动的方向性探讨,不过仅分析了主要迁入城市对迁出省份的流场模式。上述研究多将流入地和流出地孤立分析,缺乏对不同地区(城市)人口流动的流向、流量及相互联系等空间分析方面的研究。复杂网络与统计物理方法等跨学科研究方法的兴起为研究城市人口流动提供了新的分析工具与视角。复杂网络是指复杂系统中通过节点之间的相互作用或关联来构建网络,再利用统计物理学分析网络结构及其拓扑关系特征。对于中国跨区域的人口流动网络研究,直到近两年来才有部分学者尝试涉及。侯贺平等(2013)、王珏等(2014)和陈锐等(2014)分别基于乡镇尺度、长三角城市群和省级尺度,运用社会网络工具,研究了人口流动的网络特征,采用的依然是人口普查数据,主要运用重力模型来构建地区间人口迁移网络。此类方法其实是基于空间相互作用原理对地区间人口迁移的流量和流向进行估计与预测,得出的只是可能发生的“人口流动”数据,难以反映真实情况。为了弥补以往研究只对人口迁移研究的缺陷,本文采用百度LBS(基于地理位置的服务)开放平台提供的“百度迁徙”大数据进行研究。与传统的普查数据相比,该数据可以更为真实地反映个体出行移动轨迹,更偏向于分析人口的日常流动,具有典型的即时性特征,是对以往数据的一个重要补充。

如今,“大数据”已经渗透到全球范围内社会经济的各个领域,带有地理位置信息的大数据,为人口地理学研究的发展提供了新的方法与手段。百度迁徙数据是通过对比用户位置的变化,将8小时内所有位置发生变化的智能终端用户数量进行计算分析,可以全覆盖铁路、公路、航空等各类交通工具。本文数据采集于2015年9月14~21日,通过“百度迁徙”网页,依次统计中国334个城市的前10位流入和流出城市的排序。由于“百度迁徙”中列出的是相对权重比例数,如由A城市流入B城市的人数占流入B城市总人数的百分比,是个相对重要性系数,因此需要采用权重赋值法对数据进行标准化处理,将前10位流入或流出城市按照排序依次赋上数值10到1,用以表示人口流动加权网络中某城市的前10位流入或流出城市的得分系数。以纵坐标的城市作为流出城市,

横坐标为流入城市,输入前 10 位的关系数据,构建 334 个城市间人口流动网络的邻接关系矩阵表,最终得到一个  $334 \times 334$  的有向多值网络矩阵。本研究采用复杂网络分析方法,构建中国 334 个城市和地区(含 331 个地级及以上城市和香港、澳门、台湾地区)间人口流动网络,测度并分析该网络的复杂性结构特征,为今后的人口流动与城市网络研究提供了一种新思路。

### 一、地级市人口流动网络的空间格局与层级划分

在以往的人口流动研究中,方向性一直是一个很容易被忽视的特征,但也是一个重要的特征。复杂网络侧重研究节点及节点间的拓扑关系。人口流向是沿着城市节点间的连线的方向进行,存在入度与出度之分,分别表示节点在网络中的吸引力和辐射力。节点的入度值和出度值则体现出节点在网络中的地位和角色(刘军,2009)。

区域间人口的自由流动需要依托一定的外部条件,如交通工具、制度环境和社会关系等。因此,城市间人口流动与城市的经济、人口规模、地理位置等传统引力因素存在着密切关系。综合考虑人口流动网络中城市节点的度值,可以发现以下几个特征。

**(一) 行政等级越高的城市在人口流动网络中地位越高,影响范围越广;入度与出度分布呈现出相似的集聚性,整体上大部分城市的人口流入和流出保持着相对平衡**

本文主要通过复杂网络分析工具 Gephi,计算中国 334 个地级及以上城市的人口流动网络度值,发现各地级市之间的网络地位差异较大。另外,城市行政级别的高低与网络度值存在一定的相关性。从表 1 可见,4 个直辖市(北京、上海、重庆和天津)的网络总度值及排序高于副省级城市与普通省会城市,大部分副省级城市高于普通省会城市,大部分省会城市高于其他地级市。从总度值的平均值上来看,直辖市为 2 476,约为副省级城市(620)的 4 倍,是普通省会城市(231)的 11 倍,网络地位优势比较明显。

整个人口流动网络中,网络地位最高的城市是北京,入度值为 1 967,出度值为 2 034、总度值超过 4 000。网络地位最低的为临沧、大兴安岭、海北州、克拉玛依等 32 个总度值为 0 的城市,可以看出这些城市有着共同的特征:城市规模小且均位于边疆地区,人口的流入流出相对较少。在对人口流入的吸引力(入度)上看,排名前 30 位的城市包括 4 个直辖市、15 个副省级城市、7 个省会城市和 4 个其他地级市(苏州、桂林、三亚、香港),入度值均在 90 以上。而人口流出前 30 位的城市,依然包括 4 个直辖市、15 个副省级城市、8 个省会城市和 3 个其他地级市(苏州、邯郸和徐州),出度值均在 100 以上,说明在对人口流入的吸引力与人口流出的辐射能力来看,行政等级高的城市都处于明显的优势地位(见表 1)。

从城市的入度与出度空间分布看,人口流动呈现出相似的集聚性:网络地位较高的城市均主要集中分布在京津冀、长三角、珠三角、成渝与长江中游等国家级城市群内,另

表 1 不同级别地级市的人口流动网络指标统计

城市类别	入度	出度	入度 - 出度	总度值	排序
直辖市					
北京	1967	2034	-67	4001	1
上海	1344	1434	-90	2778	2
重庆	879	896	-17	1775	3
天津	611	739	-128	1350	4
副省级城市					
西安	663	686	-23	1349	5
成都	632	484	148	1116	6
广州	459	560	-101	1019	7
武汉	439	555	-116	994	8
杭州	525	392	133	917	9
深圳	396	520	-124	916	10
南京	323	223	100	546	14
济南	192	185	7	377	16
长春	136	215	-79	351	17
厦门	233	105	128	338	18
沈阳	144	180	-36	324	19
宁波	120	202	-82	322	20
哈尔滨	134	182	-48	316	21
青岛	159	68	91	227	27
大连	97	88	9	185	32
普通省会城市					
郑州	291	361	-70	652	12
长沙	315	269	46	584	13
昆明	230	149	81	379	15
兰州	105	171	-66	276	22
福州	132	143	-11	275	23
合肥	156	118	38	274	25
石家庄	98	117	-19	215	28
太原	67	134	-67	201	29
贵阳	67	113	-46	180	34
西宁	45	109	-64	154	41
银川	62	90	-28	152	42
南宁	53	79	-26	132	54
海口	58	74	-16	132	55
南昌	30	92	-62	122	57
乌鲁木齐	61	45	16	106	65
呼和浩特	29	40	-11	69	95
拉萨	27	10	17	37	159
平均度值					
直辖市	1200	1276		2476	25
副省级城市	310	310		620	15
普通省会城市	107	124		231	45

外还零散地分布在中西部地区的省会城市,如西安、郑州、兰州等。虽然城市节点在入度与出度的网络空间格局大致保持较高程度的相似性,但少数节点仍存在较大的差异,如乌鲁木齐、三亚、桂林、青岛等城市入度值明显大于出度值。从整个空间分布图来看,入度分布更加倾向于在长三角地区集聚,而出度分布倾向于在京津冀地区集聚。这表明长三角地区对于人口具有强大的吸引力,而京津冀城市群则表现出更强的人口辐射能力。

通过对照网络中 334 个节点的入度与出度,发现两组数据排序的相对一致性(见图 2)。经 SPSS 相关性检验, Pearson 相关系数高达 0.983, sig 值小于 0.05,可以看出节点的入度与出度呈现出显著的正相关关系。通过统计,有 279 个城市节点的入度与出度之差的绝对值小于等于 30,占总数的83.53%。这在一定程度上表明城市节点的入度与出度的相对平衡,也可以说人口在城市间可以自由流动,大多数城市不存在单向的阻碍流动的约束力。当然也存在部分节点入度与出度存在较大的差值。如图 2 所示,典型的入度盈余(入度-出度)较大



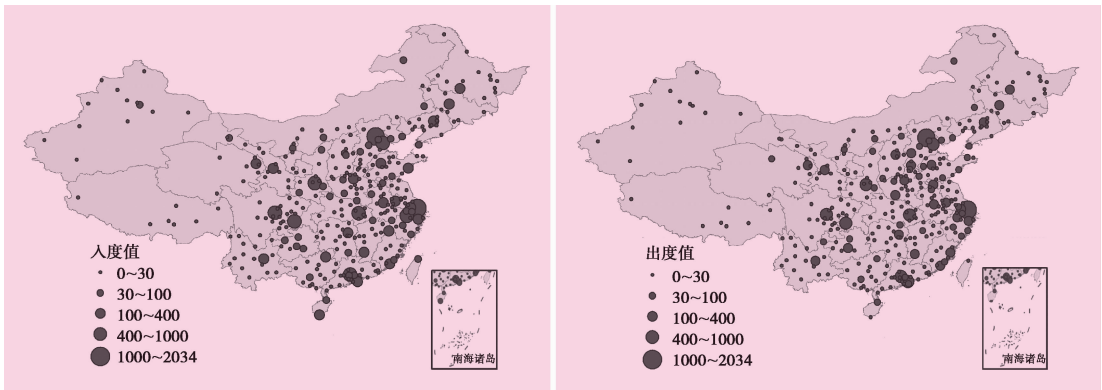


图 1 地级市人口流动网络节点度值空间分布图(入度、出度)

的节点有成都、杭州、昆明、厦门、三亚、青岛、桂林、香港、澳门等大型城市,在中小城市中也由黄山、酒泉、上饶、北海、张家界等。不难看出,这些均为典型的旅游城市,在人口流入的吸引力上有着较大的优势。出度盈余的节点有深圳、武汉、郑州、南昌和太原等,多为区域核心节点(省会城市),在人口流动网络中主要起辐射控制作用。还有部分节点处于一种入度与出度的“相对平衡”状态,如北京、上海、重庆、西安等,这些节点的总度值恰好处于整个网络中第一序列之中。这些城市人数众多,流动频繁,人口的流入和流出处于“高水平”的均衡状态。另外一些小城市(如宜昌、威海、常德、安顺、三明、衢州等)入度与出度值几乎相等且均小于 30,表明人口流动处于低水平的均衡状态。

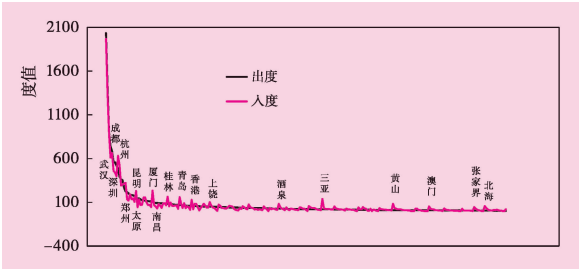


图 2 人口流动网络的入度与出度相关性分析

(二) 东部沿海三大城市群仍是人口流动的热点区域,但西部地区中心城市网络地位大幅提升,交通与地理区位也是人口流动网络的主要影响因素

根据所有城市节点的总度值进行排序,发现城市间人口流动网络显现出明显的等级层次性。为了增强同级间的同质性和各层级间的差异性,本文采用自然裂点分级法将 334 个节点分为 5 个层级(见表 2)。从表 2 的城市节点分级统计可以发现,第一序列中(全国性网络副中心以上)西部城市占有较大的比例,如重庆、西安、成都等;部分边缘城市在网络中处于较高的地位,如昆明、长春、哈尔滨、兰州等位列区域网络中心。这些都说明对西部及偏远地区而言,省会城市或区域核心城市是人口流动的第一目的地,其次才是北上广深等一线城市。这也印证了人口迁移的一般规律,尤其农村人口首先向区域内的中心城市(省会城市)集聚,然后再向更高级别的核心城市迁移,对于中西部及偏远

表 2 城市人口流动网络层级划分

层级(网络总度值)	城 市
全国性网络中心(>2000)	北京、上海
全国性网络副中心(900~2000)	重庆、天津、西安、成都、广州、武汉、杭州、深圳
区域性网络中心(250~900)	苏州、郑州、长沙、南京、昆明、宁波、长春、厦门、沈阳、济南、哈尔滨、兰州、 赣州、福州、合肥
地方性网络中心(90~250)	桂林、青岛、石家庄、太原、徐州、香港、邯郸、南阳、银川、东莞、大连、榆林、金华、 廊坊、三亚、西宁、上饶、鄂尔多斯、贵阳、临沂、泉州、嘉兴、通辽、保定、洛阳、 商丘、温州、秦皇岛、南宁、海口、运城、南昌、十堰、遵义、佛山、唐山、九江、 酒泉、呼伦贝尔、乌鲁木齐、昭通、无锡、菏泽、黔南、信阳、安阳、聊城、黄山、 阜阳、郴州、烟台、大同、张家口
地方性网络节点(<90)	海西州等 256 个城市

地区人口净流出省份的单核心城市而言,这种吸纳能力往往显得更强。另外,赣州、邯郸和南阳等经济条件并不突出的城市,凭借着庞大的人口基数和优越的交通区位条件而在网络中占有较高的地位。这些都与传统的人口迁移研究结论并不完全一致,其中的一个主要原因在于,传统意义上的人口迁移主要影响因素为迁出地与迁入地的经济推力影响,而即时性的人口日常流动由于受到空间距离的限制和出行方式的影响,人们更倾向于向邻近大城市和交通枢纽流动,这就造成拥有广阔腹地的西部地区核心城市重庆、成都和西安在网络中占有重要地位,而偏远地区的省会城市和一些中小城市凭借优越的交通区位条件成为区域内的人口集散中心。这在一定程度上也反映出基于人口流动的城市网络体系并不完全遵循已有的基于经济规模的城市等级体系,城市在人口流动网络的地位高低和作用大小不仅取决于经济与人口规模,也受地理区位和交通条件等因素的共同影响。

二、人口流动网络流向分布的空间格局与核心城市的流场分布

(一) 地级市人口流动网络流向分布上形成“东南密集,西北稀疏”的态势,并呈现“两大四小”多中心网络空间格局

借助 ArcGIS 中的网络分析工具,提取 334 个节点城市间边权值(两个城市之间人口流动的得分系数)大于 0 的联系流。依据前文的节点总联系分级,将边权值分为“高、中、低”3 个层级,以便进一步分析中国城市人口流动网络联系的空间格局(见图 3)。

从图 3 可以看出,中国城市人口流动向东部沿海城市群和少数大城市集中的态势依然维持,流向分布上呈现明显的“东南密集,西北稀疏”的态势。由此可见,“胡焕庸线”不仅反映了两端人口空间分布的相对稳定性,也反映了城市间的人口日常流动的空间格局。虽然西部地区的人口流动日趋活跃,但东部地区作为主要人口集散地的局面未发

生根本改变,大部分西部地区和偏远地区的城市的首位联结城市均位于东南沿海。这也揭示了经济吸引力在人口流动中的核心作用。此外,综合考虑城市节点层级分布与城市间网络联系强度(边权值),发现城市人口流动呈现“两大四小”多中心网络空间格局,尤其是在高值联系流上更为明显。具体来讲,“两大”指北京和上海,“四小”指重庆、西安、广州和武汉,分别为京津冀、长三角、成渝、关中、珠三角和长江中游城市群的龙头城市。这六大节点的网络联系覆盖了全部334个城市,可以说6个中心节点在整个网络中处于绝对主导地位,是全国范围内主要的人口流动集散中心。

为了进一步体现“两大四小”核心节点在人口流动网络中的重要地位,通过提取这6个核心城市的网络联系边,分析它们各自联系覆盖的区域范围与强度(见图4)。由于

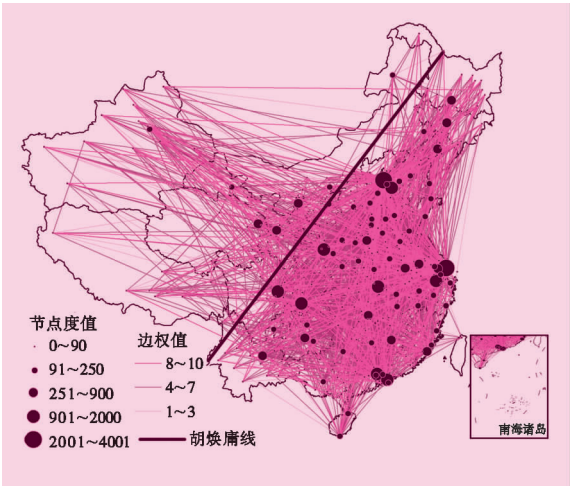


图3 中国城市人口流动网络空间格局

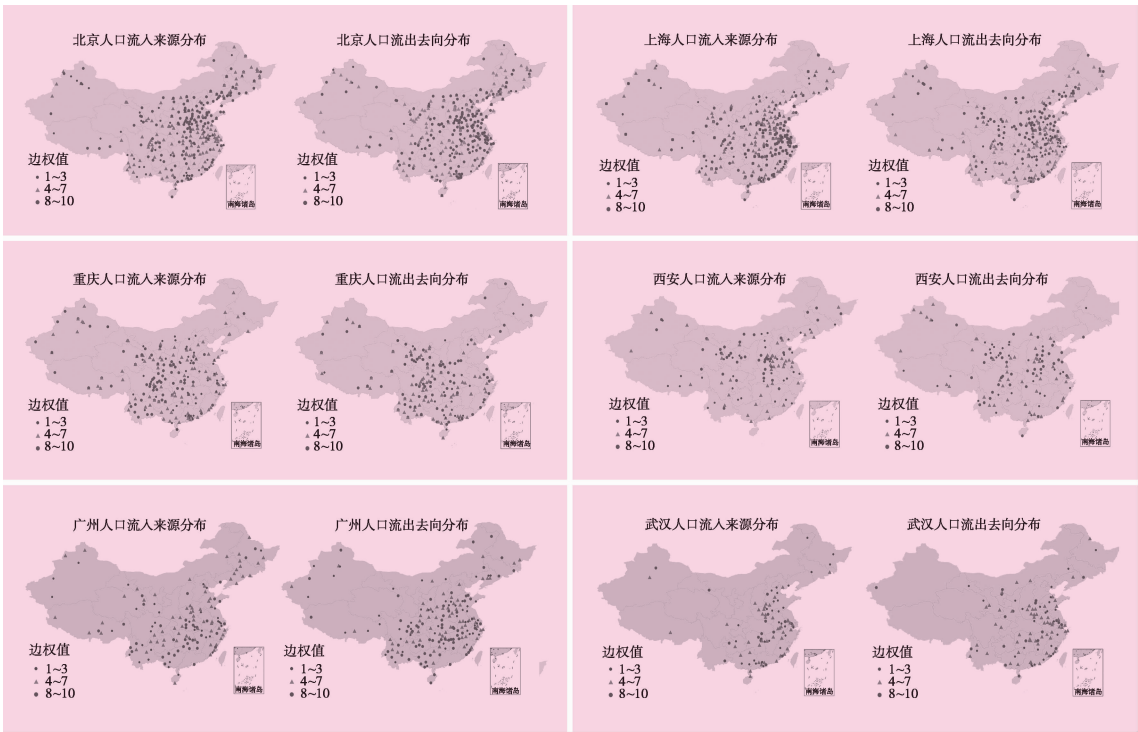


图4 “两大四小”中心节点人口流动分布

人口流动网络是个典型的有向网络,因此,本文将网络联系分为入度网络与出度网络两种,以便对比讨论分析。通过对比观察,发现六大核心城市在人口流动分布具有以下特征。

1. 各中心节点均有明确的辐射范围界限,等级越高的节点城市,辐射范围越大。通过对比观察,可以看出北京与上海辐射范围最大,基本上覆盖了全国范围,但两者的辐射重点又有所差异。如北京主要覆盖华北地区、东北与西北等北方地区和华东;上海主

表 3 六大中心节点人口流动网络联结城市数量

中心 节点城市	流动方向 (个数)	高值联系 (个数)	中值联系 (个数)	低值联系 (个数)
北京	流入(277)	山东(15)、河南(13)、山西(11)、河北(10)、吉林(7)、湖南(7)、内蒙古(6)	江苏(10)、四川(10)、黑龙江(8)、安徽(8)、云南(6)、江西(6)	广东(8)、新疆(4)、青海(4)、云南(3)
	流出(270)	山东(13)、河南(13)、安徽(13)、江苏(9)、浙江(8)、四川(8)、山西(7)、广东(7)	黑龙江(10)、辽宁(7)、新疆(7)、甘肃(5)、吉林(5)	共 16 个城市,处于零散分布
上海	流入(210)	安徽(16)、江苏(11)、浙江(8)、江西(7)、河南(7)、广东(5)、湖南(5)	广东(10)、山东(9)、河南(8)、河北(8)、云南(8)、四川(8)、湖北(6)	广东(5)、四川(4)、福建(3)、河南(3)、山东(3)
	流出(240)	河南(11)、山东(8)、辽宁(6)、河北(5)、山西(5)	安徽(8)、江苏(8)、浙江(6)、湖北(6)、广东(5)、云南(5)、四川(5)	安徽(7)、浙江(6)、江西(5)、四川(5)、甘肃(4)
重庆	流入(153)	四川(17)、陕西(6)、贵州(5)、云南(4)	云南(5)、湖南(5)、广东(5)、广西(4)	甘肃(6)、广东(5)、云南(3)、广西(3)
	流出(133)	四川(7)、云南(7)、甘肃(6)、陕西(4)、广东(4)	四川(13)、云南(7)、贵州(5)、湖南(4)	广西(4)、湖南(4)、广东(3)
西安	流入(120)	甘肃(8)、河南(5)、青海(3)	河南(7)、山东(3)、新疆(3)	西藏(4)、山东(3)
	流出(119)	四川(6)、河南(5)、甘肃(3)	甘肃(6)、新疆(6)、四川(4)	甘肃(5)、西藏(3)
广州	流入(80)	湖南(4)、广西(3)、江西(3)	河南(6)、吉林(6)、安徽(4)	江苏(4)、云南(3)
	流出(87)	四川(5)、云南(4)、山东(4)	湖南(5)、浙江(3)、湖北(3)	湖南(5)、江苏(3)
武汉	流入(90)	湖南(4)、河南(2)、江苏(2)	河南(8)、安徽(5)、广东(4)	安徽(4)、河北(4)、广东(3)
	流出(106)	江苏(3)、江西(2)、陕西(2)	安徽(8)、河南(7)、湖南(5)	广东(4)、江苏(3)、山东(3)

注:括号内数字表示该省中与中心节点联结的城市数量,这里仅列出居前几位的省份。



要覆盖华东地区、中原地区和华南;而“四小”中心节点的辐射范围相对要小很多,主要覆盖各自区域内部城市和邻近省份城市。

2. 对照流入来源分布和流出去向分布(高值联系 8~10),发现各中心节点在人口流向分布上并非是均衡的,而是呈现明显的空间分异性。例如,北京流动人口的来源主要为华北东北等北方城市,而流出地分布出现了明显的南移;上海则相反,流出地向北移动;重庆和西安节点流出地相对于流入地出现了小幅向东南偏移现象;而广州节点则是向西南方向偏移;武汉由于主要辐射周边省份城市,人口流向分布相对一致,未出现明显的偏移现象。这些重心节点的人口流向分布基本上互为补充契合,未出现明显的重叠现象,佐证了本文关于六大节点的网络联系覆盖了全部 334 个城市的观点。

3. 通过统计六大中心节点人口流动联结城市数目来看,如河南、山东、四川、广东、江苏、河北、湖南、安徽和湖北前 10 位人口大省是中国人口流动的主要流出地。其中河南和四川与 6 个中心节点均保持人口流动联系,这也符合现实中“哪都有河南人和四川人”的实际情形。作为传统的劳务输出大省,河南与四川人流动性强,分布广泛,同时这也与两省庞大的人口基数存在一定联系。

(二) 人口流动网络中城市节点等级与平均流动距离存在一定的正相关,空间距离仍然是人口流动的重要束缚,边缘节点总是趋向于与核心节点优先建立联系

人口流动与两个城市的经济、人口规模、地理区位等传统引力因素存在着密切关系。因此,需要考虑空间距离对于人口流动的束缚力。通过 GIS 空间分析工具来测算 6 个中心节点连接边线的距离(两个城市行政中心的直线距离),用以表示人口在两个城市之间的流动距离。结果发现,人口流动的平均距离随着节点度值的降低而降低(见表 4),这表明在人口流动网络中,节点等级与人口流动距离存在一定的正相关性,网络地位越高的城市影响的范围越广,如北京与上海两个全国性的网络中心城市,能够与新疆、西藏等边缘节点保持高值联系,平均流动距离高达 1 000 公里以上。另外,高值联系城市数目的比例也随着节点度值的降低而降低,这就反映出人口流动网络连接的择优

表 4 “两大四小”中心节点网络联系边的平均距离

度值	节点名称	总连接边数		平均距离(公里)		高边权(8~10)			
		流入	流出	流入	流出	流入平均距离	流出平均距离	流入连接边数	流出连接边数
4001	北京	277	270	1179.77	1167.04	983.96	1075.34	125(45.13)	155(57.41)
2778	上海	210	240	1156.87	1100.91	933.75	1048.08	92(43.81)	85(35.42)
1775	重庆	153	133	929.46	937.21	791.54	926.41	52(33.99)	56(39.16)
1349	西安	120	119	916.11	879.18	795.43	791.21	40(33.33)	40(33.62)
1019	广州	80	100	874.82	837.35	731.38	790.32	21(26.25)	31(31.00)
994	武汉	90	106	776.85	780.78	638.28	754.26	16(17.78)	19(17.93)

注:括号内数字为百分比。

取向特征,即节点总是趋向于与高等级节点(核心节点)优先建立联系,体现出核心节点(高度值)强大的吸引力与控制力。但是对比发现,平均距离与边权值也呈现一定的负相关关系,连接线的边权越高,线两端的节点之间的距离越短,体现了网络连接的邻近取向特征。在图4中可以看出,相对于中低值联系分布,高值联系分布更倾向于在核心节点的周围聚集,这在一定程度上表明空间距离依然是人口流动的重要束缚之一。

### 三、地级市人口流动网络的复杂性分析

(一) 地级市人口流动网络的入度与出度累计概率分布均呈现明显的幂律分布,体现出“二八定律”,符合无标度网络特性,反映出人口流动空间分布的非均衡性

度分布函数是研究复杂网络系统结构的宏观统计特征的重要指标。为了研究中国城市人口流动网络的复杂性特征,本文分别对全网络入度与出度的累计概率分布进行拟合分析(见图5)。结果表明,该网络的入度与出度累计概率分布均呈现明显的幂律分布( $R^2=0.9312$  和  $0.9319$ ),而采用指数函数拟合,拟合度不及幂函数( $R^2=0.6389$  和  $0.6469$ )。这说明城市间人口流动网络是一种典型的复杂网络,度分布指数( $\gamma=0.836$  和  $0.834$ )低于一般复杂网络的平均水平(2~3),说明该网络的节点度数存在更高的异质性:分布跨度大(典型的长尾分布)、呈现出严重的不平衡性。其中有近80%的节点中心度不足90,体现出帕累托分布的“二八定律”,这就说明网络符合无标度网络特性。也就是说节点的度越大,其网络服务能力也越强,造成人口流动分布出现集聚效应和极端分化现象。

(二) 整个网络表现出较高的易达性和运行效率,并呈现典型的小世界网络特征,表明人口可以在城市间自由有序流动

利用复杂网络分析工具 pajek,对整个网络的平均路径长度和聚类系数进行计算,以考察该网络结构的可达性和集聚性。计算结果显示,整个城市网络的平均路径长度为1.995,在全部可能存在联系的111 222条连接路径中,不需中介就直接产生联系(长度为1)的最短路径有6 388条,占5.74%;长度为2(只需要通过1次中介就可连接)的最短路径有98 978条,占88.99%,因此,累计有近95%的节点之间最多只需要一次中转就可产生联系。整个网络表现出较高的易达性和运行效率。

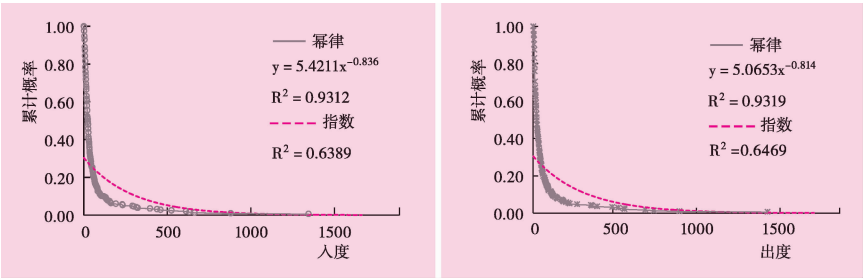


图5 网络入度、出度累计概率分布

聚类系数反映的是网络中节点的互联水平,即与某节点相连接的两个节点是否也存在连接关系。结果显示,整个城市网

络的平均聚类系数为 0.341,表现出较强的集聚性。通过分析网络中每个节点的聚类系数,发现聚类系数最大值( $C=1$ )的节点的度数均低于 10,而高度值节点如北京、上海、重庆、西安、广州和武汉,聚类系数均小于 0.1。这表明在网络中低度节点主要倾向于与高度节点连接,从而导致产生较高的聚类系数。

本文通过构建一个 334 个节点的随机网络来对比分析(见表 5),结果显示,随机网络的平均路径长度为 2.05,略大于实际网络,而聚类系数为 0.118,远低于实际网络,这也表明实际的城市网络呈现出典型的“小世界网络”特征。这种网络的无标度性与小世界特征说明网络具有较高的连通性和可达性。同时也表明,随着中国城市化进程的加速,交通设施的不断完善,引导着人口在城市间自由有序地流动。

表 5 实际网络与随机网络特征指标对比

指 标	实际网络		随机网络
	长度	比例(%)	
平均路径长度	1	5.74	2.050
	2	88.99	
	3	5.27	
聚类系数	0.341		0.118

四、结论与讨论

本文通过“百度迁徙”海量个体居民的出行信息数据,来构建城市之间的联系网络,采用复杂网络方法从整体层面上研究中国城市人口流动网络结构,构建了中国城市人口流动网络模型,分析其复杂的空间结构与网络特征。结果发现以下特征。

第一,运用复杂网络分析工具 Gephi 来计算中国 334 个地级及以上城市的人口流动网络度值,发现各地级市之间的网络地位差异较大,城市行政级别的高低与网络度值存在一定的相关性。从对人口流入的吸引力与人口流出的辐射能力看,行政等级高的城市都处于明显的优势地位。城市人口流动的入度与出度网络在空间分布上呈现相似的集聚性,大部分城市的入度与出度之间保持相对平衡,表现出人口流入流出的相对平衡性。

第二,基于城市节点的网络总度值对地级市进行分级,城市间人口流动网络显现出明显的层次性,其中东部沿海三大城市群仍是人口流动的热点区域,但西部地区中心城市(重庆、成都和西安)的网络地位大幅提升;层级分布也反映出基于人口流动的城市网络体系并不完全遵循基于经济条件的城市体系,城市在人口流动网络的地位和作用高低不仅取决于经济与人口规模,也是与地理区位和交通条件等因素共同影响的结果。

第三,综合考虑城市节点层级分布与城市间网络联系强度,发现中国地级市人口流动网络流向分布上呈现“两大四小”多中心网络空间格局;尤其是在高值联系流上更为明显。六大中心节点在整个网络中处于绝对主导地位,但各自的流场分布又有所差异,主要体现在流入来源与流出去向分布区域、高中低值联系影响区域等两方面。

第四,对六大中心节点连接边的距离测算,人口流动网络中城市节点网络等级与平均流动距离存在一定的正相关性,在一定程度上表明空间距离仍然是人口流动的重要束

缚,边缘节点总是趋向于与更高等级节点(核心节点)优先建立联系,反映出人口流动网络连接的首选取向以及局部的邻近连接特征。

第五,通过对人口流动网络节点的度累计概率分布、平均路径长度和聚类系数等指标进行测度分析,发现地级市人口流动网络的入度与出度累计概率分布均呈现明显的幂律分布,体现出帕累托分布的“二八定律”,这表明网络符合无标度网络特性,并呈现出典型的“小世界”网络特征,表明了人口在城市间自由有序地流动。

由于后台原始数据难以获取,本文采用权重赋值法将“百度迁徙”可视化数据进行标准化处理,虽然获取的是人口流动的相对数量,但可以真实反映出人类活动在地理空间上的投影,以此来研究基于人口流动的加权有向型城市网络的空间特征与相互作用是完全可行的,也得到了有效的验证。当然,基于个体行为轨迹记录的“大数据”使构建更加复杂的模型成为可能,然而,人口流动的影响因素复杂多元,因此,寻找一个数据筛选收集的合适尺度(时间与空间)就显得十分重要。由于考虑到后台运算量等问题,本文数据覆盖对象有一定的局限性,另外,数据中难以获取人口的社会属性(如职业、年龄、出行目的、逗留时间等),因此,本文仅对城市间人口流动的宏观特征和空间格局进行了尝试性研究,人口流动的内在动力机制,还有待后续的探索研究。

#### 参考文献:

1. 鲍常勇(2007):《我国 286 个地级及以上城市流动人口分布特征分析》,《人口研究》,第 6 期。
2. 陈锐等(2014):《基于改进重力模型的省际流动人口的复杂网络分析》,《中国人口·资源与环境》,第 10 期。
3. 丁金宏等(2005):《中国人口流动的区域差异与流场特征》,《地理学报》,第 1 期。
4. 段成荣、杨舸(2009):《我国流动人口的流入地分布变动趋势研究》,《人口研究》,第 6 期。
5. 侯贺平等(2013):《基于改进辐射模型的乡镇人口流动网络研究》,《中国人口·资源与环境》,第 8 期。
6. 劳昕、沈体雁(2015):《中国地级以上城市人口流动空间模式变化——基于 2000 和 2010 年人口普查数据的分析》,《中国人口科学》,第 1 期。
7. 刘军(2009):《整体网分析讲义: UCINET 软件实用指南》,格致出版社。
8. 刘望保等(2012):《中国省际人口迁移流场及其空间差异》,《经济地理》,第 2 期。
9. 孙峰华等(2006):《2005 年中国流动人口分布的空间格局及其对区域经济发展的影响》,《经济地理》,第 6 期。
10. 王桂新等(2012):《中国省际人口迁移区域模式变化及其影响因素:基于 2000 和 2010 年人口普查资料的分析》,《中国人口科学》,第 5 期。
11. 王国霞等(2012):《20 世纪末中国迁移人口空间分布格局——基于城市的视角》,《地理科学》,第 3 期。
12. 王珏等(2014):《基于社会网络分析的长三角地区人口迁移及演化》,《地理研究》,第 2 期。
13. 严善平(2007):《中国省际人口流动的机制研究》,《中国人口科学》,第 1 期。
14. 于涛方(2012):《中国城市人口流动增长的空间类型及影响因素》,《中国人口科学》,第 4 期。
15. 张永庆等(2006):《城市人口迁移的网络特征》,《东北大学学报(自然科学版)》,第 2 期。

(责任编辑:朱 犁)