



## 基于手机大数据的中国人口迁徙模式及疫情影响研究

戴碧涛 谭索怡 陈洒然 蔡梦思 秦烁 吕欣

## Measuring the impact of COVID-19 on China's population migration with mobile phone data

Dai Bi-Tao Tan Suo-Yi Chen Sa-Ran Cai Meng-Si Qin Shuo Lu Xin

引用信息 Citation: *Acta Physica Sinica*, 70, 068903 (2021) DOI: 10.7498/aps.70.20202084

在线阅读 View online: <https://doi.org/10.7498/aps.70.20202084>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

### 您可能感兴趣的其他文章

#### Articles you may be interested in

评估新型冠状病毒地区防控效果的一种近似方法

Approximate method to evaluate the regional control efficacy of COVID-19

物理学报. 2020, 69(10): 100201 <https://doi.org/10.7498/aps.69.20200441>

大气压氦气介质阻挡放电单-多柱演化动力学

Evolutionary dynamics of single-multiple columns in atmospheric helium dielectric barrier discharge

物理学报. 2020, 69(14): 145203 <https://doi.org/10.7498/aps.69.20200473>

新型冠状病毒肺炎早期时空传播特征分析

Analysis on early spatiotemporal transmission characteristics of COVID-19

物理学报. 2020, 69(8): 080701 <https://doi.org/10.7498/aps.69.20200285>

基于连续感染模型的新冠肺炎校园传播与防控策略分析

Analysis of COVID-19 spreading and prevention strategy in schools based on continuous infection model

物理学报. 2020, 69(24): 240201 <https://doi.org/10.7498/aps.69.20201106>

基于Hadoop大数据平台和无简并高维离散超混沌系统的加密算法

Encryption algorithm based on Hadoop and non-degenerate high-dimensional discrete hyperchaotic system

物理学报. 2017, 66(23): 230503 <https://doi.org/10.7498/aps.66.230503>

基于上证指数高频数据的中国资本市场微观特性研究

Microscopic characteristics of Chinese capital market based on the high frequency data of Shanghai composite index

物理学报. 2017, 66(12): 120203 <https://doi.org/10.7498/aps.66.120203>

# 基于手机大数据的中国人口迁徙模式及疫情影响研究\*

戴碧涛<sup>1)</sup> 谭索怡<sup>1)</sup> 陈洒然<sup>2)</sup> 蔡梦思<sup>1)</sup> 秦烁<sup>2)</sup> 吕欣<sup>1)†</sup>

1) (国防科技大学系统工程学院, 长沙 410073)

2) (盲信号处理国家重点实验室, 成都 610041)

(2020年12月8日收到; 2021年2月2日收到修改稿)

新型冠状病毒感染的肺炎 (COVID-19) 可通过人员接触与流动迅速传播, 因此研究人类迁徙和出行模式的变化对疫情防控至关重要. 本文基于手机运营商 2020 年春运及疫情暴发前后连续两个月的全国地级市之间的人口流动数据, 运用时序网络分析方法构建人口流动网络拓扑结构指标, 并通过引入地理衰减因子提出 Spatial-Louvain 社团检测算法, 研究平时、春运、疫情防控隔离和生产复工四阶段的人口迁徙模式的时空演化规律. 研究发现: 受各地疫情防控措施影响, 武汉封城后全国城市间人口流量急剧下降, 并持续至 2 月中旬. 疫情期间人口流动网络结构呈现四阶段的时空演化模式; 本文提出的空间网络社团检测算法比传统 Louvain 算法平均模块度值提高了 14%; 中国城市分布以经济交互和地理位置为基础, 形成了以核心城市为中心, 向周边辐射的城市群格局; 疫情因素仅能在短暂时间内改变部分城市的城市群归属, 当该因素消失或减弱后, 城市群能迅速恢复原有格局.

**关键词:** COVID-19, 移动大数据, 人口流动, 时空演化**PACS:** 89.65.Cd, 89.65.-s, 89.75.Kd, 64.60.aq**DOI:** 10.7498/aps.70.20202084

## 1 引言

人口迁徙是疫情传播的重要媒介, 突发、大规模和分散的人类迁徙可将疾病的局部暴发转变为广泛流行<sup>[1-3]</sup>. 一方面, 频繁的人口流动为远距离、大规模传播病毒感染创造了条件, 另一方面, 人群在不同地区间的往返活动, 对潜在感染者与易感人群的充分接触传播起到极大的促进作用. 新型冠状病毒感染的肺炎 (COVID-19) 可在人与人之间持续传播, 在 2020 年初新冠肺炎疫情的暴发期间, 随着春节及返程高峰的到来, 各地区间大规模的人口流动增加了交叉接触传播的风险, 为疫情防控带

来了极大的挑战. 截至 2020 年 12 月 6 日 24 时, 我国累计报告确诊病例达 86634 例, 累计死亡病例达 4634 例. 分析人口迁徙模式, 挖掘其背后的时空演化规律有助于分析疫情传播路径和控制疫情传播范围, 对疫情防控至关重要.

针对新冠肺炎的相关研究和防控工作已经迅速展开, 现有研究主要侧重于流行病学的传播特征<sup>[4-6]</sup>、基本再生数<sup>[7-9]</sup>等主要参数的估计以及基于传播动力学进行感染人数预测<sup>[10,11]</sup>. 但传统的动力学模型并不能模拟疫情在地域间的传播, 此外, 动力学模型的预测结果通常显著依赖于初始系数的取值, 初始参数的设置差异可能导致研究结果大相径庭<sup>[12]</sup>. 因此考虑人口迁徙的疫情传播模型得

\* 国家杰出青年科学基金 (批准号: 72025405)、国家自然科学基金 (批准号: 82041020, 71901067, 72001211)、国家自然科学基金重大项目 (批准号: 91846301)、四川省科学技术厅新冠肺炎应急专项 (批准号: 2020YFS0007)、湖南省自然科学基金 (批准号: 2020JJ5679) 和湖南省科学技术厅重点研发计划 (批准号: 2019GK2131) 资助的课题.

† 通信作者. E-mail: [xin\\_lyu@sina.com](mailto:xin_lyu@sina.com)

到了广泛关注. Brockmann 和 Helbing<sup>[13]</sup> 指出疾病的传播与城市间的“有效距离”有关, 其中有效距离的核心构成因子是城市间的交通流量. Wu 等<sup>[14]</sup> 通过使用确诊病例数据、全球飞往武汉的航空网络数据以及腾讯数据库定位服务记录获取的人口流动数据, 首次对本次疫情的时空传播特征进行建模. 而王聪等<sup>[15]</sup> 结合了百度迁徙数据和公开新闻报道, 刻画了武汉市 COVID-19 早期时空传播特征. Gross 等<sup>[16]</sup> 研究了 COVID-19 在中国的时空传播模式, 研究发现各省感染人数与相应省份人口数量、距湖北省的距离均符合幂律分布, 疾病的传播与湖北人口的迁徙高度相关, 得益于有效的隔离措施, 感染率在后期呈指数级下降. 许小可等<sup>[17]</sup> 对疫情暴发前期武汉外流人口去向的地理分布进行了统计分析. Tian 等<sup>[18]</sup> 基于人口流动大数据, 研究了武汉封城对疫情阻断的效果, 结果表明武汉封城极大地减少了全国各省的输入和感染人数, 且在封城之前, 受感染的个体数量与湖北的人口迁移高度相关. Jia 等<sup>[2]</sup> 使用海量手机定位数据, 基于在疫情暴发前武汉输入到全国各地的人口流动数据, 构建“人口流动-风险源模型”准确地预测新冠疫情暴发的时间、强度和地理分布.

综合来看, 大规模人口流动是疫情得以迅速扩散的重要媒介, 城市间密集的人口迁徙对疫区疫情向外传播具有重要推动作用. 基于大数据挖掘人口出行模式可以预测传染病如何在地区间扩散, 并在灾难性流行病暴发或再次暴发之前利用数据分析技术对其进行防控.

然而, 已有研究使用的人口流动数据大多基于百度迁徙数据<sup>[19]</sup> 和中国疾病预防控制中心公布的病例数据, 而百度迁徙指数构造方法目前并未公开, 其与真实人口迁徙的关系并不完全明确<sup>[15]</sup>. 同时, 确诊病例等数据存在滞后性和统计误差. 这些因素会导致人口流动数据在时效性和准确性上均有所欠缺. 大量研究表明, 移动大数据在刻画疫情期间传播风险方面具有突出能力, 利用移动手机数据能从时空两方面精确提取出突发事件情况下的人口流动数据<sup>[20,21]</sup>, 进而有效地预测大规模灾难之后的人口迁移轨迹<sup>[22]</sup>. 根据国家工信部的数据, 2018 年我国的手机普及率高达 112.2 部/百人<sup>[23]</sup>, 2019 年移动电话用户总数已超 16 亿户<sup>[24]</sup>. 因此手机通信数据凭借覆盖广、定位准确的特性可以在出现突发事件情况下对大规模人群进行定位和追踪,

进而及时、有效地提取出人口的分布和流动信息, 进一步为应急管理决策提供有效支撑.

本文基于疫情期间 2020 年 1 月 1 日—2 月 29 日的匿名手机定位大数据, 构建全国范围的人口流动网络, 从时空大数据分析的角度挖掘不同阶段全国人口流动的演化规律, 为疫情发展态势预测、防控措施评价等提供科学决策支撑. 本文的主要贡献在以下几个方面: 1) 基于手机定位大数据研究人口迁徙, 发现我国城市分布随人口流动而呈现出区域集聚的城市群格局; 2) 为了探究人口流动网络背后蕴含的空间作用机理以及城市间人口流动的演化规律, 提出了一种引入地理衰减因子的 Spatial-Louvain (SL) 算法, 解决了现有社团检测算法不能很好地集成空间信息的问题; 3) 本文的研究框架有助于进一步理解突发事件对大规模人口迁徙的影响、人口流动性与区域经济发展的关系, 同时为疫情防控效果评价, 尤其是出行限制措施的实施效果评估提供了科学、定量的分析范式.

## 2 数据与方法

### 2.1 手机定位数据

本文使用的人口流动数据从国内某手机运营商提供的经过匿名处理的手机定位数据中提取, 数据统计了 2020 年 1 月 1 日—2 月 29 日中国大陆 366 个地级市 (不包含港澳台) 之间的人口流动情况. 原始数据包括日期、出发省份、出发城市、抵达省份、抵达城市、运营商统计的人口移动量、由运营商使用机器学习方法估计的总人口移动量, 共 7 个字段, 总计 8350643 条数据, 累计记录了 2195986035 人次的流动数据 (数据记录了不同城市间的人口流量, 不包含用户 id). 本文使用由运营商基于机器学习方法估计的总人口移动量来研究人口迁徙. 尽管基于机器学习的外推技术已经相当成熟, 为了确保数据的准确性, 本文在中国几个地级市以及中国所有省份, 通过国内其他两家手机运营商的实际覆盖率验证了这种计数推断, 证明了使用该数据来预测整个网络中的用户数量具有很高的准确性.

### 2.2 数据处理与网络构建

本文基于时序网络分析方法将数据集按照日期划分为 60 个时间窗口, 每个时间窗口均为 1 天, 将同一天内起始城市和结束城市分别相同的人口

表 1 网络中心性指标  
Table 1. Network centrality metrics.

中心性指标	公式	描述
度	$k_i = \sum_{j=1}^N a_{ij}$	$N$ 为网络节点总数, $a_{ij}$ 表示节点 <i>i</i> 与节点 <i>j</i> 之间的连接, 如果连接存在, 则 $a_{ij} = 1$ ; 否则 $a_{ij} = 0$ .
加权重	$k_i^w = \sum_{j=1}^N w_{ij}$	$w_{ij}$ 表示节点 <i>i</i> 与节点 <i>j</i> 之间的连接权重, 在本文中指两个城市间的人口流动数量.
密度	$\rho = \frac{M}{N(N-1)}$	$M$ 为网络中实际存在的边数, $N$ 为节点总数. 该指标用来衡量网络疏密.
集聚系数	$C_i = \frac{T(i)}{k_i(k_i-1) - 2k_i^{-1}}$	其中 $T(i)$ 是经过节点 <i>i</i> 的有向三角形的数量, $k_i$ 是节点 <i>i</i> 的入度和出度之和, $k_i^{-1}$ 是 $k_i$ 的倒数.
介数	$B_i^w = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}$	$\sigma_{st}$ 为从节点 <i>s</i> 到节点 <i>t</i> 的最短路径数量, $\sigma_{st}(i)$ 为从节点 <i>s</i> 到节点 <i>t</i> 并且经过节点 <i>i</i> 的最短路径数量.
接近中心性	$C_i = \frac{N}{\sum_{j=1}^N \text{dis}_{ij}}$	$N$ 为节点总数, $\text{dis}_{ij}$ 为节点 <i>i</i> 到节点 <i>j</i> 的距离.

移动量合并, 构建成一个  $366 \times 366$  的城市间人口迁徙 (OD) 矩阵. 然后将每个时间窗口均构建成为有向加权网络<sup>[25]</sup>. 网络中的每个节点代表在数据集中出现的城市, 共计 366 个, 每个城市由对应的编号标识. 如果用户在目的地城市中停留半个小时以上, 则将其计为一次出行. 矩阵元素  $w_{i,j}$  表示由城市  $i$  到城市  $j$  的人口流动量.

### 2.3 网络时序拓扑结构分析

本文采用 6 个中心性指标 (见表 1) 对构建的人口流动网络进行结构性分析, 主要包括度 (用来衡量城市间的人口交互流动的频度以及节点城市的辐射力和吸引力)、加权重 (用来衡量城市间的人口流动总量)、密度 (用来衡量网络中节点联系的紧密程度)、集聚系数 (测量人口流动的集聚性)、介数与接近中心性 (衡量节点城市在网络中的枢纽连接作用以及网络整体连通度).

### 2.4 社团检测

人口流动往往与城市的地理位置、经济体量息息相关, 为了深入挖掘人口流动网络中蕴含的组织结构信息和城市分布特征, 本文使用社团检测技术对联系紧密的城市群进行分类和识别. 目前, 复杂网络的社团检测算法主要有层次聚类算法、基于模块度优化方法、图分割等算法<sup>[26]</sup>. 然而针对时空网络, 现有算法大多仅考虑网络的拓扑结构, 不能很好地将空间信息集成到模型中. 为解决这一问题, 本文基于重力模型的地理衰减思想<sup>[27]</sup>, 对 Louvain

算法<sup>[28]</sup>进行改进.

Louvain 算法中模块度<sup>[29]</sup>定义公式为

$$Q = \frac{1}{2m} \sum_{i,j} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \delta(C_i, C_j), \quad (1)$$

基于重力模型改进的 Louvain 算法中模块度的计算公式为

$$Q = \frac{1}{2m} \sum_{i,j} \left( \frac{A_{ij}}{d_{ij}^\beta} - \frac{k_i k_j}{2m} \right) \delta(C_i, C_j), \quad (2)$$

其中  $m$  代表网络的边数,  $A_{ij}$  为节点  $i$  与节点  $j$  之间连边的权重,  $k_i = \sum_j A_{ij}$  为节点  $i$  的所有连边的权重之和, 同理  $k_j = \sum_i A_{ji}$ .  $d_{ij}$  为节点  $i$  与节点  $j$  之间的地理距离,  $\beta$  为衰减系数,  $C_i$  与  $C_j$  分别表示节点所属的社团, 如果两个节点属于同一社团,  $\delta$  取值为 1, 否则取值为 0. 本文选取  $A_{ij}/d_{ij}^\beta$  作为网络边权, 修正 Louvain 算法中的权重参数, 引入重力模型一方面可以避免  $d_{ij}$  直接与  $A_{ij}$  线性组合和消除量纲的影响, 另一方面引入地理距离强化了短程边效应, 弱化了新疆、拉萨等边缘地区的长程且人口流量少的边, 防止边缘地带的城市社团在各阶段反复变动. 本文将衰减因子  $\beta$  设置为 1.

## 3 结果分析

### 3.1 人口流量时序特征

本文基于 60 天的城市人口流动数据, 通过划分时间窗口的方法分析网络元素 (节点和边)、中心性指标的变化, 进而分析网络的演化过程, 结果如

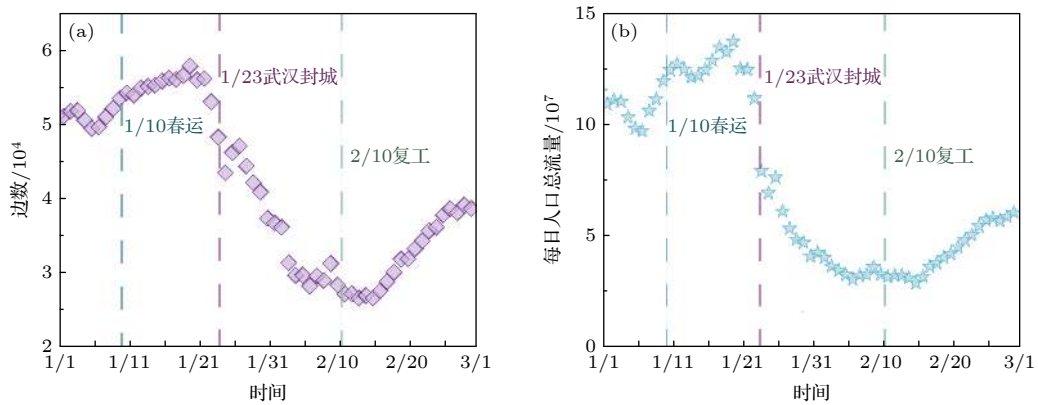


图 1 (a) 人口流动网络边数时序特征; (b) 人口总流量时序特征

Fig. 1. (a) Temporal characteristics of the number of edges in the population mobility network; (b) temporal characteristics of total population flow.

图 1 所示. 可以看出, 2020 年 1 月—2 月间城市间的人口流动网络边数变化和人口总流量演化趋势一致, 呈现明显的阶段性特征. 自 1 月 10 日春运以来, 人口流动规模迅速增加, 城市间人口转移的路径逐步多元化. 1 月 20 日有关疫情的媒体报道引发社会热议, 随后人口流动规模开始下降. 1 月 23 日 10 时武汉正式封城, 各地群众积极支持防疫, 居家隔离, 自 24 日起人口流动规模呈现断崖式下跌, 随后以高速率持续降低, 2 月 15 日达到谷值. 疫情期间人口流动网络中单天最大边数为 57888, 最小边数为 26830, 降低 53.7%, 单天最大人口流量为 137494000, 最小人口流量为 29157400, 降低 78.8%, 这些数据表明, 武汉封城及各地疫情防控策略对控制人口流动和人员接触起到了显著作用.

### 3.2 网络拓扑时序特征

城市间的人口流动网络拓扑结构变化也呈现明显的阶段性特征: 1 月 1 日—1 月 9 日, 图 2 中的平均度、加权度、密度、集聚系数、接近中心性相对平稳, 1 月 10 日—1 月 20 日对应指标稳步上升, 20 日达到最大值, 23 日有大幅度下降, 然后在 24 日骤降, 呈现断崖式下跌. 这是因为 1 月 10 日为春运开始日, 1 月 23 日为武汉封城日, 春运期间, 城市间人口流动变得频繁且规模庞大, 网络连通性增大, 平均连接数提高了 16.7%, 密度提高了 10.3%, 集聚系数提高了 9.2%. 网络中节点城市之间的连接有多条路径, 介数降低了 7.1%. 由于 1 月 23 日武汉封锁交通, 政府进行干预, 各省市陆续启动一级响应, 城市间人口流动大大减少 (高达 80.4%), 网络连通性降低, 平均连接数降低了 54.9%,

密度降低了 53.5%, 集聚系数降低了 29.8%. 同时 2 月 10 日开始, 各地陆续开始复工, 人口流动开始增加 (提高了 110%), 逐步朝常态化阶段恢复, 平均连接数增加了 54.9%, 密度提高了 50.0%, 集聚系数提高了 26.1%, 人口流动提高了 110.0%, 介数降低了 12.5%.

为了更好地反映新冠肺炎疫情的发展动态和相应防疫措施效果, 结合网络拓扑结构的阶段性特征, 本文划分了 4 个人口迁徙阶段. 第一阶段是 2020 年 1 月 1 日—1 月 9 日, 没有采取具体针对新冠肺炎干预措施的常态化阶段. 第二阶段是 2020 年 1 月 10 日—1 月 22 日, 发生大规模人口流动, 加速新冠肺炎蔓延的春运阶段. 第三阶段是 2020 年 1 月 23 日—2 月 10 日, 武汉封城到各地复工, 恢复经济生产之前的居家隔离阶段. 第四阶段是 2020 年 2 月 11 日—2 月 29 日, 社会生产恢复, 各地开始复工的返流阶段. 将 60 天的人口流动网络合成 4 个静态网络, 相关网络基础特征如表 2 所列. 常态化阶段网络连接紧密, 人口流动频繁, 到了春运阶段进一步增强. 然而居家隔离阶段网络边数、平均度和密度均稍高于常态化阶段, 这是因为 1 月 24 日和 25 日是春节时期, 仍然有部分人口在返乡途中, 本文为了突出武汉封城这个关键事件 (1 月 24 日全国人口总流量呈现断崖式下跌), 将 1 月 23 日作为春运与隔离阶段的分割点. 同时在隔离阶段, 各地政府响应国家卫健委紧急号召令, 调动救援队驰援武汉, 也增加了城市间的连通路径. 而平均日流量出现了大幅度下降则说明了行人出行模式大多转变为居家隔离, 人口迁徙得到了有效控制.

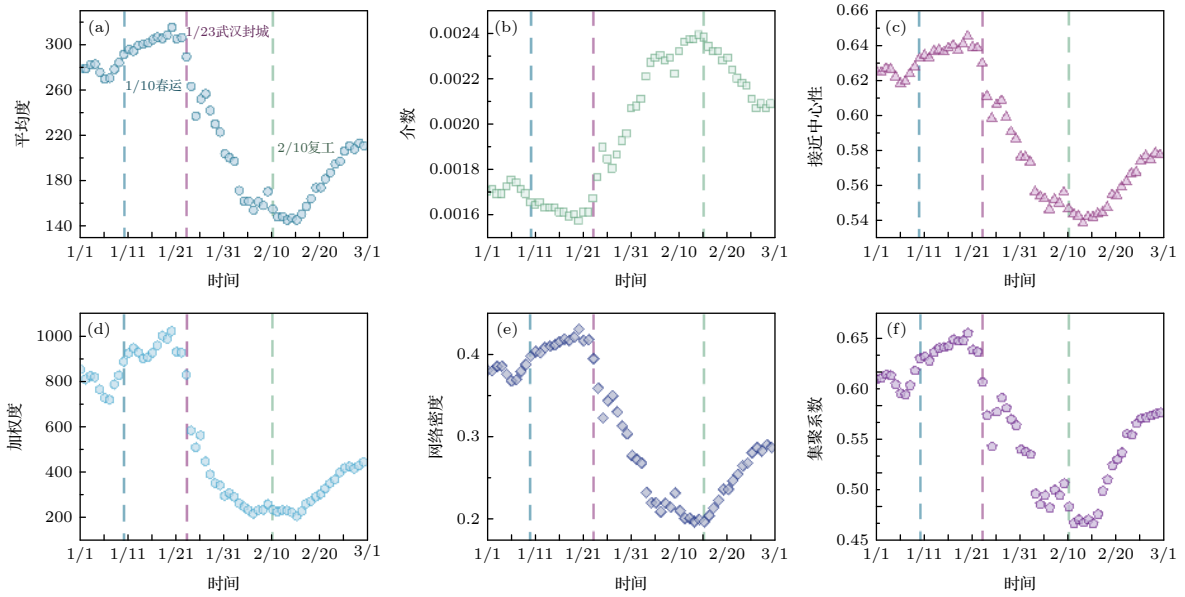


图 2 人口流动网络中心性指标时序特征 (a) 平均度; (b) 介数; (c) 接近中心性; (d) 加权度; (e) 网络密度; (f) 集聚系数  
 Fig. 2. Temporal characteristics of population mobility network centrality metrics: (a) Average degree; (b) betweenness centrality; (c) closeness centrality; (d) weighted degree; (e) density; (f) clustering coefficient.

表 2 人口流动网络四阶段网络基础特征  
 Table 2. Basic characteristics of the population mobility network in four stages.

网络密度	时间	边数	平均日流量	平均度	图密度	集聚系数
常态化	1.1—1.9	83918	107060322	255.016	0.625	0.779
春运	1.10—1.23	93591	125834714	255.016	0.697	0.818
居家隔离	1.24—2.10	84969	46081244	231.523	0.633	0.775
复工	2.11—2.29	77009	44067415	209.834	0.573	0.738

### 3.3 社团检测算法对比

为了体现 SL 算法的优越性, 首先比较了两种算法四阶段网络的模块度, 结果如图 3(c) 所示, SL 算法在各阶段均显著高于传统的 Louvain 算法, 平均模块度值提高了 14%. 然后进一步比较了两种算法在常态化阶段的社团检测结果. 如图 3(a) 和图 3(b) 中区域 1 所示, 传统的 Louvain 算法仅考虑城市间即边上的人口流量, 忽略了现实的空间信息和距离要素, 从而将一些偏远地区的城市也与中心城市合并为一个社团, 如乌鲁木齐市和北屯市. 另一方面, SL 算法弱化了长程及人口流量少的边, 强化了短程边效应, 使社团划分更加合理. 区域 2—5 则说明即使地理邻近的城市也可能因为其他因素不属于同一个社团, SL 挖掘了人口流动网络潜藏分布规律, 具体内容将在社团演化部分进行深入讨论.

### 3.4 人口流动网络社团演化分析

人口流动可以作为衡量城市之间经济联系的

一种指标, 本文使用 SL 算法在四个阶段中分别对人口流动网络进行社团划分 (图 4), 为了体现经济因素, 图例仅展现省会城市、一线、新一线和二线城市<sup>[30]</sup>. 本文算法所检测的城市社团的边界与城市经济区的边界具有高度相似性, 且四个阶段的社团数目均为 18. 社团分布呈现以省会城市、一线、新一线和二线城市为核心, 以点带面, 向四周辐射的区域集聚城市群分布特征 (由人口流动的差异性造成, 不同于传统的城市群定义). 图 5 则细致地展示了各阶段社团之间的成员变化及转移路径, 图中“1.”—“4.”分别对应表 2 中的四个阶段.

#### 3.4.1 常态化阶段分布情况

在常态化人口流动阶段, 由于地理上的邻近性和城市之间的经济交互联系, 除了 1—2 个地处新疆西藏的边远社团, 因为地理位置过于偏僻, 尚未能与周边城市加强联系, 其余绝大多数社团都体现了中心城市和省会城市耦合的集群现象. 以重庆、成都、贵阳和昆明为纽带, 构成了成渝社团, 以沈阳、

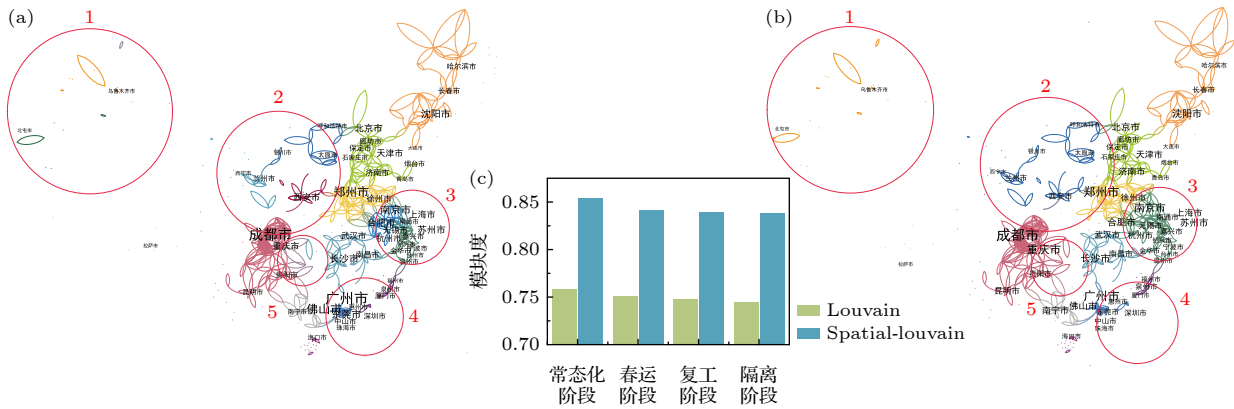


图3 SL算法与Louvain算法的人口流动网络社团检测结果对比 (a) SL算法社团检测结果; (b) Louvain算法社团检测结果; (c) 模块度值

Fig. 3. Comparison between SL algorithm and Louvain algorithm: (a) Results of SL; (b) results of Louvain; (c) modularity value.

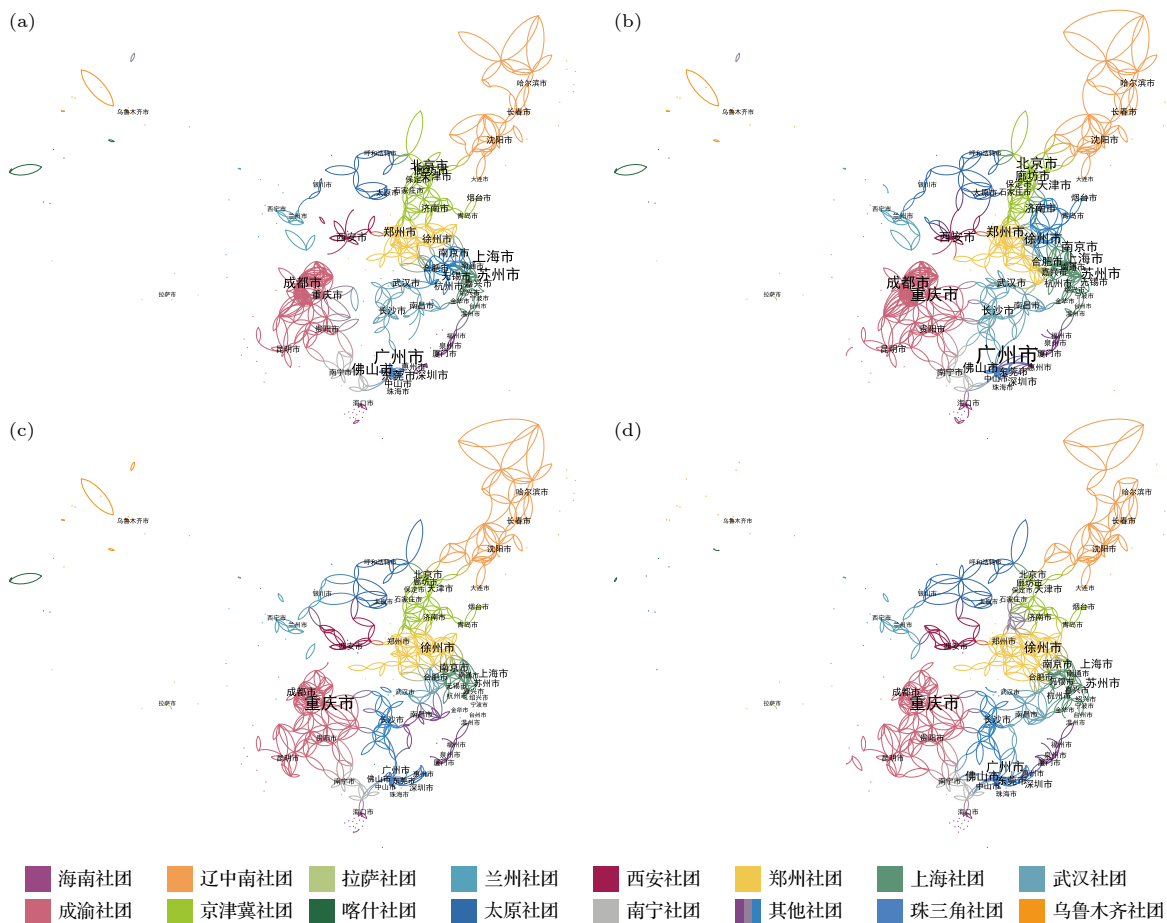


图4 基于人口流动网络的四阶段城市集群演化 (a) 常态化阶段; (b) 春运阶段; (c) 隔离阶段; (d) 复工阶段

Fig. 4. Four-stage evolution of urban agglomeration based on population mobility network: (a) Normal times; (b) chunyun migration; (c) epidemic interventions; (d) recovery times.

长春、哈尔滨、大连为核心的辽中南社团，以北京、天津、济南、石家庄为核心的京津冀社团，以广州、深圳、佛山、珠海为核心的珠三角社团，以及以武汉、长沙为核心的中部社团，具体社团构成信息详见图5和补充材料 (online)。

### 3.4.2 常态化阶段到春运阶段社团演化

在春运时期绝大部分城市群与常态化阶段保持一致，仅有47个城市发生了社团转移。以郑州和徐州为核心的郑州社团发生了分离，徐州等城市形成了新的徐州社团。本文构建的是有向网络，且

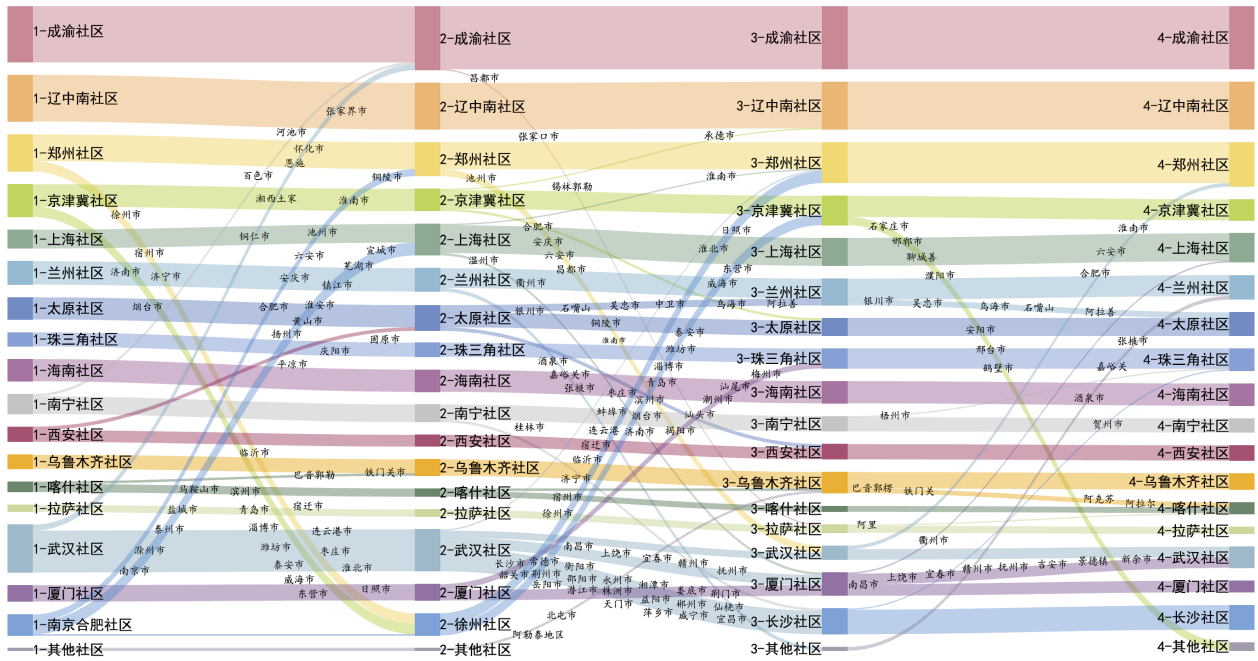


图5 基于人口流动网络的四阶段城市群演化桑基图

Fig. 5. Sankey diagram of four-stage evolution of urban agglomeration based on population mobility network.

两阶段对应城市地理距离一致, 因此人口流量变化是影响社团成员转移的主导因素. 这说明该社团的人口常年在郑州社团工作, 由于春运返乡, 大规模人口回流, 郑州社团人口流动发生较大的变化, 导致社团分离.

京津冀社团里以济南、烟台和青岛为核心的山东半岛社团也加入了徐州社团, 作用机理与徐州社团相似. 同时山东半岛社团和徐州社团的合并也说明山东和江苏一直存在经济和人口交互, 但是由于彼此在地理上分别邻近京津冀社团和郑州社团, 这种效应只有在春运时期才有所体现.

广西社团的百色和河池、武汉长沙社团的铜仁、湘西土家族苗族自治州、恩施土家族和苗族自治州加入了成渝社团; 南京合肥社团分离, 合肥及其周边城市加入郑州社团, 南京及其周边城市加入上海社团. 这说明常态化阶段合肥南京社团人口流动、经济交互以及城市耦合的程度超越其与郑州社团和上海社团的交互程度.

总体来看常态化阶段到春运阶段的演化能够体现社团间邻近城市的人口和经济效应. 社团内部城市保持整体稳定, 邻近社团的部分接壤城市会发生社团转移, 且大多是由春运返乡, 大规模人口迁徙导致的.

### 3.4.3 春运阶段到隔离阶段社团演化

自武汉封城开始, 各地纷纷启动公共卫生一级

响应, 号召公民居家隔离, 因此隔离阶段人口流动大幅度下降. 然而大部分社团仍保持稳定, 社团内成员变动不大, 共有 83 个城市改变了其社团归属, 且大多为城市等级较低的城市.

其中合肥等城市社团变动最大, 脱离郑州社团加入了武汉社团. 这是因为在居家隔离阶段, 合肥等城市人口流动规模大幅度减少, 行人出行方式转变为邻近地区的小规模流动, 因此加入了武汉社团.

银川等城市脱离太原社团加入兰州社团, 而常态化、春运阶段银川等城市一直在太原社团, 并不在地理位置与其更接近的兰州社团, 说明其与太原社团城市连接紧密, 经济因素占据社团构成主导地位. 这表明, 极端突发情况下, 在经济欠发达地区, 地理的邻近性超越经济的交互, 地理位置成了影响人口流动的主导因素. 此前, 在常态化阶段和春运阶段, 长沙武汉社团一直很稳定, 天然的地理邻近、直通高铁以及源远流长的历史联系让武汉和长沙一直联系十分紧密. 然而, 在隔离阶段, 长沙等城市脱离了武汉社团, 构成了新的长沙社团. 这是因为, 武汉是重灾区、中国疫情的暴发地, 加之武汉封城, 行政上的封锁和灾害性疾病导致了社团分离.

南昌等城市脱离武汉社团加入厦门社团的原因则与银川加入兰州社团 (地理邻近性) 的原因一致, 长沙离开武汉社团则是由于行政封锁与重大防疫

所导致. 而在春运阶段形成的济南徐州社团在此时瓦解, 各自分别回到常态化阶段的京津冀社团和郑州社团 (一方面春运阶段大部分人口已经从务工城市返回家乡, 理论上社团应当与春运阶段保持一致, 另一方面国家建议公民居家隔离, 人口流动已经大大减少), 这一现象说明即使由于疫情的影响, 也无法完全隔绝城市群间的交互.

#### 3.4.4 隔离阶段到复工阶段社团演化

与上一阶段相比, 复工阶段大部分社团仍然保持稳定, 共有 35 个城市发生社团转移. 银川等城市离开了上一阶段加入的太原社团回归了兰州社团; 合肥等离开武汉社团回归郑州社团; 南昌等城市离开了厦门社团回归武汉社团; 其他城市陆续回归到常态化阶段的状态. 这说明中国城市社团以城市群为核心且具有鲁棒性和恢复性, 在干扰因素消除后能迅速恢复原有格局.

#### 3.4.5 社团演化规律总结

城市辐射力是解释城市在城市网络层次结构中的功能、作用和相对位置的重要因素<sup>[31]</sup>. 本文利用改进后的 SL 算法对四阶段网络进行分析, 发现人口流动网络中大部分城市社团以多个一线城市或省会城市为核心, 且具有强烈的区域集聚效应, 形成了以核心城市为点向周边辐射的分布格局. 城市群在四阶段大体保持稳定, 侧面反映出中国城市间的人口流动以城市群内部流动为主. 此外, 春运和疫情的因素并不能使城市社团发生大规模分离. 经济和地理因素对人口流动以及城市社团的影响不会因为突发性传染病事情而消失, 城市群体现了中国城市之间的动态关系.

## 4 总结和讨论

人口迁徙模式对疫情传播起到了举足轻重的作用, 本文使用手机匿名定位大数据, 构建人口流动网络, 通过 6 个中心性指标以及集成空间信息提出的 SL 算法, 从时空大数据分析的角度探究疫情期间人口流动网络的演化模式及空间作用机理.

本文的分析结果表明, 我国城市分布随人口流动而呈现出区域集聚的城市群耦合特征, 分为京津冀、珠三角、成渝等 18 个社团, 其中经济交互和地理位置是城市群社团构建的基础. 居家隔离阶段, 银川等城市离开太原社团加入兰州社团, 说明在突

发应急情况下, 偏远地区人口流动的主导因素是地理位置而不是经济交互. 同时四阶段大部分社团保持稳定, 共有 113 个城市发生社团转移, 其中 63 个城市转移一次 (仅占城市总数的 17%), 47 个城市转移 2 次, 3 个城市转移 3 次. 春运、疫情等事件仅能在一定程度上影响少部分社团演变, 说明中国城市格局具有鲁棒性. 复工返流阶段城市陆续回到常态化阶段的社团, 说明经济和地理因素对人口流动以及城市社团的影响不会因为突发性传染病事情而消失, 城市群体现了中国城市之间的动态关系.

本文基于手机定位大数据, 利用空间可视化和网络分析模型, 从时空演化的角度挖掘了疫情期间全国人口流动背后潜藏的城市群效应以及大规模人口流动在平时与疫情期间的流动模式差异, 验证了控制人口出行模式对疫情防控起到至关重要的作用. 为疫情防控、分析疾病传播路径以及应对突发性公共卫生事件开辟了一种通用的分析范式.

## 参考文献

- [1] Halloran M E, Vespignani A, Bharti N, Feldstein L R, Alexander K, Ferrari M, Shaman J, Drake J M, Porco T, Eisenberg J N 2014 *Science* **346** 433
- [2] Jia J S, Lu X, Yuan Y, Xu G, Jia J, Christakis N A 2020 *Nature* **582** 389
- [3] Colizza V, Barrat A, Barthélemy M, Vespignani A 2006 *Proc. Natl. Acad. Sci. U.S.A.* **103** 2015
- [4] Li Q, Guan X H, Wu P, Wang X Y, Zhou L, Tong Y Q, Ren R Q, Leung S, Lau E, Wong J, Xing X S, Xiang N J, Wu Y, Li C, Chen Q, Li D, Liu T, Zhao J, Li M, Tu W X, Chen C D, Jin L M, Yang R, Wang Q, Zhou S H, Wang R, Liu H, Luo Y B, Liu Y, Shao G, Li H, Tao Z F, Yang Y, Deng Z Q, Liu B X, Ma Z T, Zhang Y P, Shi G Q, Lam T, Wu J, Gao G, Cowling B, Yang B, Leung G, Feng Z J 2020 *N. Engl. J. Med.* **382** 1199
- [5] Guan W J, Ni Z Y, Hu Y, Liang W H, Ou C Q, He J X, Liu L, Shan H, Lei C L, Hui D S 2020 *N. Engl. J. Med.* **382** 1708
- [6] Liu Y X, Yang Y, Zhang C, Huang F M, Wang F X, Yuan J, Wang Z Q, Li J X, Li J M, Feng C 2020 *Sci. China, Ser. C Life Sci.* **63** 364
- [7] Zhou T, Liu Q, Yang Z, Liao J, Yang K, Bai W, Lu X, Zhang W 2020 *J. Evid.-Based Med.* **13** 3
- [8] Luan R S, Wang X, Sun X, Zhou T, Liu Q H, Lu X, Wu X P, Gu D Q, Tang M S, Cui H J, Shan X F, Ouyang J, Zhang B, Zhang W 2020 *J. Sichuan Univ.* **51** 131 (in Chinese) [栾荣生, 王新, 孙鑫, 陈兴蜀, 周涛, 刘权辉, 吕欣, 吴先萍, 谷冬晴, 唐明霜, 崔慧杰, 单雪峰, 欧阳净, 张本, 张伟 2020 *四川大学学报 (医学版)* **51** 131]
- [9] Zhou T, Liu Q, Yang Z, Liao J, Yang K, Bai W, Lu X, Zhang W 2020 *J. Evid.-Based Dent. Pract. Med. China* **20** 359 (in Chinese) [周涛, 刘权辉, 杨紫陌, 廖敬仪, 杨可心, 白薇, 吕欣, 张伟 2020 *中国循证医学杂志* **20** 359]
- [10] Tan S Y, Cao Z Q, Qin S, Chen S R, Sai B, Guo S H, Liu C

- C, Cai M S, Zhou T, Zhang W, Lu X 2020 *J. Univ. Electron. Sci. Technol. China* **49** 788 (in Chinese) [谭索怡, 曹自强, 秦烁, 陈洒然, 赛斌, 郭淑慧, 刘楚楚, 蔡梦思, 周涛, 张伟, 吕欣 2020 电子科技大学学报 **49** 788]
- [11] Wei Y Z, Lu Z Z, Du Z C, Zhang Z J, Zhao Y, Shen S P, Wang B, Hao Y T, Chen F 2020 *Chinese J. Epidemiology* **41** 470 (in Chinese) [魏永越, 卢珍珍, 杜志成, 张志杰, 赵杨, 沈思鹏, 王波, 郝元涛, 陈峰 2020 中华流行病学杂志 **41** 470]
- [12] Cohen J 2020 <https://www.sciencemag.org/news/2020/02/scientists-are-racing-model-next-moves-coronavirus-thats-stillhard-predict> [2021-1-11]
- [13] Brockmann D, Helbing D 2013 *Science* **342** 1337
- [14] Wu J T, Leung K, Leung G M 2020 *The Lancet* **395** 689
- [15] Wang C, Yan J, Wang X, Li M 2020 *Acta Phys. Sin.* **69** 080701 (in Chinese) [王聪, 严洁, 王旭, 李敏 2020 物理学报 **69** 080701]
- [16] Gross B, Zheng Z, Liu S, Chen X, Sela A, Li J, Li D, Havlin S 2020 *arXiv*: 2003.08382 [Statistical Physics]
- [17] Xu X K, Wen C, Sun G Y, Sun H C, Liu B, Wang X W 2020 *J. Univ. Electron. Sci. Technol. China* **49** 324 (in Chinese) [许小可, 文成, 张光耀, 孙皓宸, 刘波, 王贤文 2020 电子科技大学学报 **49** 324]
- [18] Tian H Y, Liu Y H, Li Y D, Wu C H, Chen B, Kraemer M U, Li B Y, Cai J, Xu B, Yang Q Q 2020 *Science* **368** 638
- [19] Xu C H, Yu Y G, Yang Q C, Lu Z Z 2020 *arXiv*: 2004.12541 [physics. soc-ph]
- [20] Bengtsson L, Gaudart J, Lu X, Moore S, Wetter E, Sallah K, Rebaudet S, Piarroux R 2015 *Sci. Rep.* **5** 8923
- [21] Wilson R, Erbach S E, Albert M, Power D, Tudge S, Gonzalez M, Guthrie S, Chamberlain H, Brooks C, Hughes C 2016 *PLoS. Curr.* **1** 8
- [22] Lu X, Bengtsson L, Holme P 2012 *Proc. Natl. Acad. Sci. U.S.A.* **109** 11576
- [23] 中华人民共和国工业和信息化部无线电管理局 (miit.gov.cn) [https://www.miit.gov.cn/jgsj/wgj/gzdt/art/2020/art\\_87ace87acac0426a99f4213e4d578cac.html](https://www.miit.gov.cn/jgsj/wgj/gzdt/art/2020/art_87ace87acac0426a99f4213e4d578cac.html) [2021-1-11]
- [24] 中华人民共和国工业和信息化部无线电管理局 (miit.gov.cn) [https://www.miit.gov.cn/jgsj/wgj/gggs/art/2020/art\\_9212747d4c794f919bd4aa03a4ca2fcf.html](https://www.miit.gov.cn/jgsj/wgj/gggs/art/2020/art_9212747d4c794f919bd4aa03a4ca2fcf.html) [2021-1-11]
- [25] Holme P, Saramaki J 2012 *Phys. Rep.* **519** 97
- [26] Fortunato S 2010 *Phys. Rep.* **486** 75
- [27] Zipf G K 1946 *Am. Sociol. Rev.* **11** 677
- [28] Blondel V D, Guillaume J L, Renaud L, Etienne L 2008 *J. Stat. Mech.: Theory Exp.* **10** P10008
- [29] Newman M E J 2004 *Phys. Rev. E* **70** 056131
- [30] 新一线城市研究所 (yicai.com) <https://www.maigoo.com/news/550235.html> [2021-1-11]
- [31] Barbosa H, Barthelemy M, Ghoshal G, James C R, Lenormand M, Louail T, Menezes R, Ramasco J J, Simini F, Tomasini M 2018 *Phys. Rep.* **734** 1

# Measuring the impact of COVID-19 on China's population migration with mobile phone data\*

Dai Bi-Tao<sup>1)</sup> Tan Suo-Yi<sup>1)</sup> Chen Sa-Ran<sup>2)</sup> Cai Meng-Si<sup>1)</sup>  
Qin Shuo<sup>2)</sup> Lu Xin<sup>1)†</sup>

1) (*College of Systems Engineering, National University of Defense Technology, Changsha 410073, China*)

2) (*State Key Laboratory on Blind Signal Processing, Chengdu 610041, China*)

(Received 8 December 2020; revised manuscript received 2 February 2021)

## Abstract

Population migration is an essential medium for the spread of epidemic, which can accelerate localized outbreaks of disease into widespread epidemic. Large-scale population movements between different areas increase the risk of cross-infection and bring great challenges to epidemic prevention and control. As COVID-19 can spread rapidly through human-to-human transmission, understanding its migration patterns is essential to modeling its spreading and evaluating the efficiency of mitigation policies applied to COVID-19. Using nationwide mobile phone data to track population flows throughout China at prefecture-level, we use the temporal network analysis to compare topological metrics of population mobility network during two consecutive months between before and after the outbreak, i.e. January 1st to February 29th. To detect the regions which are closely connected with population movements, we propose a Spatial-Louvain algorithm through adapting a gravity attenuation factor. Moreover, our proposed algorithm achieves an improvement of 14% in modularity compared with the Louvain algorithm. Additionally, we divide the period into four stages, i.e. normal time, Chunyun migration, epidemic interventions, and recovery time, to describe the patterns of mobility network's evolution. Through the above methods, we explore the evolution pattern and spatial mechanism of the population mobility from the perspective of spatiotemporal big data and acquire some meaningful findings. Firstly, we find that after the lockdown of Wuhan and effective epidemic interventions, a substantial reduction in mobility lasted until mid-February. Secondly, based on the economic interaction and geographic location, China has formed an urban agglomeration structure with core cities centering and radiating toward the surroundings. Thirdly, in the extreme cases, the dominant factor of population mobility in remote areas is geographic location rather than economy. Fourthly, the urban agglomeration structure of cities is robust so that when the epidemic weakens or disappears, the city clusters can quickly recover into their original patterns.

**Keywords:** COVID-19, mobile big data, population flow, spatio-temporal evolution

**PACS:** 89.65.Cd, 89.65.-s, 89.75.Kd, 64.60.aq

**DOI:** [10.7498/aps.70.20202084](https://doi.org/10.7498/aps.70.20202084)

\* Project supported by the National Science Fund for Distinguished Young Scholars of China (Grant No. 72025405), the National Natural Science Foundation of China (Grant Nos. 82041020, 71901067, 72001211), the Major Program of the National Natural Science Foundation of China (Grant No. 91846301), the Science and Technology New Coronary Pneumonia Emergency Program of Sichuan Province, China (Grant No. 2020YFS0007), the Natural Science Foundation of Hunan Province, China (Grant No. 2020JJ5679), and the Key R&D Program of Hunan Provincial Science and Technology Department, China (Grant No. 2019GK213).

† Corresponding author. E-mail: [xin\\_lyu@sina.com](mailto:xin_lyu@sina.com)