

目 次

论文

平台企业的灵活积累机制

——一个合约理论的视角 ..... 滕 飞( 1 )

信息逆差:代码权力膨胀与公众权力让渡 ..... 施颖婕( 21 )

大语言模型能用来给主题模型中的主题编码吗?

——兼论大语言模型在社会科学领域应用的前景  
..... 李 代 张博伦 周浥莽( 33 )

数字时代的科技公司与国家治理

——一个综合性的分析框架 ..... 钟瑞雪( 60 )

新闻大数据视角下测度区域一体化水平的新路径

——以长三角地区为例 ..... 李 军 董方杰( 84 )

解密数字社会中的幸福密码

——不同互联网使用程度下社会信任感和社会公平感对主观幸福感的  
影响机制研究 ..... 郭媛媛 张 腾( 103 )

## 研究报告

- 数字时代安全科技价值报告 …… 陆菲菲 秦 昊 张友红 吕 鹏(133)
- 问题触发的算法模型响应机制探索 …… 许正军 袁 岳(153)

## 译文

数字健康的社会技术伦理

——生命伦理学进路的批判和延伸

…………… 詹姆斯·肖 约瑟夫·多尼娅 著 何 丽 译(166)

## 书评

数字资本主义和智能技术的当代批判

——评《过度智能》 …… 周银知(186)

## 访谈

通过人类行动思考机器智能

——专访人类学家露西·萨奇曼 …… 露西·萨奇曼 胡万亨(198)

# CONTENTS

## THESIS

|                                                                                                                                                                                                       |                                        |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------|
| The Flexible Accumulation Mechanism of Platform Enterprises: A Contract Theory Perspective .....                                                                                                      | Teng Fei( 1 )                          |
| “Information Deficit”: Code Power Inflation and Public Power Cession ...                                                                                                                              | Shi Yingjie( 21 )                      |
| Can LLM Label Topics from A Topic Model: The Future of LLM in Social Science .....                                                                                                                    | Li Dai, Zhang Bolun, Zhou Yimang( 33 ) |
| Technology Companies and National Governance in the Digital Age: A Comprehensive Analytical Framework .....                                                                                           | Zhong Ruixue( 60 )                     |
| A New Approach to Measuring Regional Integration from the Perspective of Big Data in News: A Case Study of the Yangtze River Delta Region ...                                                         | Li Jun, Dong Fangjie( 84 )             |
| Decrypting the Happiness Code in the Digital Society: A Study on the Influence Mechanism of Social Trust and Fairness Perception on Subjective Well-Being under Different Internet Usage Levels ..... | Guo Yuanyuan, Zhang Teng( 103 )        |

## RESEARCH REPORTS

|                                                                                   |                                                    |
|-----------------------------------------------------------------------------------|----------------------------------------------------|
| Report on the Value of Security Technology in the Digital Age .....               | Lu Feifei, Qin Hao, Zhang Youhong, Lyu Peng( 133 ) |
| Exploration of Algorithmic Model Response Mechanisms for Problem Triggering ..... | Xu Zhengjun, Yuan Yue( 153 )                       |

## TRANSLATED TEXT

The Sociotechnical Ethics of Digital Health: A Critique and Extension of Approaches from

Bioethics ..... written by J. Shaw, J. Donia; trans. by He Li( 166)

## BOOK REVIEW

A Contemporary Critique of Digital Capitalism and Smart Technologies: Review of *Too*

*Smart* ..... Zhou Yinzhi( 186)

## INTERVIEW

Thinking Machine Intelligence Through Human Actions: An Interview with Anthropologist

L. Suchman ..... L. Suchman, Hu Wanheng( 198)



# 新闻大数据视角下测度区域 一体化水平的新路径 ——以长三角地区为例

李 军 董方杰<sup>\*</sup>

**摘要:** 本文是利用新闻大数据对区域一体化水平进行测度的首次尝试。本文采集并处理了长三角地区 41 座城市的新闻媒体在 2016—2022 年发布的所有新闻报道,利用大数据文本分析和社会网络分析技术测度长三角一体化水平,并以长三角一体化指数呈现历时态(时间)与共时态(空间)下一体化的趋势与差异,探索了一条借助新闻大数据测度区域一体化水平的新路径。

**关键词:** 新闻大数据 社会网络分析 长三角一体化 长三角城市群

## 一、引言

城市化是经济发展的关键动力(Bairoch, 1991; Bertinelli & Black, 2004),对现代社会的基本架构产生了深远影响。经济地理学揭示了城市作为各类生产要素聚集中心的重要性(Storper, 2013)。尤其是自 20 世纪中叶以来,随着经济全球化的快速推进,区域一体化的进程也在不断加速。在民族国家内部,出现了一种新型区域联合体,被称为“城市群”(megalopolis),它展示了经济和城市发展从点到面的跨越,也标志着城市化进入了更为发达的空间形态(Fang & Yu, 2017),从相互竞争转变为协同发展(Fang & Yu, 2020)。

全球范围内有五个被广泛承认的世界级城市群,分别是美国东北部大西洋沿岸城市群、北美五大湖城市群、日本太平洋沿岸城市群、英伦城市群以及欧洲

---

<sup>\*</sup> 李军,南京大学社会学院。董方杰,南京大学社会学院。

西北部城市群(Gottmann, 1976)。随着中国改革开放后城市化的快速推进,城市群逐渐成为中国新一轮城镇化的重要推动力(Zhao, Zhang & Zhao, 2021)。目前,长三角城市群已成为六个世界级城市群之一(方创琳、姚士谋、刘盛和等,2011;胡雅龙,2010),充分展示了长三角城市群在经济和战略上的重要地位。

相较于其他世界级城市群,长三角城市群展现出明显的国家规划特色。2010年5月,中国批准实施《长江三角洲地区区域规划》,正式确定了长三角地区的范围为江浙沪,并明确了其发展战略定位:成为亚太地区重要的国际门户、全球重要的现代服务业和先进制造业中心,以及具有较强国际竞争力的世界级城市群。该规划主要集中在区域规划方面,核心内容是划定各城市的发展重点。其目的在于避免这些城市之间的恶性竞争,明确它们各自的产业定位,以解决区域城市发展过程中可能出现的同质化问题。然而,该规划并未充分体现对区域一体化和协同发展的思考。

直至2018年6月,国家主席习近平在首届中国国际进口博览会上宣布支持长江三角洲地区的一体化发展,并将其提升为国家战略。紧随其后,2019年12月1日,中共中央、国务院印发《长江三角洲区域一体化发展规划纲要》作为指导长三角地区当前及今后一个时期一体化发展的纲领性文件,这是中央层面首次将安徽省整体纳入长三角城市群的范畴。文件指出,长三角地区是我国经济发展最活跃、开放程度最高、创新能力最强的区域之一,在国家现代化建设大局和全方位开放格局中具有举足轻重战略地位。推动长三角一体化发展,增强长三角地区的创新能力和竞争能力,提高经济集聚度、区域连接性和政策协同效率,对于引领全国高质量发展、建设现代化经济体系具有重大意义。在产业集聚效应和政策规划引领的双重推动下,长三角城市群的一体化进程进入了一个新的发展阶段。

与此同时,学界对于测度区域一体化水平的讨论正在蓬勃发展,主要有两条研究路径。其中一条路径遵循传统的经济学研究范式,另一条则源自网络社会的崛起和“流动空间”(space of flow)(Castells, 2000)概念的提出。传统测度区域一体化水平的路径主要使用现有的经济统计数据,尽管在某种程度上可以反映区域一体化的程度,但这些统计数据存在一个重要问题,即

缺乏方向性尺度,无法真实地测量区域内各城市因各类互动而形成的网络关系。不过,近年来大数据研究的兴起为弥补上述研究的缺陷提供了可能性。

## 二、文献回顾

本文的主要目的是利用新闻大数据的方法,为测度区域一体化水平提供一条新的研究路径。在过去关于测度区域一体化水平的研究中,可以明显看出两条清晰的研究路径,这也反映了区域研究一体化的范式转型。

第一条研究路径采用传统的经济学研究方法。该路径利用所在区域的各类社会经济指标,主要考察人口、GDP、财政支出等经济活动在空间距离上的集聚。这类研究路径通常将属性数据(data of city attributes)(李涛、周锐、苏海龙等,2015; Beaverstock, Smith & Taylor, 2000; Taylor, 1999)按研究重点划分为不同的指标维度,然后借助各类经济学模型来测度区域经济一体化水平。例如,引力模型(gravity model)广泛应用于城市之间联系强度的测算(李涛、周锐、苏海龙等,2015;许静涵、林礼娜、武义天等,2021; Chen, 2009; Chen & Huang, 2018; Fang, Sheng & Li, 2018; Han, Cao & Liu, 2018; Zhao, Zhang & Zhao, 2021),也可用于测算边界效应,衡量地区边界导致市场分割的程度(李晓、王小彬,2017;王晔倩、林理升,2006;肖海平、谷人旭,2011;行伟波、李善同,2009;赵永亮、徐勇、苏桂富,2008;赵永亮、才国伟,2009; Broker, 1984; Engel & Rogers, 1994; McCallum, 1995; Poncet, 2003);拟合分析模型(邓淑芬、朱佳翔、钟昌宝,2018;刘干、陆叶,2018; Ibrahim, 2008; Kim, 2001; König & Ohr, 2013; Wenjie, 2022)通常使用聚类方法对各类指标进行层次化,以得出城市群一体化发展水平的差异。具体到长三角一体化水平的相关研究中,孙红梅等人采用熵值法构建了2010—2017年长三角一体化协同评价指标体系,并通过全局自相关与局部空间自相关散点分析发现长三角一体化协同具有较强的空间正相关性(孙红梅、姚书淇,2020);刘志彪等人提出了一种综合指标体系,涵盖了十个维度,即空间、市场、产业、创新、交通、信息、贸易、公共服务、生态环境和制度,以此来测量长三角一体化程度(刘志彪、孔令池,2019);李世奇等人采用基于专家咨

询的主成分分析评价方法构建了长三角一体化指标体系,该方法侧重于市场统一性、要素同质性、发展协同性和制度一致性,为理解长三角一体化程度提供了新的视角(李世奇、朱平芳,2017)。

第二条研究路径源自卡斯特尔(M. Castells)在《网络社会的崛起》一书中提出的“流动空间”概念。他认为,社会由资本、信息、技术等“流”组成,城市网络位于流动空间中。换言之,城市之间的网络关系而非等级结构塑造了世界和区域城市网络,使得流动空间成为信息社会的主导性空间组织形式(Castells, 2000)。这一观点直接推动了城市研究从场所空间(space of place)转向流动空间(space of flow),城市网络的研究范式得到了众多学者的关注(Beaverstock, Smith & Taylor, 2000; Taylor, 1999)。为了弥补属性数据无法动态展现城市网络关系的缺陷,学者们开始使用关系型数据(relational data)。在测度区域一体化水平与城市网络关系方面,国内外学者运用交通流(physical transportation)(Derudder & Witlox, 2005; Mahutga, Ma & Smith et al., 2010; Smith & Timberlake, 2002; Van Nuffel, Derudder & Witlox, 2010)、贸易流(Afesorgbor, 2017; Lee, Owen & Mensbrughe, 2009; Rossi, Beaverstock & Taylor, 2007; Wilson & Otsuki, 2007)和电信流(Graham & Marvin, 1995; Hall & Pain, 2006; Rutherford, Gillespie & Richardson, 2004)等流量数据指标来进行。

在长三角城市群研究中,不同学者采用了不同方法来探索该地区的特点。朱查松等人使用2001年、2009年长三角企业分支机构数据,通过分析企业联系网络来研究长三角城市网络的区域空间组织特征(朱查松、王德、罗震东, 2014)。唐子来等人基于企业区位数据,采用总部分支法,分析了京津冀、长三角和珠三角地区的城市关联网。他们发现,北京、上海、广州和深圳不仅是中国的核心城市,还分别是三个地区的“门户城市”,承担着连接全球经济网络和辐射区域腹地的双重作用(唐子来、李涛,2014)。李涛等人以关系型大数据为基础,选择了企业分支数据、信息流数据和交通流数据,运用关联网分析和多变量分析方法,对长三角地区30个城市的区域经济一体化指数进行了测量(李涛、周锐、苏海龙等,2015)。这种方法不仅考虑了企业间的联系,还综合考虑了信息和交通流动,为评估长三角地区的经济一体化程度提供了多维度视角。此外,一些学者还通过探索城市网络结构(唐子来、赵渺希,2010;赵渺希、唐子来,

2010)、创新协同网络(白俊红、蒋伏心,2015;徐宜青、曾刚、王秋玉,2018;张贵、温科,2017)等方式,对长三角城市群进行了研究。

近年来,大数据研究的兴起为我们从流动空间的角度更细致、翔实地理解区域一体化进程提供了新的视角。大数据的优势在于充分利用互联网中的大量数据,以近乎全样本的方式对社会事实进行全面观察、计算和分析,为研究大空间的宏观或微观经济现象提供了全新的路径。虽然曾有学者从事过相关研究,如熊丽芳等人采用创新方法构建了中国东部、中部和西部三大经济区的城市网络。他们利用百度指数中的用户关注度数据,反映了经济区内各城市之间的互联互通程度。通过模拟计算百度信息流,可以描绘出中国三大经济区内城市之间的网络关系(熊丽芳、甄峰、王波等,2013;熊丽芳、甄峰、席广亮等,2014)。甄峰等人通过分析微博用户数据,探究了微博社会空间视角下的城市网络结构,发现中国城市网络在微博平台上呈现明显的等级关系和层级区分,表明城市的网络连接度与城市等级存在一定的相关性,揭示了城市在数字空间中的社会互动模式,为理解城市间的互联互通和影响力分布提供了新的路径(甄峰、王波、陈映雪,2012)。然而,这种基于个体行为的大数据虽然能够从某个侧面“复刻”区域一体化的程度,但无法完整呈现区域一体化的真实实践,而新闻大数据则弥补了上述研究的不足。

### 三、数据来源与研究方法

#### (一) 新闻大数据的数据来源

本文所采用的数据来源为锐研中文新闻数据库。作为国内学术领域知名的中文新闻数据库之一,该数据库目前覆盖了全网各类电子报纸、政府官方新闻、大型新闻门户等超过1300家互联网新闻媒体来源的全文索引数据。该数据库还提供了新闻内容主题分类,并能提取各类结构化变量,如时间、地点、人物、机构、主题、情感和影响力。借助该数据库,我们可以进一步搜集指定媒体范围和时间范围的全文数据。

长三角城市群涉及三省一市,共有40个地级市和1个直辖市。本文选取

的媒体主要为各地级市的日报,如《南京日报》《杭州日报》《合肥日报》;个别城市的日报不以城市名称命名,如六安《皖西日报》、宿州《拂晓报》。此外,上海市选取了一份日报《文汇报》和一份晚报《新民晚报》,江苏省、浙江省和安徽省也各选取了一份省级日报和省级晚报。总计涉及 48 家报纸从 2016 年 1 月 1 日至 2022 年 12 月 31 日的所有新闻报道数据,共计超过 505 万篇(表 1)。

| 表 1 2016—2022 年长三角城市群新闻报道总数 |         |         |         |         |         |         |         |           |
|-----------------------------|---------|---------|---------|---------|---------|---------|---------|-----------|
| 年份                          | 2016 年  | 2017 年  | 2018 年  | 2019 年  | 2020 年  | 2021 年  | 2022 年  | 总计        |
| 新闻报道总数                      | 616 565 | 633 153 | 748 157 | 778 321 | 860 171 | 719 992 | 703 459 | 5 059 818 |

我们从 505 万余篇新闻报道中进一步筛选出涉及超过 2 座城市的新闻文本,最终得到了 669 346 篇共现新闻报道作为研究的数据总集(表 2)。

| 表 2 2016—2022 年长三角城市群共现新闻报道总数 |        |        |         |         |         |        |        |         |
|-------------------------------|--------|--------|---------|---------|---------|--------|--------|---------|
| 年份                            | 2016 年 | 2017 年 | 2018 年  | 2019 年  | 2020 年  | 2021 年 | 2022 年 | 总计      |
| 共现新闻报道总数                      | 89 582 | 94 540 | 107 156 | 113 110 | 107 248 | 85 573 | 72 137 | 669 346 |

(二) 长三角一体化指数的构建策略

区域一体化发展首先在于加强区域内不同城市之间的连接,这种“连接”可以体现在多个方面,包括交通互联、文化交流、科创协同、招商引资、环保协作、医疗卫生和教育资源共享等。新闻是对新近发生、正在发生或仍能产生社会效应的事实进行的报道,城市之间的合作交流事件通常会被新闻所记录。当一篇新闻报道同时提及 2 个城市时,这两个城市往往存在某种“连接”。因此,我们可以通过文本分析技术统计一段时间内同时提及 2 个城市的文本次数,从而近似地测量这两个城市之间的绝对连接强度。表 3 展示了 2016 年上海与长三角各城市在新闻大数据集中的共现次数。

表 3                    2016 年上海与长三角各城市在新闻大数据集的共现次数

| 城市   | 共现<br>次数 | 城市  | 共现<br>次数 | 城市  | 共现<br>次数 | 城市   | 共现<br>次数 | 城市  | 共现<br>次数 |
|------|----------|-----|----------|-----|----------|------|----------|-----|----------|
| 南京市  | 7789     | 盐城市 | 895      | 嘉兴市 | 1714     | 合肥市  | 3327     | 黄山市 | 762      |
| 无锡市  | 2445     | 扬州市 | 1028     | 湖州市 | 1429     | 芜湖市  | 1265     | 滁州市 | 792      |
| 徐州市  | 1808     | 镇江市 | 985      | 绍兴市 | 1709     | 蚌埠市  | 499      | 阜阳市 | 488      |
| 常州市  | 1822     | 泰州市 | 1191     | 金华市 | 1236     | 淮南市  | 672      | 宿州市 | 275      |
| 苏州市  | 4469     | 宿迁市 | 469      | 衢州市 | 911      | 马鞍山市 | 562      | 六安市 | 462      |
| 南通市  | 1891     | 杭州市 | 9380     | 舟山市 | 815      | 淮北市  | 240      | 宣城市 | 592      |
| 连云港市 | 613      | 宁波市 | 3495     | 台州市 | 890      | 铜陵市  | 404      | 亳州市 | 250      |
| 淮安市  | 429      | 温州市 | 1728     | 丽水市 | 815      | 安庆市  | 742      | 池州市 | 389      |

如果一篇报道同时提及超过 2 个城市,那么每个城市两两之间都会被计算一次。考虑到新闻报道的总数量并不是恒定的(尽管报纸的版面总数大致相同,但文章数量通常会有变化),因此我们可以将某一年的绝对连接强度除以该年提及多个城市的报道总数(共现新闻报道总数),以得到相对连接强度。

然而,需要注意的是,并非每次新闻报道提及多个城市都意味着这些城市之间存在较强或明显的“连接”关系。例如,一篇报道可能只是介绍了一位来自江苏泰州的上海人。这类新闻数量相对较少,而且在一定程度上代表了两个城市之间的弱联系。对于海量的文本大数据而言,这种误差对最终的统计结果影响并不大。因此,使用上述统计方法来近似地测量不同城市之间的连接强度是可行的。

在特定区域内,我们可以将一个城市与该区域内其他城市的所有相对连接强度求和,以衡量该城市与整个区域的一体化程度;将所有城市与整个区域的融合程度再次汇总,就可以得到整个区域的一体化指数。

在一个具有  $N$  个城市的区域内,如果用  $S_{ij}$  来表示同时报道城市  $i$  和城市  $j$  的相对连接强度,用  $CI_i$  表示城市  $i$  与整个区域的一体化程度,用  $RI$  表示区域的整体一体化程度,则:

$$CI_i = \sum_{j=1}^N S_{ij} (j \neq i)$$

$$RI = 1/2 \sum_{i=1}^N CI_i$$

区域一体化指数  $RI$  的可解释意义在于:对于每一篇涉及多个城市的新闻报道,其平均包含的“边”(两两城市对)的数量。如果所有的新闻报道仅包含 2 个不同城市,则  $RI=1.0$ 。正是因为一篇新闻报道可能涉及 2 个以上的城市,“边”的数量会呈现非线性放大(3 个城市共有 3 条边,4 个城市共有 6 条边,5 个城市共有 10 条边,依此类推),因此  $RI>1.0$ 。如果单篇新闻报道普遍涉及越来越多的城市,则一定程度上说明城市之间的连接得到了加强,区域融合也更加深入。虽然单篇新闻平均报道的城市数量也可以用来作为区域融合度的测量指标,但“边”的平均数量更能反映其变化幅度,从而该指标具有更好的可表示性。

## 四、新闻大数据下长三角一体化进程的时空演变

### (一) 长三角一体化程度与共现新闻报道总数

为了研究长三角一体化程度的历时态变化情况,我们选择了 2016—2022 年连续 7 年的新闻大数据进行研究。以年为单位,首先计算了每年涉及多个城市的新闻报道占当年所有新闻报道的比例,以及长三角城市群的区域一体化指数。如图 1 所示,2016—2019 年的多城共现报道比例在波动中增大,表明长三角城市间互动的新闻比例在增加。这一现象在一定程度上验证了区域融合的持续发展,与长三角一体化指数的波动上升趋势相一致。2020 年后,多城共现报道比例有所下降,表明区域内城市间的合作和联动在新闻报道中的关注度有所降低。这一趋势在长三角一体化指数中也得到了验证。不过,令人欣喜的是,2022 年长三角一体化指数再次上升,表明区域整体融合再次展开。

### (二) 长三角各城市与区域一体化程度

长三角一体化指数的核心计算逻辑在于长三角 41 个城市之间的共现互动。



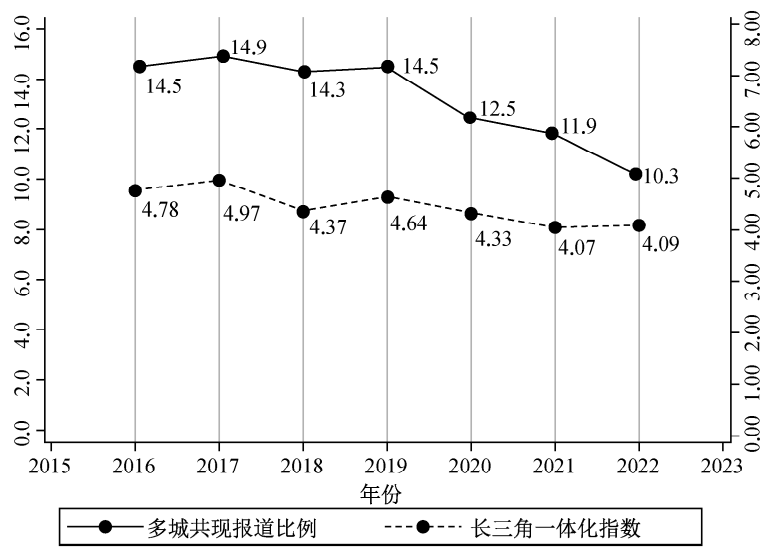


图 1 多城共现报道比例与长三角一体化指数

在前文中,我们定义了  $CI_i$  表示城市  $i$  与整个区域的一体化程度,接下来呈现的长三角各城市一体化指数,以 2016 年上海的一体化指数为基准(设为 100),进而标准化了 2016—2022 年各城市的指数值。我们采用空间分布和时间序列的双重维度来展现这一指数,其中表 4 根据历年指数平均值划分了城市的层级。

从空间分布来看,上海的一体化指数显著高于其他城市,凸显了其在长三角一体化进程中的中心地位,可视作区域融合的引擎。此外,南京和杭州的指数紧随其后,成为长三角融合的两大支点,而合肥、苏州、宁波等城市则作为次中心地带,在整个一体化进程中扮演着重要角色。第四层级包括无锡、徐州等 7 座城市,第五层级包括扬州、安庆等 17 座城市,第六层级包括衢州等 11 座城市。可以看到,长三角地区的一体化逐渐形成了一个“中心—两翼—多次中心”的层级格局。在各省内部,江苏省呈现的地理分布趋势为各城市的一体化指数自南至北、自东至西逐渐降低,浙江省各城市一体化指数也呈现出自北向南、自东向西逐步降低的趋势,江浙两省一体化指数的分布差异展现的是上海的辐射带动作用。安徽则较为特殊,合肥—南京的夹心地带一体化指数显著高于其他地区,体现的是南京在安徽诸城中独特的城市影响力。从时间序列角度来看,中心地带

的城市如上海、南京、杭州、苏州、宁波等,其一体化指数稳中有进;边缘地带城市的指数波动较大。

| 表 4            | 长三角各城市一体化指数一览表                                                                                                                                                             |
|----------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 长三角一体化<br>指数等级 | 城市(一体化指数)                                                                                                                                                                  |
| 第一层级(100 以上)   | 上海市(107.8)                                                                                                                                                                 |
| 第二层级(70—100)   | 南京市(92.3)、杭州市(83.7)                                                                                                                                                        |
| 第三层级(40—70)    | 合肥市(60.9)、苏州市(57.8)、宁波市(46.0)                                                                                                                                              |
| 第四层级(30—40)    | 无锡市(36.9)、徐州市(34.6)、芜湖市(32.7)、南通市(31.3)、温州市(30.8)、嘉兴市(30.4)、常州市(30.1)                                                                                                      |
| 第五层级(20—30)    | 扬州市(29.5)、安庆市(27.8)、绍兴市(27.7)、金华市(26.8)、湖州市(26.5)、滁州市(26.1)、马鞍山市(25.8)、盐城市(25.0)、镇江市(24.8)、阜阳市(24.5)、宣城市(23.4)、台州市(23.3)、泰州市(22.2)、六安市(22.1)、蚌埠市(21.7)、黄山市(21.3)、淮南市(20.6) |
| 第六层级(0—20)     | 衢州市(19.7)、连云港市(19.4)、淮北市(19.4)、丽水市(19.0)、铜陵市(18.8)、宿迁市(18.6)、舟山市(17.9)、池州市(17.7)、宿州市(17.6)、淮安市(17.5)、亳州市(16.8)                                                             |

(三) 长三角一体化进程中的城市社会网络分析

在前文中,我们定义了  $S_{ij}$  来表示同时报道城市  $i$  和城市  $j$  的相对连接强度,因此通过  $S_{ij}$  建立  $41 * 41$  的网络共现矩阵来呈现长三角各城市之间的网络互动。<sup>①</sup> 我们将从网络密度、中心度、凝聚子群三个层面展开分析。

首先,从网络密度分析来看。网络密度衡量了网络中节点之间联系的密集程度。在数学上,它被定义为实际连接数与最大连接数之间的比率。在社会网络的实际运用中,网络密度越大,网络中节点之间的联系越紧密,网络对节点的影响可能性越大。一般来说,在真实的世界中,网络密度总是低于 0.5 (Mayhew & Levinger, 1976)。在表 5 中,我们运用 Ucinet 对长三角各城市之间的共现互动网络进行了测算,发现长三角城市网络的密度一直居于高

<sup>①</sup> 期刊版面有限,无法全面呈现图片信息,有意者请向作者索要。

位,意味着其一体化程度很高。虽然在新冠疫情影响下有所波动,但依然稳步上升,展现了长三角各城市之间的融合韧性,一体化进程始终在稳步推进。

| 表 5                      2016—2022 年长三角各城市的共现网络密度 |        |        |        |        |        |        |        |
|---------------------------------------------------|--------|--------|--------|--------|--------|--------|--------|
| 年份                                                | 2016 年 | 2017 年 | 2018 年 | 2019 年 | 2020 年 | 2021 年 | 2022 年 |
| 数值                                                | 0.3439 | 0.3610 | 0.3561 | 0.3427 | 0.3341 | 0.3256 | 0.3841 |

其次,我们从四大中心度的角度去考察长三角城市网络中的各个节点(城市)。由于涉及城市众多,并且中心度的分析主要用于评估节点的重要性,因此我们在图 2 中列举了 2016—2022 年长三角城市中四大中心度位列前五的城市,事实上这也是长三角一体化指数最高的几个城市,验证了我们构建长三角一体化指数的科学性。

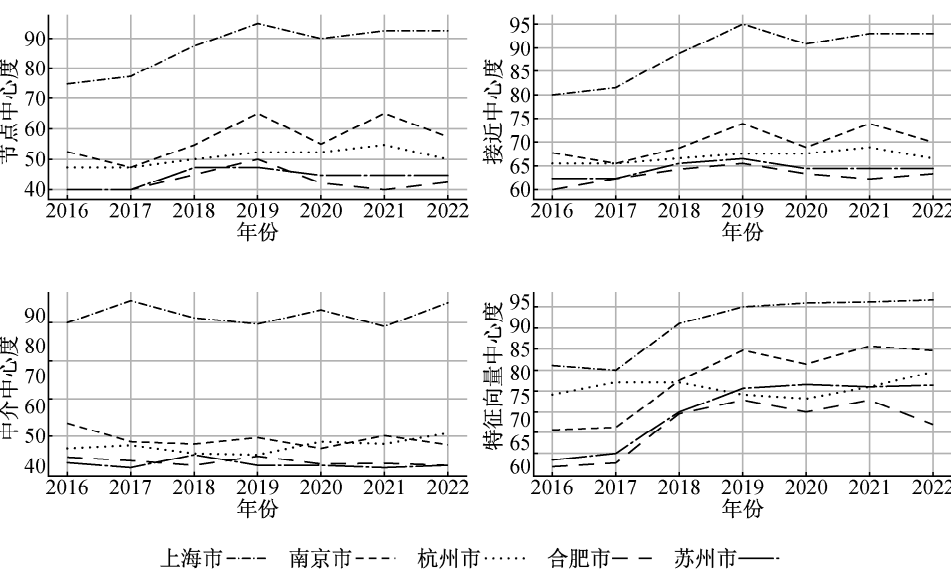


图 2    2016—2022 年长三角五大城市的中心度演变趋势

在社会网络分析中,通常使用节点中心度(degree centrality)或接近中心度(closeness centrality)来评估节点的重要性。节点中心度衡量了节点直接相连的数量,节点中心度越高,节点在网络中的重要性就越大,即位于网络的中心位置。然而,节点中心度忽略了节点之间的距离因素,因此引入接近中心度来弥补这个

局限。接近中心度通过计算节点到其他所有节点的最短距离来反映节点靠近中心的程度。接近中心度越高,节点在网络中的位置越靠近中心,其重要性也越显著。无论是节点中心度还是接近中心度,2016—2022 年期间,上海都是长三角地区最重要的网络节点,其次是南京、杭州、苏州和合肥。需要特别注意的是,2018—2019 年合肥的节点中心度和接近中心度有了显著提升,逼近甚至超过苏州。这说明自 2018 年以来,长三角一体化进程首次全面纳入安徽,以合肥为代表的安徽省域全面加强了与其他地区的联系与互动。从各城市的数值趋势来看,五大城市的节点中心度和接近中心度都呈增加趋势。这说明在长三角一体化大背景下,中心节点的功能性和重要性进一步增强。

此外,我们在评估某一特定节点的依赖性时,常使用中介中心度(betweenness centrality)作为评估标准,这类节点在实际生活中通常被称为“结构洞”。中介中心度的核心计算逻辑在于衡量节点在多大程度上扮演“中介”的角色。当一个节点经常出现在其他节点之间的最短路径上时,它的中介中心度就较高。这种中心度反映了节点在网络中的控制和影响力。中介中心度越高,节点在众多路径中越扮演必经之地,具有更强的传递或控制其他节点向外连接的能力。在长三角城市网络中,上海是最为重要的中介枢纽。南京和杭州之间存在着激烈的次中心竞争;2020 年之前,南京稳压杭州;2020 年之后,杭州的中介作用逼近甚至超过了南京。另外,在苏州和合肥之间的比较中,大部分年份中合肥的中介中心度都高于苏州。这是因为安徽的大部分城市需要依靠合肥作为枢纽来参与整个长三角的网络,而苏州因为与上海紧密相连,再加上南京作为重要的中介枢纽存在,导致苏州的中介能力相对较弱。同时我们还注意到,2020 年后合肥的中介中心度有所下降,这表明长三角一体化战略推动了安徽诸城市与长三角各城市之间的直接联系与互动。合肥中介中心度的下降,间接说明了长三角一体化程度的提高。

相较于前述三种中心度考虑数量和距离的因素,特征向量中心度(eigenvector centrality)更加注重节点连接的质量,而不仅仅考虑数量。特征向量中心度高的节点是那些连接了许多其他重要节点的节点,其重要性是通过相邻节点的重要性来评估的。从特征向量中心度的角度来看,上海依然是最为重要的城市,其次是南京、杭州、苏州和合肥。从数值的演变趋势来看,这五个主要城市的中

心度值大幅增加。这表明在长三角一体化进程中,各节点之间的联系质量进一步加强。

如果说我们通过四大中心度考察了重要中心节点的变化来侧面刻画长三角一体化的进程,那么进一步,我们用凝聚子群来完成“由点到面”的城市网络聚类。我们通过 Ucinet(最大分割深度为 2,集中标准为 0.2)对长三角城市网络进行了计算,最终将其划分为 4 个子群,分别为子群一(上海+江苏全域)、子群二(浙江全域)、子群三(合肥、芜湖、六安、宣城、马鞍山、池州、铜陵、安庆、滁州)、子群四(黄山、阜阳、宿州、蚌埠、淮南、亳州、淮北)。其中,在时空演变上,较大的变化在于 2016—2017 年,上海更多与浙江全域组成子群,但从 2018 年开始归类于江苏全域;相较于浙江、江苏较为稳定的子群结构,安徽全域则是一分为二,且这两个子群结构城市变化较大。<sup>①</sup>

## 五、总结与讨论

大数据研究方法的兴起,为以跨学科方式进行学术研究提供了可能(陈云松、吴青熹、黄超,2015; McFarland, Lewis & Goldberg, 2016; Ruppert, 2013)。这也导致了許多分支方向的出现,无论是计算社会学、计算经济学还是计算传播学,甚至更广义的“计算社会科学”,尽管它们在理论架构、计算方法和研究路径上存在差异,但核心思想都是基于大数据理念,挖掘传统定量研究无法触及的大量非结构化电子印迹数据(迈尔-舍恩伯格、库克耶,2013),以实现对社会现象和规律更精确的描述和解释(周涛、高馨、罗家德,2022)。本文正是基于这一核心思想和方法,收集并处理了长三角地区三省一市 41 座城市 2016—2022 年新闻媒体发布的所有报道。通过一系列数据处理方法,将百万级的非结构化数据转化为可进行有效分析的结构化数据。最终,我们成功构建了拥有 66 万余篇新闻文本的长三角新闻数据库,以测度长三角一体化水平。通过长三角一体化指数,我们呈现了长三角一体化在历时态(时间)和共时态(空间)下的趋势和差异。

---

<sup>①</sup> 期刊版面有限,无法全面呈现图片信息,有意者请向作者索要。

从方法层面来看,本文探索了一种新的途径,利用新闻大数据来测度区域一体化水平,以展示城市之间的实践和互动,并将其指数化、结构化和网络化。这也是计算传播学、经济地理学和社会网络分析的三重交叉尝试。从现实意义上讲,我们从时间和空间两个维度展示了长三角一体化的进程,为推动长三角一体化提供了重要的依据。研究发现,在整体一体化融合的趋势上,2017—2019年长三角地区的一体化水平稳步递增,融合程度持续提高;然而,由于新冠疫情的影响,2020年一体化程度有所减缓;但在2022年,这种下降趋势得到逆转,一体化指数再次上升,区域整体融合再次展开。这一点在后续的社会网络分析中得到了进一步验证,印证了长三角一体化指数构建的科学性。与此同时,本文对长三角41座城市进行了一体化水平的指数化操作。依据各城市的一体化指数,我们将长三角各城市划分为六大层级,可以看到长三角地区的一体化逐渐形成了一个“中心—两翼—多次中心”的层级格局。这样的层级格局体现为两个方面:首先是核心城市如上海、南京、杭州、苏州和合肥,在长三角网络节点中愈发重要的定位,这一点我们通过四大中间度进行了说明;其次在城市网络的内部子群中,核心城市主导了长三角内部子群的划分,由于核心城市的重要性愈发突出,长三角一体化的进程逐渐从“点—点”互动进程转向了“面—面”互动的子群融合。从政策后续来说,习近平总书记在深入推进长三角一体化发展座谈会上指出:“长三角一体化发展有许多深层次问题有待进一步破解,发展质量效率和辐射带动作用仍需提升,重点领域、重点区域一体化尚需努力,产业链供应链分工协作水平有待提升,建立全国统一大市场的龙头带动作用有待进一步发挥,改革开放还需进一步向纵深拓展,超大特大城市治理和发展还有不少短板。”本文认为,如果要推动长三角一体化往深层次发展,必须要利用长三角“中心—两翼—多次中心”的城市层级格局,以龙头带动高质量发展,以两翼衔接重点区域,以多次中心提升一体化水平,完成城市连接实现从“点—点”到“面—面”的过渡。

最后,本文在国内属于首次利用新闻大数据对区域一体化水平进行测度,得到的各城市一体化指数具有很好的后续研究价值。虽然在本文中,我们只对一体化指数进行了描述性的研究,但在接下来的长三角一体化讨论中,我们将利用各城市的一体化指数和各类政治、经济、社会、文化和环境等维度变量构建一个

多年面板数据。通过运用计量经济学模型,我们将深入探讨影响长三角一体化水平的因素,以提供更为实际的研究结论和政策建议。

## 参考文献

- 白俊红、蒋伏心,2015,《协同创新、空间关联与区域创新绩效》,《经济研究》第7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。
- 邓淑芬、朱佳翔、钟昌宝,2018,《长三角城市群物流产业空间关联格局演变——基于聚类、探索性空间数据和 $\sigma$ 收敛性的分析》,《商业经济研究》第20期。
- 方创琳、姚士谋、刘盛和等,2011,《2010中国城市群发展报告》,北京:科学出版社。
- 胡雅龙,2010,《世界第六大城市群:长江三角洲城市群崛起之路》,上海:上海社会科学院出版社。
- 李世奇、朱平芳,2017,《长三角一体化评价的指标探索及其新发现》,《南京社会科学》第7期。
- 李涛、周锐、苏海龙等,2015,《长三角区域经济一体化水平的测度:以关系型大数据为基础》,《新常态:传承与变革——2015中国城市规划年会论文集》。
- 李晓、王小彬,2017,《粤港澳经济一体化:基于边界效应视角的分析》,《湖北社会科学》第11期。
- 刘干、陆叶,2018,《我国区域经济发展水平的聚类分析》,《生产力研究》第1期。
- 刘志彪、孔令池,2019,《长三角区域一体化发展特征、问题及基本策略》,《安徽大学学报》(哲学社会科学版)第3期。
- 迈尔-舍恩伯格、库克耶,2013,《大数据时代:生活、工作与思维的大变革》,盛杨燕、周涛译,杭州:浙江人民出版社。
- 孙红梅、姚书淇,2020,《长三角城市群一体化协同发展的空间实证研究——以财政水平差异分析为支撑》,《城市发展研究》第8期。
- 唐子来、李涛,2014,《京津冀、长三角和珠三角地区的城市体系比较研究——基于企业关联网络的分析方法》,《上海城市规划》第6期。
- 唐子来、赵渺希,2010,《经济全球化视角下长三角区域的城市体系演化:关联网络和价值区段的分析方法》,《城市规划学刊》第1期。
- 王晔倩、林理升,2006,《引力模型与边界效应分析——以长三角和珠三角服务贸易为例》,《上海经济研究》第8期。

- 肖海平、谷人旭,2011,《沪苏浙边界区县域经济合作与协调发展中的边界效应实证研究》,《经济研究导刊》第15期。
- 行伟波、李善同,2009,《本地偏好、边界效应与市场一体化——基于中国地区间增值税流动数据的实证研究》,《经济学(季刊)》第4期。
- 熊丽芳、甄峰、王波等,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。
- 徐宜青、曾刚、王秋玉,2018,《长三角城市群协同创新网络格局发展演变及优化策略》,《经济地理》第11期。
- 许静涵、林礼娜、武义天等,2021,《基于引力模型和聚类分析的城市群经济结构研究——以长三角23市为例》,《浙江金融》第2期。
- 张贵、温科,2017,《协同创新、区域一体化与创新绩效——对中国三大区域数据的比较研究》,《科技进步与对策》第5期。
- 赵渺希、唐子来,2010,《基于网络关联的长三角区域腹地划分》,《经济地理》第3期。
- 赵永亮、才国伟,2009,《市场潜力的边界效应与内外部市场一体化》,《经济研究》第7期。
- 赵永亮、徐勇、苏桂富,2008,《区际壁垒与贸易的边界效应》,《世界经济》第2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。
- 周涛、高馨、罗家德,2022,《社会计算驱动的社会科学研究方法》,《社会学研究》第5期。
- 朱查松、王德、罗震东,2014,《中心性与控制力:长三角城市网络结构的组织特征及演化——企业联系的视角》,《城市规划学刊》第4期。
- Afesorgbor, S. 2017, "Revisiting the Effect of Regional Integration on African Trade: Evidence from Meta-Analysis and Gravity Model." *The Journal of International Trade & Economic Development* 26(2).
- Bairoch, P. 1991, *Cities and Economic Development: From the Dawn of History to the Present*, Chicago: University of Chicago Press.
- Beaverstock, J., R. Smith & P. Taylor 2000, "World-City Network: A New Metageography?" *Annals of the Association of American Geographers* 90(1).
- Bertinelli, L. & D. Black 2004, "Urbanization and Growth." *Journal of Urban Economics* 56 (1).



- Broker, J. 1984, *How Do International Trade Barriers Affect Interregional Trade? Regional and Industrial Theories*, Amsterdam: North-Holland Press.
- Castells, M. 2000, *The Rise of the Network Society*, Oxford: Wiley-Blackwell Press.
- Chen, Y. & L. Huang 2018, "A Scaling Approach to Evaluating the Distance Exponent of the Urban Gravity Model." *Chaos, Solitons & Fractals* 109.
- Chen, Y. 2009, "Urban Gravity Model Based on Cross-Correlation Function and Fourier Analyses of Spatio-Temporal Process." *Chaos, Solitons & Fractals* 41(2).
- Derudder, B. & F. Witlox 2005, "An Appraisal of the Use of Airline Data in Assessing the World City Network: A Research Note on Data." *Urban Studies* 42(13).
- Engel, C. & J. Rogers 1994, *How Wide Is the Border?* Cambridge: National Bureau of Economic Research Cambridge.
- Fang, C. & D. Yu 2017, "Urban Agglomeration: An Evolving Concept of An Emerging Phenomenon." *Landscape and Urban Planning* 162.
- Fang, C. & D. Yu 2020, *China's Urban Agglomerations*, Singapore: Springer Singapore.
- Fang, C. , Q. Sheng & S. Li 2018, "Economic Links Between the Cities in North Anhui Based on Gravity Model." *Areal Research & Development* 37(2).
- Gottmann, J. 1976, "Megalopolitan Systems Around the World." *Ekistics* 41(243).
- Graham, S. & S. Marvin 1995, *Telecommunications and the City: Electronic Spaces, Urban Places*, London: Routledge.
- Hall, P. & K. Pain 2006, *The Polycentric Metropolis: Learning from Mega-City Regions in Europe*, London: Routledge.
- Han, R. , H. Cao & Z. Liu 2018, "Studying the Urban Hierarchical Pattern and Spatial Structure of China Using A Synthesized Gravity Model." *Science China Earth Sciences* 61(12).
- Ibrahim, S. 2008, "A Study of Optimum Currency Area in East Asia: A Cluster Analysis." *Journal of Economic Integration* 23(4).
- Kim, J. 2001, "Economic Integration of Major Industrialized Areas: An Empirical Tracking of the Continued Trend." *Technological Forecasting and Social Change* 67(2).
- König, J. & R. Ohr 2013, "Different Efforts in European Economic Integration: Implications of the EU Index." *Journal of Common Market Studies* 51(6).
- Lee, H. , R. Owen & D. Mensbrugghe 2009, "Regional Integration in Asia and Its Effects on the EU and North America." *Journal of Asian Economics* 20(3).

- Mahutga, M. , X. Ma & D. Smith et al. 2010, “Economic Globalisation and the Structure of the World City System: The Case of Airline Passenger Data.” *Urban Studies* 47(9).
- Mayhew, B. & R. Levinger 1976, “Size and the Density of Interaction in Human Aggregates.” *American Journal of Sociology* 82(1).
- McCallum, J. 1995, “National Borders Matter: Canada-US Regional Trade Patterns.” *The American Economic Review* 85(3).
- McFarland, D. , K. Lewis & A. Goldberg 2016, “Sociology in the Era of Big Data: The Ascent of Forensic Social Science.” *The American Sociologist* 47(1).
- Poncet, S. 2003, “Measuring Chinese Domestic and International Integration.” *China Economic Review* 14(1).
- Rossi, E. , J. Beavertstock & P. Taylor 2007, “Transaction Links Through Cities: ‘Decision Cities’ and ‘Service Cities’ in Outsourcing by Leading Brazilian Firms.” *Geoforum* 38(4).
- Ruppert, E. 2013, “Rethinking Empirical Social Sciences.” *Dialogues in Human Geography* 3(3).
- Rutherford, J. , A. Gillespie & R. Richardson 2004, “The Territoriality of Pan-European Telecommunications Backbone Networks.” *Journal of Urban Technology* 11(3).
- Smith, D. & M. Timberlake 2002, *Hierarchies of Dominance among World Cities: A Network Approach*, [https://xueshu.baidu.com/usercenter/paper/show?paperid=941323ecde756d3bd80bad14a5dbeb9f&site=xueshu\\_se](https://xueshu.baidu.com/usercenter/paper/show?paperid=941323ecde756d3bd80bad14a5dbeb9f&site=xueshu_se).
- Storper, M. 2013, *Keys to the City: How Economics, Institutions, Social Interactions, and Politics Shape Development*, Princeton: Princeton University Press.
- Taylor, P. 1999, “So-Called ‘World Cities’: The Evidential Structure Within A Literature.” *Environment and Planning A* 31(11).
- Van Nuffel, N. , B. Derudder & F. Witlox 2010, “Even Important Connections Are Not Always Meaningful: On the Use of A Polarisation Measure in A Typology of European Cities in Air Transport Networks.” *Tijdschrift voor economische en sociale geografie* 101(3).
- Wenjie, L. 2022, “Evaluation of Urban-Rural Development Integration Level Based on Cluster Analysis.” 2022 14th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA).
- Wilson, J. & T. Otsuki 2007, “Regional Integration in South Asia: What Role for Trade Facilitation?” World Bank Policy Research Working Paper.

Zhao, Y. , G. Zhang & H. Zhao 2021, “Spatial Network Structures of Urban Agglomeration Based on the Improved Gravity Model: A Case Study in China’s Two Urban Agglomerations.” *Complexity*.