

文章编号:2097-4558(2024)06-1508-13

# 城市社区尺度下 COVID-19 时空扩散特征与影响因素 ——以上海疫情为例

李周平<sup>1</sup>, 葛如一<sup>1</sup>, 郭晓爽<sup>2</sup>

(1. 上海商学院 商务信息学院, 上海 200235; 2. 上海理工大学 中英国际学院, 上海 200093)

**【摘要】**针对新型突发传染病疫情在大型城市传播过程中的区域防控问题,提出了一种能够有效识别局部传播源、划分扩散边界的时空事件聚类模型。以2022年3月上海市新冠疫情为例,研究结果表明:在0.8、1.5和2.5 km这3个尺度上制定区域防控策略能有效控制疫情的近邻扩散与迁移扩散;而热点区域内公交站数量、购物点数量和是否存在三甲医院3个空间风险因子可作为区域防控决策的辅助因素,地铁站数量、有无大型商业中心和有无大型超市3个空间风险因子可辅助预判疫情的迁移扩散路径。

**关键词:**疫情防控;密度峰值聚类;城市社区;时空扩散

中图分类号:C 939

文献标志码:A

DOI:10.3969/j.issn.2097-4558.2024.06.010

## Spatiotemporal Spreading Characteristics and Influence Factors of COVID-19 in Urban Community Level: Evidence from Shanghai

LI Zhouping<sup>1</sup>, GE Ruyi<sup>1</sup>, GUO Xiaoshuang<sup>2</sup>

(1. Faculty of Business Information, Shanghai Business School, Shanghai 200235, China;

2. Sino-British College, University of Shanghai for Science and Technology, Shanghai 200093, China)

**【Abstract】**For the problem of regional prevention and control of infectious diseases in large cities, this paper proposes a novel spatiotemporal clustering model which can effectively identify local diffusion sources and divide diffusion boundaries. Taking the COVID-19 in Shanghai in March 2022 as an example, it is shown that developing regional prevention and control strategies at the scale of 0.8 km, 1.5 km, and 2.5 km can effectively control the adjacent diffusion and relocation diffusion of the pandemic. The spatial risk factors of bus stop, shopping spot, and large-scale hospital can assist the formulation of regional prevention and control strategies. The spatial risk factors of metro station, shopping mall, and supermarket can help predict the migration path of the pandemic.

**Keywords:** pandemic prevention and control; density peak clustering; urban community; spatiotemporal diffusion

收稿日期:2022-09-08 修订日期:2023-03-15

基金项目:国家社会科学基金支持项目(18ZDA088);上海市“科技创新行动计划”软科学研究项目(23692106100, 22692198700)

作者简介:李周平(1981-),男,博士,副教授。研究方向为商业数据分析、数据治理与决策支持。

通信作者:葛如一(1978-),女,博士,副教授。

E-mail: gery@sbs.edu.cn

自2019年12月湖北武汉爆发新冠疫情以来,新型冠状病毒肺炎(COVID-19)疫情的多轮大规模传播对我国公共卫生安全与社会经济发展造成了严重危害。爆发于2022年3月的上海新冠疫情,持续时间达3个月,累计报告感染人数62万余人。无论从疫情的感染规模、扩散速度,还是防控的复杂程度来看,该轮上海疫情都非常具有代表性。因此,作为新型突发传染病在大型城市爆发的典型案例,上海

新冠疫情防控的时空特征具有很高的研究价值。

在应对 COVID-19 早期疫情的过程中,基于输入感染病例及其密切接触人群出行轨迹追踪隔离的点式防控政策取得了较好的防控效果<sup>[1-2]</sup>。但是,在上海疫情发现首例本土感染者后的第 16 天,风险人群筛查阳性感染者占比就高达 59.6%,这意味着大部分感染者的传播链条难以追踪。究其原因,主要是随着变种病毒传播强度的提升,加之大型城市人群流动的复杂性,使得新冠疫情的扩散过程呈现出随机迁移的扩散特征<sup>[3-4]</sup>。因此,在城市空间的尺度上,上海疫情表现出较高的随机性与复杂性,难以追踪预判。另一方面,大量研究已证实,在城市居民密集接触的社区及其周边生活空间中,新冠疫情的扩散过程呈现出显著的近邻扩散特征<sup>[4-5]</sup>。因此,将整个城市疫区化整为零,在更微观的社区空间尺度上,识别疫情扩散的局部传播源,划分扩散边界,探究局部区域的扩散规律,对于了解大型城市突发传染病疫情的传播机制,优化防控政策,更具可行性与实践意义<sup>[6]</sup>。

目前,国内外对传染病疫情传播的研究主要聚焦于从时间和空间两个维度探究传染病扩散的特征与机制。在时间维度方面,基于经典的 SIR 模型,学者们扩展了 SI、SIS 和 SEIR 模型,从不同的传染机制入手,刻画了感染人群数量随时间变化的规律<sup>[7-9]</sup>。在空间维度方面,国内外文献主要聚焦于传染病疫情的空间聚集性特征、空间分布的影响因素以及空间传播机制研究。Huang 等<sup>[10]</sup>通过莫兰指数分析了疫情传播过程中的全局聚集特征。Spassiani 等<sup>[11]</sup>通过核密度热力图、层次聚类算法分析了疫情传播的局部聚集性分布特征。另一些学者通过多元回归或空间回归模型在网格范围或街道乡镇尺度上,研究了感染者分布与空间风险因子的相关关系<sup>[3,12-13]</sup>。荆磊等<sup>[4]</sup>和 Medrek 等<sup>[14]</sup>通过多智能体与元胞自动机,仿真分析了传染病传播机制与防疫策略对空间扩散的影响机理。

但是,已有研究成果表明,传染病疫情作为典型的时空事件,其时空关系紧密联系不可分割<sup>[4,15]</sup>。结合时间与空间两个维度挖掘疫情的时空扩散特征,对于探究疫情的传播机制及防疫政策的制定更具现实意义。Hu 等<sup>[16]</sup>基于手足口病例的时空数据,通过数学模型,在县域尺度下对比分析了青海省与上海市 2009~2016 年间手足口病例的时空扩散特征。Backer 等<sup>[17]</sup>基于动力学模型和引力模型,分析了 2014~2016 年间埃博拉病毒在西非地区城市间的时空扩散特征与动力学机制。近期,关于 COVID-19 疫情的时空扩散研究也正逐步展开。

Alkhamis 等<sup>[18]</sup>和 Kim 等<sup>[19]</sup>通过时空扫描统计方法,在国家与城市尺度上复盘了新冠疫情暴发中心的时空迁移模式。Ye 等<sup>[20]</sup>通过新冠感染病例的空间质心的时序变化规律,在城市尺度上分析了 2020 年中国长三角地区城市间新冠疫情的时空扩散趋势。吕振华等<sup>[21]</sup>通过地理空间分析技术,在城市尺度上分析了 2020 年湖北省疫情的时空聚集特征。刘勇等<sup>[22]</sup>利用探索性数据分析方法,在区县尺度上从病例总量、输入性扩散性病例数量和扩散比 3 个维度探究了河南省新冠肺炎疫情的时空扩散过程。

综上所述,当前针对传染病时空扩散特征的研究,主要结合空间统计方法与时序分析方法,在各级行政区域尺度上(国家、城市、区县、街道乡镇)分析疫情的时空聚集趋势与疫情中心的迁移规律。这些研究能较好地复盘及分析传染病疫情的时空扩散趋势,有助于决策者在宏观上把握疫情扩散规律,但对于城市社区尺度上的疫情扩散过程关注不够。而更微观尺度上的扩散规律与影响因素研究,对于大型城市应对突发疫情制定防控政策更具指导意义。另外,当前文献大多基于行政区域的边界来分析疫情的时空扩散规律与影响因素,而传染病的扩散过程并不受行政边界的约束。因此,打破行政区域边界的限制,识别局部传播源,划分扩散边界,更有助于掌握传染病疫情的真实扩散情况。

本文的学术贡献在于:

(1) 基于疫情传播的近邻扩散与迁移扩散机制,提出时空加权密度峰值聚类模型(STW-DPC),首次将密度峰值聚类算法应用于局部区域传播源的识别与近邻扩散边界的划分。STW-DPC 模型能够辅助决策者在城市社区尺度上识别疫情的局部传播源及其扩散范围,从而制定区域防控策略。

(2) 基于 STW-DPC 模型复盘 2022 年 3 月上海新冠疫情的扩散过程,在不同的参数组合下分析疫情的区域扩散特征及其防控效果,为区域防控的有效性提供科学依据。

(3) 基于上海新冠疫情的复盘数据,探索疫情的区域扩散特征与城市设施空间分布的相关关系。决策者可以在 STW-DPC 模型基础上,结合区域内城市设施的空间分布情况,以此来预判局部疫情的扩散趋势。

## 1 研究思路与数据

### 1.1 研究区域概况

如图 1 所示,自 2022 年 3 月 1 日上海市在社会面发现首例本土新冠病毒感染者,到 5 月 31 日居住

区有序解除静态管理,全市累计报告感染人数 626 814 人。其间,3 月 17 日社会面筛查阳性占比达到峰值 59.6%,3 月 28 日浦东、浦南及毗邻区域实施静态管理,4 月 1 日浦西区域实施静态管理。

4 月 6 日单日新增感染者居住地数量达到峰值 1 651 个,4 月 13 日单日新增感染者数量达到峰值 27 605 例,5 月 17 日全市实现社会面清零,6 月 1 日解除静态管理。

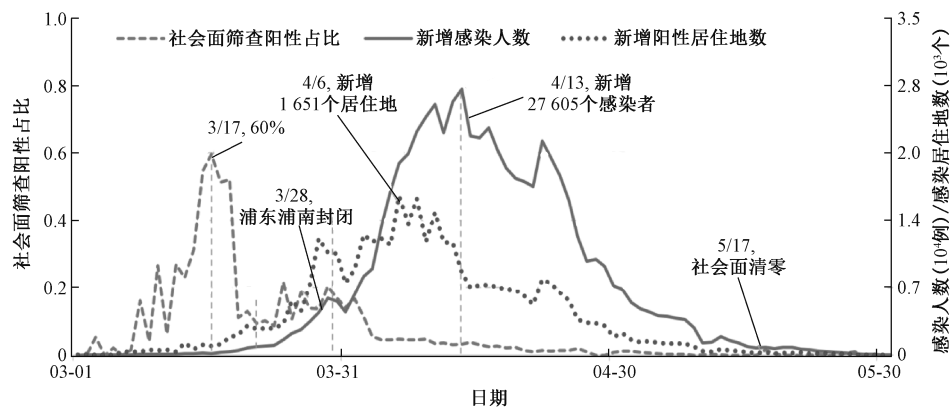


图 1 2022 年 3~5 月上海市新冠疫情概况  
Fig. 1 Overview of COVID-19 in Shanghai from March to May 2022

上海市该轮新冠疫情的防控政策总体分为两个阶段。第 1 阶段,3 月 1 日至 3 月 28 日,疫情防控主要以点式防控为主。该阶段防疫部门通过溯源调查,按照密接、次密接、高风险逐层筛查,对于密接者涉及的所在居住小区、工作单位或学校的相关人员实行“2 天封闭管理+12 天社区健康管理”的管控措施。第 2 阶段,疫情防控以区域防控为主,按照封控区、管控区、防范区的分类标准,分别实施“7 天封闭管理+7 天社区健康管理”“7 天社区健康管理”和“7 天自我健康管理”的管控措施。

1.2 研究思路

为研究城市社区尺度上的传染病疫情扩散的时空特征与影响因素,首先需要界定城市社区,进而明晰传染病疫情在社区尺度上的时空扩散机制。《社区生活圈规划技术指南》指出,社区是聚居在一定地域范围内的人们所组成的社会生活共同体,其核心空间范围与一般意义的居民小区空间范围一致。而以居民居住地为中心,步行 5~15 min 可达的城市空间是满足居民日常生活需求、提供公共基础服务的基本社区生活圈空间<sup>[23]</sup>。如图 2 所示,本文将居民小区界定为基本社区单元  $C_i$ ,将以社区  $C_i$  为中心、 $R$  为半径的圆形区域界定为  $C_i$  的社区生活圈。根据城市居民的出行特点,社区居民的生活圈半径  $R$  一般不超过 1.5 km<sup>[24-25]</sup>。

研究表明,传染病疫情的传播是在一定时空尺度上的扩散过程,扩散过程主要表现为时空相对连续的近邻扩散与时空跳跃的迁移扩散<sup>[4-5]</sup>。在城市社区尺度上,传染病疫情的扩散过程也表现为社区

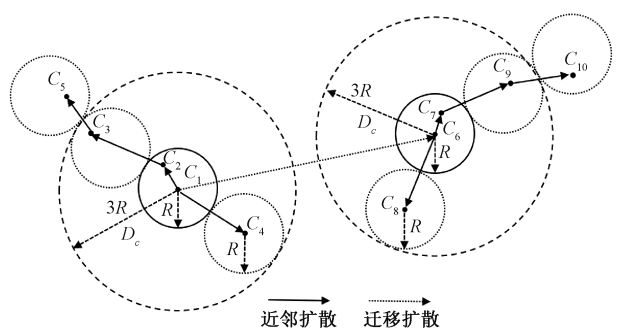


图 2 城市社区尺度上疫情的扩散机制  
Fig. 2 The diffusion mechanism of the epidemic at the urban community scale

近邻扩散与社区迁移扩散两种模式。其中,社区近邻扩散主要是由社区居民日常生活圈出行的交叉接触而造成的疫情传播。由图 2 可见: $R$  为社区居民步行可达的生活圈半径; $2R$  为两个社区生活圈发生重叠的最远距离,距离小于等于  $2R$  的社区为近邻社区; $3R$  为社区  $C_i$  的近邻社区的生活圈可达的最远距离。本文将  $3R$  设定为社区  $C_i$  通过其近邻社区传播的最远距离  $D_c$ 。社区迁移扩散则是社区居民搭乘各类交通工具,在城市空间范围内的较远距离出行而引起的疫情传播。因此,在城市社区尺度上,如何刻画感染社区间的时空近邻性,从而甄别两种扩散模式是揭示疫情扩散规律的关键。本文通过引入时空柱状体来计算感染社区的局部密度,从而刻画感染社区间的时空近邻性。

为进一步分析迁移扩散与近邻扩散特征,还需解决两个关键问题。其一,如何识别局部区域的传播源社区?局部传播源社区是指该社区的疫情由迁



移扩散导致,并且由其引发疫情在近邻社区间的扩散传播。如图 2 中的  $C_1$ 、 $C_6$  社区为局部传播源社区。其二,识别局部传播源社区后,如何划分由该传播源社区通过近邻扩散引发感染的社区?如图 2 中的  $C_2$ 、 $C_3$ 、 $C_4$  和  $C_5$  由传播源社区  $C_1$  通过近邻扩散传播引发感染, $C_7$ 、 $C_8$ 、 $C_9$  和  $C_{10}$  由传播源社区  $C_6$  通过近邻扩散传播引发感染。本文将由近邻扩散引发感染的社区定义为蔓延社区。基于密度峰值聚类算法,构建时空加权密度聚类模型,从而识别局部区域的传播源社区,并通过簇的划分算法确定从属于局部传播源的蔓延社区。

最后,以 2022 年 3 月的上海新冠疫情数据为例,通过时空加权密度聚类模型,识别局部传播源并划分蔓延社区,复盘疫情防控效果,并结合地图兴趣点数据,探究城市设施的空间分布与该轮上海新冠疫情扩散特征的关系。

1.3 数据来源

如表 1 所示,研究数据来自上海市卫健委发布的每日本土新增确诊病例、无症状感染者居住地信息,居住地理位置及城市设施数据来自链家网公布的小区数据及高德地图的兴趣点数据。鉴于 3 月 28 日后上海市全域陆续实施静态管理,人员基本不再流动,疫情在社区间的扩散传播被阻断。本文仅采用了 03-01~03-27 的疫情数据。另外,考虑到在 3 月 28 日之前,如果社区发现首例感染者,该社区将被立即实施静态管理,疫情通过该社区的扩散传播被阻断。因此,本文的数据集仅保留同一社区发现首例感染者的日期数据。基于卫健委发布的场所地址数据,进一步通过高德地图与链家网小区信息获取每个场所地址的地理坐标。最终获取 4 391 个疫情感染场所的地址数据,剔除 486 个不能定位居民小区坐标的场所,剩余 3 905 个感染社区。其中,486 个不能精确定位居民小区坐标的场所基本为工地、村庄与办公地等非居民小区,不属于本文的研究范围,故剔除后对论文的研究结论影响不大。

表 1 研究数据来源

Tab. 1 The source of research data

数据名称	数据来源	数据集字段
感染者居住地	上海市卫健委每日发布的感染者数据高德地图地理编码接口	场所首例阳性感染者感染日期、场所地址、坐标
兴趣点数据 (POI)	高德地图 POI 搜索接口	场所名称、类型、坐标
居民小区	链家网居民小区数据	居民小区地址、坐标、居民户数

2 基本模型

本文基于密度峰值聚类算法 (Density Peaks Clustering, DPC),结合传染病疫情的时空扩散机制,引入密度加权策略与时间约束,构建时空加权密度峰值聚类模型 (STW-DPC)。

2.1 密度峰值聚类算法

密度峰值聚类算法 (DPC)是由 Rodriguez 等<sup>[26]</sup>提出的基于密度的空间聚类算法。该算法的聚类思想是簇中心的密度应该在局部范围内最大,即簇中心点被密度不超过它的邻居包围,并且簇中心与密度更大的空间点之间的距离应该相对较远。DPC 算法根据空间样本点  $i$  的局部密度  $\rho_i$  与该样本点的相对距离  $\delta_i$ , 决定该样本点是否为簇中心。局部密度的定义为

$$\rho_i = \sum_j^n \exp\left(-\frac{d_{ij}^2}{D_c^2}\right)$$

(1)

式中:  $d_{ij}$  为样本点  $i$ 、 $j$  的欧式距离,通常引入高斯核函数来计算  $\rho_i$ ;距离阈值  $D_c$  是该算法需要输入的唯一超参数。

相对距离  $\delta_i$  为距离样本点  $i$  最近且局部密度大于  $\rho_i$  的样本点与样本点  $i$  之间的欧式距离,其定义如下式所示。对于数据集中局部密度最大的样本点,其相对距离默认为最大,

$$\delta_i = \min_{\rho_j > \rho_i} (d_{ij})$$

(2)

DPC 算法基于以下两种假设实施聚类:① 簇中心点应该是局部范围内的密度峰值点;② 非中心点应与其最近的高密度点属于同一个簇。基于上述两种假设,DPC 算法的聚类过程分为两个步骤:首先,根据样本点的  $\rho$  和  $\delta$ , 辅以决策图,将局部密度与相对距离均较大的空间样本点判定为簇中心点;其次,将非中心点划分到局部密度比它高且距离它最近的样本点的簇中。

图 3(a)所示为根据 28 个空间样本点的局部密度与相对距离绘制的决策图。可以发现,分布在决策图右上角样本点 1 和样本点 10 具有高  $\rho$  和高  $\delta$ , 表明这两个样本点在较大的空间范围内为密度峰值点。选取样本点 1 和 10 为簇中心。图 3(b)显示,根据簇划分算法,划分出以样本点 1 和 10 为簇中心的两个簇,并将样本点 26、27 和 28 识别为噪声点。

2.2 时空加权密度峰值聚类模型 (STW-DPC)

DPC 算法结合了密度聚类和划分聚类算法的优点,不仅可以通过局部密度与相对距离的决策图辅助识别簇中心,还可以基于局部密度的空间近邻性划分非中心点所属的簇,有助于本文在城市社区

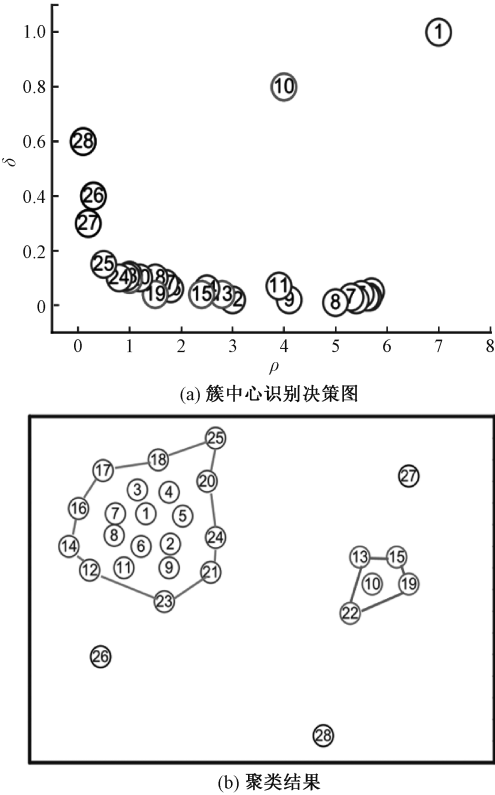


图 3 DPC 算法在二维空间的聚类示例

Fig. 3 The example of DPC algorithm clustering in two-dimensional space

尺度下,识别疫情传播过程中的局部传播源社区(簇中心点)与蔓延社区(非中心点)所属的簇。但是,DPC 算法在选取簇中心与划分非中心点时,仅考虑了样本点的空间近邻关系,并不能有效刻画传染病传播这类时空事件扩散过程中的时空近邻关系,也不能反映传播源事件的时序特征。

Dodds 等<sup>[27]</sup>在经典的 SIR 模型基础上引入了接触感染机制,建立了通用传染病模型。该模型认为易感者 S 被感染的概率与其和每个感染者 I 接触的概率的总和正相关。在感染者数量随时间指数增长的传播网络中,一个新增的感染者与系统中已经存在的某一个特定感染者之间的直接传播关系不仅随两者之间距离的增大而减弱,同时也应随着系统中已有感染者数量的增加而减弱。Sun 等<sup>[28]</sup>从传播网络视角论证了节点数量指数增长的动态网络中,新老节点间的联系随时序呈指数衰减的规律。在传染病疫情暴发初期,新老感染的社区之间的传播关系同样符合这一规律。鉴于此,引入时空柱状体概念,并在 DPC 算法的基础上加入时空密度加权策略与时间间隔约束,构建了一种基于加权密度峰值的时空聚类模型(STW-DPC)。

首先,如图 4 所示,STW-DPC 模型将时空事件

$e_i$  定义为在时间点  $t_i$ 、空间位置点  $x_i$  发生的事件。对应本文的数据集,时空事件  $e_i$  为新增感染社区,  $x_i$  为该社区对应的地理坐标,  $t_i$  为该社区首次出现感染病例的日期。

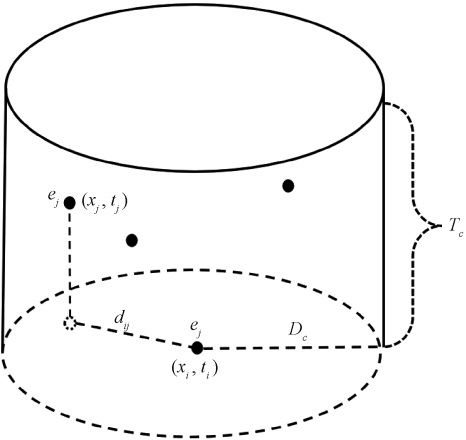


图 4 时空柱状体

Fig. 4 The spatiotemporal cylinder

其次,时空事件的近邻关系不仅随空间距离而衰减,同时新老事件  $e_j$ 、 $e_i$  间传播关系的强度也随时间推移呈指数衰减的规律<sup>[27-28]</sup>。鉴于此,STW-DPC 模型引入时间衰减函数来度量新老事件之间传播关系的强度随时间衰减的特征。新发生的事件  $e_j$  在  $t_j$  时刻发生时与已经发生的事件  $e_i$  的传播关系强度为  $w_{ij}$ , 即

$$w_{ij} = \exp(-\alpha t_j), \quad \alpha > 0 \tag{3}$$

该强度随着新事件发生时间  $t_j$  的推移呈指数衰减,其中,  $\alpha$  为时间衰减系数。

最后,考虑疫情传播过程中病毒感染存在一定潜伏期的特点(见图 4),STW-DPC 模型引入时间截止距离参数  $T_c$ <sup>[29-30]</sup>。模型中,事件  $e_i$  仅可能与  $[t_i, t_i + T_c]$  时间段内发生的感染事件存在传播链上的直接关联。《新型冠状病毒肺炎诊疗方案(试行第九版)》明确指出新冠感染者的潜伏期多为 3 ~ 7 天,因此,本文将  $T_c$  设定为 7 天。

综上所述,STW-DPC 模型将 DPC 算法计算局部密度式(1)调整为

$$\rho_i = \sum_{t_j \in [t_i, t_i + T_c]} w_{ij} \exp\left(-\frac{d_{ij}^2}{D_c^2}\right) \tag{4}$$

式中:  $\rho_i$  为时空事件  $e_i$  的局部密度;  $D_c$  为社区间发生近邻扩散的距离阈值。该式在 DPC 模型中引入新老事件传播关系强度  $w_{ij}$  与时间截止距离参数  $T_c$ 。进一步,将式(3)代入式(4),可得

$$\rho_i = \sum_{t_j \in [t_i, t_i + T_c]} \exp\left(-\alpha t_j - \frac{d_{ij}^2}{D_c^2}\right) \tag{5}$$

式(5)通过调整时间衰减系数 $\alpha$ 与近邻扩散距离阈值 $D_c$ 来刻画不同传播场景下各事件点的局部密度的时空加权特征。当 $\alpha$ 较大且 $D_c$ 较大时,局部密度侧重于反映事件点时序特征的聚集性;当 $\alpha$ 较小且 $D_c$ 较小时,局部密度侧重于反映事件点的空间聚集性。

### 3 社区聚类与聚类有效性

#### 3.1 社区聚类

将 STW-DPC 模型应用于疫情传播数据集,而在社区尺度上识别传播源社区并划分蔓延社区。基于以下假设实施聚类:

(1) 传染病疫情是由高密度点向距离最近的低密度点扩散。

(2) 与 DPC 算法的簇中心识别机制相同,传播源社区也应该是局部区域内的密度峰值点,可通过改变近邻扩散的距离阈值参数 $D_c$ 来改变局部区域覆盖的范围。

(3) 蔓延社区发生的疫情是由其最近的高密度点通过近邻扩散引发,应该与其最近的高密度点同属一个簇,即由同一个传播源社区通过近邻扩散引发感染。

由图 2 可见, $D_c$ 作为近邻扩散的距离阈值,满足 $\delta_i > D_c$ 条件的事件点,是近邻扩散空间范围内局部密度最大的社区。根据假设 1 和 2,所有符合 $\delta_i > D_c$ 条件的事件点为传播源社区(簇中心)。因此,利用 STW-DPC 模型实施聚类的步骤为:首先,根据式(5)计算每个感染社区的局部密度 $\rho_i$ ,进而计算相对距离 $\delta_i$ ;其次,将满足 $\delta_i > D_c$ 条件的社区设定为传播源社区;最后,根据假设 3,将除传播源社区之外的所有蔓延社区划定到各自的簇中。

#### 3.2 近邻扩散可控性与迁移扩散可控性

通过 STW-DPC 模型对时空事件实施聚类,还需确定参数 $D_c$ 与 $\alpha$ 的值。应用聚类算法实施聚类时,可以根据聚类有效性函数来评价聚类效果,从而选取最优的参数。常用的聚类有效性函数包括轮廓系数、分裂函数以及 Dunn 函数等。这些聚类有效性函数均通过簇内凝聚度和簇间分离度两个指标的加权组合来评价聚类的有效性<sup>[31]</sup>。但对于传染病疫情扩散这种时空事件聚类,难以从时空两个维度同时评价簇内凝聚度和簇间分离度。因此,本文从疫情的防控视角,提出近邻扩散可控性与迁移扩散可控性两个指标来评价 STW-DPC 模型的聚类有效性,从而选取聚类后防控效果较优的参数组合。随着疫情的传播,簇内因近邻扩散而感染的社区数量

会增加,感染社区分布的空间范围也会扩大,但在该簇发生疫情感染的初期,热点区域内感染事件的可控程度将直接决定该簇后续近邻扩散的控制效果。

基于 STW-DPC 模型应用于疫情扩散研究的基本假设,在不考虑近邻扩散的方向性差异情况下,将以传播源社区为圆心、 $D_c$ 为半径所覆盖的圆形区域,定义为各个簇的热点区域。而某个簇在其发生首例感染事件的第 1 个 $T_c$ 时间窗口内出现的感染社区中,热点区域所捕获的感染社区数占比,反映了该簇近邻扩散初期的可控程度。基于此,引入加权平均命中率指标对聚类后近邻扩散的整体可控性进行定量评价。平均命中率是用来衡量稀疏时空事件热点区域预测的评估指标。该指标为热点区域准确捕获的时空事件数量 $n$ 占时空事件总数的比例<sup>[32]</sup>。考虑到样本期内最终发生时空事件越多的簇的近邻扩散可控性对系统整体的近邻扩散控制的贡献应该越大,因此,在平均命中率的基础上,引入各个簇最终发生的事件占比作为权值,形成加权平均命中率指标,即

$$S(C) = \sum_{i=1}^K \frac{n_{in}(C_i)}{n_{in}(C_i) + n_{out}(C_i)} \frac{N(C_i)}{N(C)} \quad (6)$$

式中: $C = \{C_1, C_2, \dots, C_K\}$ 为聚类后簇的集合; $S(C)$ 为所有簇整体的近邻扩散可控性; $n_{in}(C_i)$ 、 $n_{out}(C_i)$ 分别为簇 $C_i$ 在其首个 $T_c$ 时间窗口期间热点区域内外感染的社区数量; $N(C_i)$ 为样本集中簇 $C_i$ 包含的感染社区数; $N(C)$ 为所有簇包含的感染社区总数。

根据 STW-DPC 模型的聚类机制,不难发现,随着参数 $D_c$ 的减小,聚类后划分的簇的数量随之增加,而单个簇的近邻扩散空间则相应减小。如果参数 $D_c$ 小到每个簇只能包含 1 个感染社区,整体的近邻扩散可控性将达到最大值 1。另一方面,随着簇的数量增加,感染社区在簇间发生迁移扩散的可能性也随之增加。因此,本文进一步提出迁移可控性指标,该指标为聚类后蔓延社区数占感染社区总数的比例。根据 STW-DPC 模型的设定,各个簇局部密度最大的事件点为该簇的传播源社区,除传播源社区之外的事件点均为近邻扩散引发感染的蔓延社区。因此,样本集中传播源社区的数量实质上等于聚类后簇的数量。迁移可控性越高,说明样本集中传播源社区的占比越小,疫情在簇间的迁移扩散越容易控制。如下式所示:

$$M(C) = 1 - \frac{K}{N(C)} \quad (7)$$

式中: $C$ 为聚类后所有簇的集合; $M(C)$ 为聚类后所



有簇整体的迁移可控性; $K$  为簇的数量。

4 结果分析

4.1 聚类分析

应用 STW-DPC 模型对 03-01~03-27 的上海疫情数据进行聚类分析。将式(5)中的参数  $T_c$  设置为 7 天,将近邻扩散的距离阈值参数  $D_c$  的取值范围设定为 0.1~4.5 km,其中,0.1 km 为覆盖 1 个社区的最小距离,4.5 km 为图 2 中 3 倍于最大社区生活圈半径的距离值。将时间衰减系数  $\alpha$  设定为 0~1 的取值范围,当  $\alpha=0$  时,将不考虑社区感染的时序特征。

分别计算  $D_c$  与  $\alpha$  不同参数组合聚类后的近邻扩散可控性与迁移扩散可控性。如图 5(a)所示,当  $D_c$  不变时,随着  $\alpha$  的增大,聚类后的近邻扩散可控性  $S(C)$  迅速增加,但当  $S(C)$  达到峰值后,随着  $\alpha$  的增大, $S(C)$  值保持平稳。当  $\alpha$  不变时,随着  $D_c$  的增大, $S(C)$  总体呈现出递减趋势。但当  $D_c$  值为 0.1、0.8、1.5、2.5 km 时,  $S(C)$  出现了 4 个波峰。其中: $D_c=0.1$  km 时,  $S(C)$  几乎不随  $\alpha$  变化,其值

均大于 0.99;当  $D_c=0.8$  km 时,  $S(C)$  值在  $\alpha=0.3$  时达到最大值 0.94;当  $D_c=1.5$  km 时,  $S(C)$  值在  $\alpha=0.3$  时达到最大值 0.92;当  $D_c=2.5$  km 时,  $S(C)$  值在  $\alpha=0.4$  时达到最大值 0.89。

由图 5(b)可见,聚类后的迁移扩散可控性  $M(C)$  几乎不受  $\alpha$  的影响,而随着  $D_c$  的增大,  $M(C)$  迅速增加,之后增长趋势显著放缓。当  $D_c$  为 0.1、0.8、1.5、2.5 km 时,  $M(C)$  分别等于 0.17、0.85、0.93 和 0.95。

综上所述,随着  $D_c$  参数值的增大,近邻扩散防控的难度总体上随之增加。当  $D_c$  一定时,可以通过调节时间衰减系数  $\alpha$  来增加首个潜伏期内热点区域的命中率,从而提升近邻扩散的防控效果。另一方面,虽然  $D_c$  越小,近邻扩散防控的难度越低,但随着簇数量的增加,簇之间发生随机迁移扩散的可能性随之增大。当  $D_c=0.1$  km 时,虽然对热点区域的防控能够有效控制 99%的近邻扩散,但所有感染社区中有 83%的社区感染由迁移扩散引发。这也能有效解释在该轮上海疫情暴发初期,为什么点式防控政策难以有效控制社会面疫情的扩散。当  $D_c$  为 0.8、1.5、2.5 km 时,近邻扩散可控性在递减趋势下呈现出 3 个波峰式增长。这说明,针对该轮上海疫情,选取 0.8、1.5、2.5 km 作为热点区域的防控半径,能够较好地兼顾近邻扩散与迁移扩散的控制。如果将 0.8 km 设定为社区居民的生活圈半径  $R$ ,而 1.5 km 与 2.5 km 约为生活圈半径的 2 倍与 3 倍距离,在一定程度上也说明了在城市社区尺度上,基于社区居民的生活圈出行距离,可以解释大型城市传染病疫情的扩散机制(见图 2),并以此制定区域防控策略是合理的。

进一步选取  $D_c=0.8$  km,  $\alpha=0.3$ ;  $D_c=1.5$  km,  $\alpha=0.3$ ;  $D_c=2.5$  km,  $\alpha=0.4$  这 3 组参数组合聚类,并按各个簇传播源社区的局部密度由高到低排序,分别选取前 5%、前 10%与前 15%的簇计算近邻扩散可控性  $S(C)$ , 计算结果如表 2 所示。

表 2 不同参数组合下近邻扩散控制效率

Tab. 2 The efficiency of neighbor diffusion control under different parameter combinations

参数组合	簇的数量	前 5%的簇 $S(C)$	前 10%的簇 $S(C)$	前 15%的簇 $S(C)$
$D_c=0.8, \alpha=0.3$	643	0.29	0.43	0.51
$D_c=1.5, \alpha=0.3$	308	0.34	0.48	0.58
$D_c=2.5, \alpha=0.4$	169	0.40	0.54	0.62

表 2 的结果表明,按局部密度大小等比例选取热点区域进行区域防控,  $D_c=2.5$  km,  $\alpha=0.4$  参数

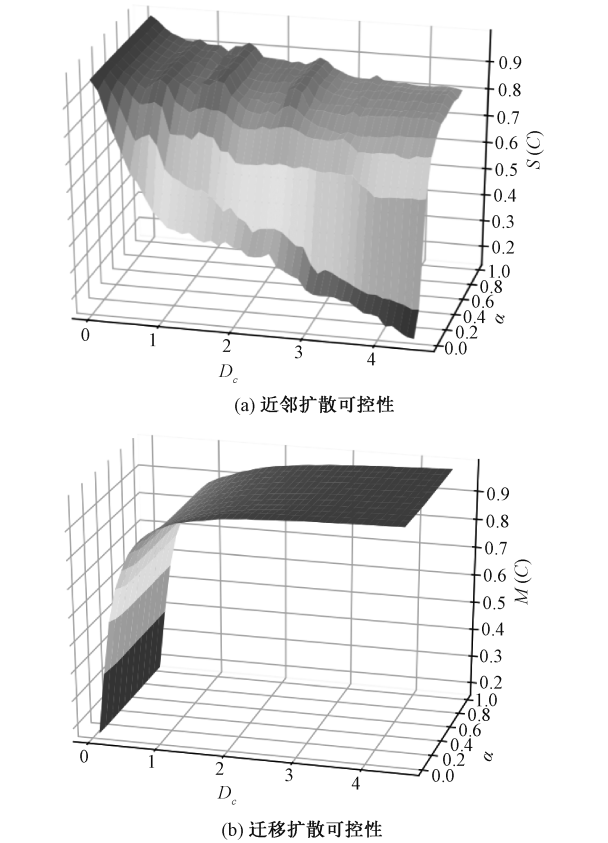


图 5 不同参数组合聚类后近邻扩散可控性与迁移扩散可控性

Fig. 5 The controllability of neighboring diffusion and migration diffusion with different clustering parameters

组合下近邻扩散的防控效率最高。在该参数组合下,从 169 个簇中选取前 5% 的 8 个簇进行区域防控,可有效控制 40% 的近邻扩散社区,选取前 15% 的 25 个簇进行区域防控,则可有效控制 62% 的近邻扩散社区。

对近邻扩散控制效率最高的参数组合  $D_c=2.5$  km,  $\alpha=0.4$  聚类后做进一步分析,如图 6 所示。

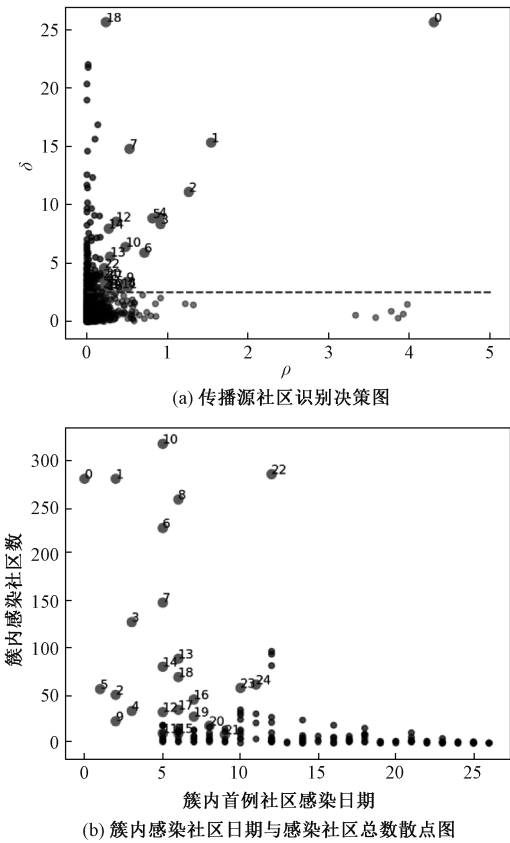


图 6 参数  $D_c=2.5$  km,  $\alpha=0.4$  聚类分析  
Fig. 6 Cluster analysis with  $D_c=2.5$  km,  $\alpha=0.4$

图 6(a) 为识别传播源社区的决策图,依据 STW-DPC 模型传播源社区的判断条件  $\delta_i > D_c$ ,将相对距离大于 2.5 km 的 169 个事件点识别为传播源社区(簇中心),即图 6(a)中虚线上方的点。虚线下方的 3 736 个事件点划分为蔓延社区。图中标注了传播源社区的局部密度排序前 15% 的簇,其中,序号越小,传播源社区的局部密度越大。图 6(b) 为簇内首例感染社区日期与感染社区总数的散点图,可以发现,排序前 15% 的 25 个簇在样本期内感染的社区总数显著高于排序靠后的簇,并且这 25 个簇的首例感染社区日期也显著早于其他簇。这说明,传播源社区的局部密度能够有效地反映各个簇的感染日期与感染社区数这两个重要特征。

在图 7 中,以传播源社区为圆心、2.5 km 为半径,绘制了排序靠前的 25 个簇的热点区域,以及各

簇在其首个 7 天时间窗口感染社区的位置分布,并按传播源社区的局部密度,由高到低标注了热点区域的序号。

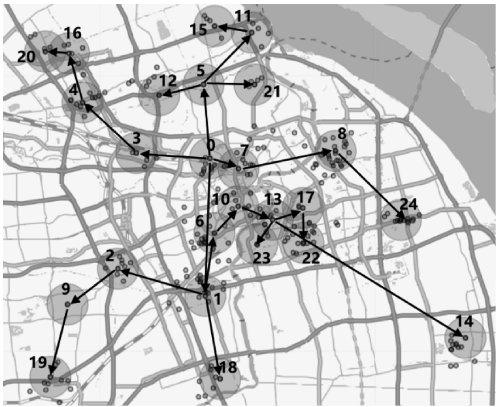


图 7 簇热点区域及首个 7 日时间窗口感染社区分布  
Fig. 7 The distribution of infected communities in the first 7-day time window of the cluster hotspot area

由图 7 可以发现,各个簇的热点区域能够有效地覆盖其首个 7 天时间窗口出现的感染社区。因此,对局部密度较高的热点区域进行区域防控,能够有效控制该簇的近邻扩散。基于疫情传播是由高密度点向距离最近的低密度点扩散的基本假设,图 7 中通过箭头线标识了疫情在簇间的迁移路径。由迁移路径可以发现,该轮上海疫情的 0 号与 1 号传播源社区分别发生在普陀区与徐汇区。以 0 号、1 号传播源社区为中心,各传播源社区呈现出局部密度由近及远递减的趋势。围绕 0 号社区,疫情向西北方向的嘉定区(3、4、16、20 号社区),东向的静安区、浦东新区(7、8、24 号社区),以及北向的宝山区(5、11、15 号社区)迁移扩散。围绕 1 号社区,疫情向西南方向的松江区(2、9、19 号社区),北向的徐汇区、黄浦区(6、10 号社区),东向的浦东新区(13、14、17 号社区),以及南向的闵行区(18 号社区)迁移扩散。综上所述,在参数组合  $D_c=2.5$  km,  $\alpha=0.4$  聚类后,观察上海疫情暴发初期的扩散过程,可以发现,疫情迁移扩散的时空特征并非完全随机,而是表现出围绕 0 号与 1 号传播源社区这两个中心向周边由近及远蔓延的轴辐式扩散模式。因此,如果能对局部密度较高的热点区域进行防控,则有可能抑制疫情的迁移扩散趋势。

4.2 回归分析

在 3 月 28 日上海实施静态管理之前,社区居民日常生活出行的目的地主要集中于社区生活圈内的城市功能设施,而城市功能设施周边居民间的近距离接触正是疫情扩散的重要原因。为进一步探究社区尺度下近邻扩散与迁移扩散的时空特征与空间风



险因子的关系。本文选取了 3 个风险指标代表近邻扩散与迁移扩散的风险程度。参考文献[3,12-13,25],从高德地图提取了 7 类城市功能设施的 POI 数据作为空间风险因子。

簇  $C_i$  出现首例感染社区后的一段时间内,进一步发生感染事件的数量与扩散范围是疫情在该区域近邻扩散的重要时空特征。感染事件越多、扩散范围越广,说明传染病疫情在该区域发生近邻扩散的风险越高,防控难度越大。因此,如表 3 所示,本文选取簇  $C_i$  在出现首例感染社区后的 7 日内发生的感染事件的数量  $N(C_i)$  与标准差圆半径  $R(C_i)$ ,这两个指标代理簇  $C_i$  的近邻扩散风险。标准差圆半径  $R$  描述了空间点相对于空间平均中心的空间偏离情况,即

$$R = \sqrt{\sum_i^n \frac{(x_i - x_{mc})^2 + (y_i - y_{mc})^2}{n - 2}}$$

(8)

式中:  $x_i$ 、 $y_i$  为事件  $i$  的二维平面坐标;  $x_{mc}$ 、 $y_{mc}$  为事件集空间分布的平均中心的二维平面坐标。

另外,簇  $C_i$  出现首例感染社区的时间越早,说明该簇发生迁移扩散的风险越高。因此,选取簇  $C_i$  出现首例感染社区的日期  $G_i$  代理其迁移扩散风险。进一步,以传播源社区为中心、 $D_c$  为半径的热点区域内提取簇  $C_i$  的 7 类城市设施的 POI 数据作为空间风险因子的代理变量。考虑区域内社区数量与人口数量也是影响疫情扩散的重要因素,因此,进一步引入热点区域内的社区数量  $L(C_i)$  和居民户数  $F(C_i)$  作为控制变量。

使用多元线性回归方法,分析表示扩散风险的

表 3 变量描述

Tab. 3 The variable description

变量	含义
$N(C_i)$	簇 $C_i$ 在其首个 7 日间窗口内感染社区的数量
$R(C_i)$	簇 $C_i$ 在其首个 7 日间窗口内感染社区分布的标准差圆半径
$G(C_i)$	簇 $C_i$ 出现首例感染社区的日期序号
$B(C_i)$	整形,簇 $C_i$ 热点区域内的公交站点数量
$S(C_i)$	整形,簇 $C_i$ 热点区域内的地铁站点数量
$K(C_i)$	整形,簇 $C_i$ 热点区域内的餐饮点数量(单位: $10^3$ 个)
$P(C_i)$	整形,簇 $C_i$ 热点区域内的购物点数量(单位: $10^3$ 个)
$A(C_i)$	布尔型,簇 $C_i$ 热点区域内有无大型商业中心,有为 1,否则为 0
$V(C_i)$	布尔型,簇 $C_i$ 热点区域内有无大型超市,有为 1,否则为 0
$H(C_i)$	布尔型,簇 $C_i$ 热点区域内有无三级甲等医院,有为 1,否则为 0
$L(C_i)$	整形,簇 $C_i$ 热点区域内的社区数量(单位: $10^3$ 个)
$F(C_i)$	整形,簇 $C_i$ 热点区域内的居民户数(单位: 万户)

因变量与空间风险因子自变量的关系。建立以  $N(C_i)$ 、 $R(C_i)$  和  $G(C_i)$  为因变量的 3 个模型,3 个模型均以 7 个空间风险因子作为自变量。为控制社区数量与人口数量对扩散风险的影响,3 个模型都引入  $L(C_i)$ 、 $F(C_i)$  作为控制变量。另外,为控制各个簇首例感染日期对近邻扩散的影响,在模型 1、2 中引入  $G(C_i)$  作为控制变量。为检验不同参数组合聚类后扩散风险与空间风险因子关系的变化,分别选取  $D_c=0.8\text{ km}$ ,  $\alpha=0.3$ ;  $D_c=1.5\text{ km}$ ,  $\alpha=0.3$ ;  $D_c=2.5\text{ km}$ ,  $\alpha=0.4$  这 3 组参数组合聚类后的簇进行回归分析。其中,模型 1、2 的观测值剔除了聚类后时间窗口不足 7 天的簇,模型 3 的观测值保留了聚类后所有的簇。回归分析结果如表 4 所示。

当参数组合  $D_c=2.5$ ,  $\alpha=0.4$  时,在模型 1、2 中,表征近邻扩散风险的 2 个因变量与公交站点数

表 4 回归分析结果

Tab. 4 The regression results

变量	$D_c = 2.5, \alpha = 0.4$			$D_c = 1.5, \alpha = 0.3$			$D_c = 0.8, \alpha = 0.3$		
	模型 1 $N(C_i)$	模型 2 $R(C_i)$	模型 3 $G(C_i)$	模型 1 $N(C_i)$	模型 2 $R(C_i)$	模型 3 $G(C_i)$	模型 1 $N(C_i)$	模型 2 $R(C_i)$	模型 3 $G(C_i)$
常数项	-0.81	-0.22	18.9***	0.68	-0.04	18.2***	-0.06	-0.02	18.0***
$B(C_i)$	0.03*	0.005**	-0.03	0.03*	0.005**	-0.07	0.02	0.003	-0.06
$S(C_i)$	0.14	0.01	-0.46**	-0.008	0.01	-0.89***	0.22	-0.002	-1.0**
$K(C_i)$	-1.4	-0.03	1.3	-0.2	-0.08	1.4	0.01	0.1	2.6
$P(C_i)$	3.3***	0.3**	-2.1	0.8*	0.2**	-1.8	0.4	-0.1	-4.3
$A(C_i)$	-1.5	-0.01	-2.8**	-1.0	-0.02	-0.6	0.2	0.01	-0.38
$V(C_i)$	0.7	-0.07	-0.5	0.6	-0.02	-1.8*	0.1	0.04	-1.3*
$H(C_i)$	1.8*	0.30**	-0.7	-0.6	-0.07	1.8	-0.76	-0.06	0.12
$L(C_i)$	-4.1	-0.3	8.1	-3.8	-0.6	5.0	3.8	0.2	-0.5
$F(C_i)$	0.02	-0.015	-0.57**	0.28*	0.024*	-0.73**	0.53**	0.05***	-2.0***
$G(C_i)$	0.11	0.024**		0.08*	0.008		0.11***	0.01***	
$R^2(\%)$	37.7	33.5	34.1	11.8	18.5	22.8	8.2	13.4	13.8
观察值	129	129	169	245	245	308	469	469	643

注: \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

量  $B$ 、购物点数量  $P$  和是否存在三甲医院  $H$  这 3 个空间风险因子显著正相关。即在控制社区数、人口户数与首例感染日期的情况下,热点区域内存在三甲医院,且公交站和购物点数量越多的簇,在首个 7 日时间窗口内感染的社区数越多,扩散范围也越大。在模型 3 中,反映迁移扩散风险的感染日期  $G$  与地铁站数量  $S$ 、是否存在大型商业中心  $A$  这 2 个空间风险因子显著负相关,说明热点区域内存在大型商业中心,且地铁站越多的簇,发生首例社区感染的日期也越早。

当参数组合  $D_c = 1.5, \alpha = 0.3$  时,在模型 1、2 中,2 个近邻扩散风险指标仍然与公交站点数量  $B$ 、购物点数量  $P$  显著正相关,但与是否存在三甲医院  $H$  相关性不再显著。在模型 3 中,感染日期  $G$  仍然与地铁站数量  $S$  显著负相关,但与是否存在大型商业中心的相关性不再显著,而与是否存在大型超市显著负相关。

当参数组合  $D_c = 0.8, \alpha = 0.3$  时,在模型 1、2 中,近邻扩散风险指标与所有空间风险因子的相关关系均不显著。在模型 3 中,感染日期  $G$  仍然与地铁站数量  $S$  显著负相关,同时与是否存在大型超市  $V$  显著负相关。

综上所述,近期研究分别以广州、重庆和武汉为例,在  $1\text{ km} \times 1\text{ km}$  网格区域或街道乡镇尺度上,通过多元回归或空间回归方法分析发现,一轮新冠疫情结束时,感染者数量与区域内人口数量、交通站点、医院、购物服务、餐饮服务等空间风险因子显著正相关<sup>[3,12-13]</sup>。本文从近邻扩散与迁移扩散两个方面,在不同距离阈值下,分析上海市新冠疫情扩散风险与热点区域内的空间风险因子分布的相关关系与前期文献有所不同。

在近邻扩散风险方面,在  $2.5\text{ km}$  尺度上,公交站数量、购物点数量和有无三甲医院这 3 个空间风险因子表现出较高显著性。在  $1.5\text{ km}$  尺度上,公交站数量与购物点数量这 2 个空间风险因子表现出较高显著性。但在  $0.8\text{ km}$  尺度上,没有空间风险因子表现出显著性。这说明,由近距离的出行接触而引起的疫情传播,与周边功能设施的分布情况并无显著关系。可能的原因是,在较小的距离尺度上,社区居民近距离、高频次的生活出行接触是造成疫情扩散的主要原因,而近距离、高频次的出行接触在空间分布上并未表现出显著的规律性。但在  $2.5\text{ km}$  和  $1.5\text{ km}$  尺度上,疫情的近邻扩散风险与空间风险因子表现出显著的相关性。可能的原因是,在较大的距离尺度上,近邻社区居民间较远距

离、低频次的生活出行接触是造成疫情扩散的重要原因,而这类生活出行围绕三甲医院、公交站点、购物点等功能设施表现出显著的聚集性。

在迁移扩散风险方面,地铁站数量在  $2.5、1.5、0.8\text{ km}$  这 3 个尺度上始终表现出较高的显著性,有无大型商业中心在  $2.5\text{ km}$  尺度上表现出较高显著性,有无大型超市在  $1.5\text{ km}$  和  $0.8\text{ km}$  这 2 个尺度上表现出较高的显著性。这说明,围绕地铁站、大型商业中心与大型超市的远距离的工作与消费出行是引起疫情迁移扩散的重要原因,地铁站与大型商业中心相较于大型超市其辐射距离更远。

## 5 结论

本文以传染病疫情的迁移扩散与近邻扩散机制为理论依据,提出时空加权密度峰值聚类模型(STW-DPC)。在城市社区尺度上,应用该模型对 2022 年 3 月上海疫情数据进行聚类分析,从而识别局部传播源社区并划分蔓延社区。进一步探究了不同参数组合下疫情防控的效果,实证分析了城市设施对扩散风险的影响。聚类分析表明:

(1) 通过 STW-DPC 模型,能够有效识别局部区域疫情扩散的传播源。传播源局部密度较大的簇,表现出更早的社区感染与更大规模的近邻扩散(见图 6)。

(2) 对上海疫情数据集聚类后发现,当距离阈值参数  $D_c$  较小时,能够相对容易地控制疫情在簇内的近邻扩散,但难以控制簇间的迁移扩散。随着距离阈值的增加,近邻扩散可控性呈减小趋势,但迁移扩散可控性大幅增加(见图 5)。这能够有效解释该轮上海疫情初期的点式防控策略难以有效控制疫情扩散的原因。

(3) 在城市社区尺度上,基于社区居民的生活圈出行距离,在  $0.8、1.5、2.5\text{ km}$  这 3 个距离尺度上,制定区域防控策略能较好地协调近邻扩散与迁移扩散风险(见图 2、图 5)。

(4) 在 3 个距离尺度上,时间衰减系数  $\alpha$  在  $0.3 \sim 0.4$  之间取值时,近邻扩散可控性  $S(C)$  达到最大值。这可能与该轮上海疫情暴发初期局部区域新增感染社区的指数增长系数有关。因此,可通过局部区域新增感染社区的指数增长情况来调整时间衰减系统  $\alpha$ ,从而提升近邻扩散的防控效果。

(5) 在  $2.5\text{ km}$  尺度上聚类发现,该轮上海疫情的迁移扩散路径并非完全随机的,而是表现出以 0、1 号感染社区为中心向周边的轴辐式扩散模式(见图 7)。

扩散风险影响因素的实证分析表明:

(1) 当传染病疫情在局部区域出现近邻扩散时,热点区域内公交站数量、购物点数量和是否存在三甲医院这 3 个空间风险因子可作为区域防控决策的辅助因素。

(2) 当传染病疫情在城市空间出现远距离迁移扩散时,可通过地铁站数量、有无大型商业中心和有无大型超市这 3 个风险因子辅助预判疫情的迁移扩散路径。

本文的研究为大型城市在社区尺度上应对突发传染病的疫情防控,拓展了一种可供借鉴的方法。依据分析结果,提出如下建议:

(1) 在城市社区尺度上,疫情传播表现出近邻扩散与迁移扩散特征,采取区域防控策略能够有效抑制这两类扩散。

(2) 在疫情暴发初期,应用 STW-DPC 模型并结合地铁站、大型商业中心、大型超市等空间风险因子的分布情况,有助于识别局部区域的传播源社区,提升近邻扩散防控效果。

(3) 以传播源社区为中心,基于社区居民的生活圈出行特征,在 1~3 倍生活圈半径尺度上划定区域防控范围,能够较好地协调近邻扩散与迁移扩散风险。公交站点、购物点、三甲医院等空间风险因子的分布情况,能够辅助划定防控范围的决策。

本文的研究方法与结论并未考虑城市外部病例的输入,以及静态管理后疫情在近邻小区间的气溶胶传播。一方面,外部输入病例的生活和出行模式与社区居民并不一致,因此,以社区为研究对象的传播机制可能并不适用于这类人群;另一方面,静态管理后,传染病疫情仍可能以气溶胶传播的模式在近邻社区间扩散传播,这使得近邻扩散的影响因素还可能与社区的感染密度以及社区的建成环境相关。本文仅在城市社区尺度上探讨了疫情的扩散机制与城市设施间的相关关系,尚不足以解释上述两个问题,有待在后续研究中进一步收集颗粒度更细的多源时空数据,从而进行多源数据的融合分析和相互验证。

参考文献:

[ 1 ] 赵宏波,魏甲晨,王爽,等. 大城市新冠肺炎疫情风险评估与精准防控对策——以郑州市为例[J]. 经济地理, 2020, 40(4): 103-109.

ZHAO H B, WEI J C, WANG S, *et al.* The risk assessment of Covid-2019 epidemic in metropolis and precise prevention and control measures: A case

study of Zhengzhou city [J]. *Economic Geography*, 2020, 40(4): 103-109.

[ 2 ] 刘郑倩,叶玉瑶,张虹鸥,等. 珠海市新型冠状病毒肺炎聚集发生的时空特征及传播路径[J]. 热带地理, 2020, 40(3): 422-432.

LIU Z Q, YE Y Y, ZHANG H O, *et al.* Spatio-temporal characteristics and transmission path of COVID-19 cluster cases in Zhuhai [J]. *Tropical Geography*, 2020, 40(3): 422-432.

[ 3 ] 黄晓慧,付迎春,张君怡,等. 广州市新冠疫情早期时空扩散特征与公共空间风险探究[J]. 热带地貌, 2020(2): 8-17.

HUANG X H, FU Y C, ZHANG J Y, *et al.* Spatial and temporal spreading characteristics and public space risk in the early stage of COVID-19 outbreak in Guangzhou[J]. *Tropical Geography*, 2020(2): 8-17.

[ 4 ] 荆磊,刘明皓,陈春,等. 基于 SEIRD-GEOCA 的 COVID-19 疫情时空分布模型构建与情景模拟[J]. 西南大学学报(自然科学版), 2022, 44(2): 207-218.

JING L, LIU M H, CHEN C, *et al.* Spatio-temporal distribution model construction and scenario simulation of COVID-19 epidemic using SEIRD-GEOCA [J]. *Journal of Southwest University (Natural Science Edition)*, 2022, 44(2): 207-218.

[ 5 ] 王姣娥,杜德林,魏冶,等. 新冠肺炎疫情的空间扩散过程与模式研究[J]. 地理研究, 2020, 39(7): 1450-1462.

WANG J E, DU D L, WEI Z, *et al.* The development of COVID-19 in China: Spatial diffusion and geographical pattern[J]. *Geographical Research*, 2020, 39(7): 1450-1462.

[ 6 ] 赵宝静,奚文沁,吴秋晴,等. 塑造韧性社区共同体: 生活圈的规划思考与策略[J]. 上海城市规划, 2020(2): 14-19.

ZHAO B J, XI W Q, WU Q Q, *et al.* Shaping a Resilient Community: Planning Methods for the Community Life Circle[J]. *Shanghai Urban Planning Review*, 2020(2): 14-19.

[ 7 ] BAI Z. Threshold dynamics of a periodic SIR model with delay in an infected compartment [J]. *Mathematical Biosciences & Engineering*, 2015, 12(3): 555-564.

[ 8 ] YANG Z, ZENG Z, WANG K, *et al.* Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions [J]. *Journal of Thoracic Disease*, 2020, 12(3): 165-174.

[ 9 ] 曹盛力,冯沛华,时朋朋. 修正 SEIR 传染病动力学模型应用于湖北省 2019 冠状病毒病(COVID-19)疫



情预测和评估[J]. 浙江大学学报(医学版), 2020, 49(2): 178-184.

CAO S L, FENG P H, SHI P P, *et al.* Study on the epidemic development of COVID-19 in Hubei province by a modified SEIR model[J]. Journal of Zhejiang University (Medical Sciences), 2020, 49(2): 178-184.

[10] HUANG R, LIU M, DING Y. Spatial-temporal distribution of COVID-19 in China and its prediction: A data-driven modeling analysis[J]. The Journal of Infection in Developing Countries, 2020, 14(3): 246-253.

[11] SPASSIANI I, SEBASTIANI G, PALU G. Spatiotemporal analysis of COVID-19 incidence data [J]. Viruses, 2021, 13(3): 463-475.

[12] 陈晓, 黄宇金, 李佳慧, 等. COVID-19 疫情时空聚集性特征及影响因素分析——以重庆市为例[J]. 地理科学进展, 2020, 39(11): 1798-1808.

CHEN X, HUANG Y J, LI J H, *et al.* Clustering characteristics of COVID-19 cases and influencing factors in Chongqing Municipality [J]. Progress in Geography, 2020, 39(11): 1798-1808.

[13] 李欣, 周林, 贾涛, 等. “新冠”疫情下的城市空间风险因素解析——以武汉为例[J]. 城市规划, 2021, 45(8): 78-86.

LI X, ZHOU L, JIA T, *et al.* Decoding the impact of potential urban risk factors on the COVID-19 situation: A case study of Wuhan[J]. City Planning Review, 2021, 45(8): 78-86.

[14] MEDREK M, PASTUSZAK Z. Numerical simulation of the novel coronavirus spreading [J]. Expert Systems with Applications, 2021, 166: 114109.

[15] GIULIANI D, DICKSON M M, ESPA G, *et al.* Modelling and predicting the spatio-temporal spread of COVID-19 in Italy[J]. BMC Infectious Diseases, 2020, 20(1): 1-10.

[16] HU Y, XU L, PAN H, *et al.* Transmission center and driving factors of hand, foot, and mouth disease in China: A combined analysis[J]. PLoS neglected Tropical Diseases, 2020, 14(3): e0008070.

[17] BACKER J A, WALLINGA J. Spatiotemporal analysis of the 2014 Ebola epidemic in West Africa [J]. PLoS Computational Biology, 2016, 12(12): e1005210.

[18] ALKHAMIS M A, ALY S, KHAJAH M M, *et al.* Spatiotemporal dynamics of the COVID-19 pandemic in the State of Kuwait [J]. International Journal of Infectious Diseases, 2020, 98: 153-160.

[19] KIM S, CASTRO M C. Spatiotemporal pattern of COVID-19 and government response in South Korea [J]. International Journal of Infectious Diseases, 2020, 98: 328-333.

[20] YE L, HU L. Spatiotemporal distribution and trend of COVID-19 in the Yangtze River Delta region of the People’s Republic of China [J]. Geospatial Health, 2020, 15(1): 25-32.

[21] 吕振华, 程绍文. 基于 Crystal Ball 和 GIS 的湖北省新冠疫情时空特征研究[J]. 华中师范大学学报(自然科学版), 2020, 54(6): 1059-1071.

LV Z H, CHENG S W. Research on the temporal and spatial characteristics of the COVID-19 in Hubei Province with the use of Crystal Ball and GIS [J]. Journal of Central China Normal University (Natural Sciences), 2020, 54(6): 1059-1071.

[22] 刘勇, 杨东阳, 董冠鹏, 等. 河南省新冠肺炎疫情时空扩散特征与人口流动风险评估——基于 1243 例病例报告的分析[J]. 经济地理, 2020, 40(3): 24-32.

LIU Y, YANG D Y, DONG G P, *et al.* The spatio-temporal spread characteristics of 2019 novel coronavirus pneumonia and risk assessment based on population movement in Henan province: Analysis of 1243 individual case reports [J]. Economic Geography, 2020, 40(3): 24-32.

[23] 中华人民共和国自然资源部. 社区生活圈规划技术指南 [EB/OL]. (2021-06-09) [2022-08-01]. [https://www.mnr.gov.cn/zt/zh/gtkjgh/jsbz/bzjd/202111/t20211129\\_2708429.html](https://www.mnr.gov.cn/zt/zh/gtkjgh/jsbz/bzjd/202111/t20211129_2708429.html). Ministry of Natural Resources of the People’s Republic of China. Community life circle planning technical guide [EB/OL]. (2021-06-09) [2022-08-01]. [https://www.mnr.gov.cn/zt/zh/gtkjgh/jsbz/bzjd/202111/t20211129\\_2708429.html](https://www.mnr.gov.cn/zt/zh/gtkjgh/jsbz/bzjd/202111/t20211129_2708429.html).

[24] 上海市规划和自然资源局. 上海市 15 分钟社区生活圈规划导则 [EB/OL]. (2016-09-02) [2022-08-01]. <https://ghzyj.sh.gov.cn/ldft/20210715/cc2adc4ffff244418d2a99fd6d7756ad.html>. Shanghai Municipal Bureau of Planning and Natural Resources. Shanghai planning guidance of 15-minute community-life circle [EB/OL]. (2016-09-02) [2022-08-01]. <https://ghzyj.sh.gov.cn/ldft/20210715/cc2adc4ffff244418d2a99fd6d7756ad.html>.

[25] 郑天铭, 刘合林. 影响社区新冠肺炎疫情的建成环境要素探究——以武汉市为例[J]. 现代城市研究, 2020, 35(10): 20-29.

ZHENG T M, LIU H L. Exploration of the built-environmental elements that influence the spread of covid-19 pandemic on community scale: A case study

of Wuhan, China [J]. Modern Urban Research, 2020, 35(10): 20-29.

[26] RODRIGUEZ A, LAIO A. Clustering by fast search and find of density peaks [J]. Science, 2014, 344 (6191): 1492-1496.

[27] DODDS P S, WATTS D J. Universal behavior in a generalized model of contagion [J]. Physical Review Letters, 2004, 92(21): 218701.

[28] SUN J, STAAB S, KARIMI F. Decay of relevance in exponentially growing networks [C]//Proceedings of the 10th ACM Conference on Web Science. New York, United States: Association for Computing Machinery, 2018: 343-351.

[29] PEI T, ZHOU C, ZHU A X, *et al.* Windowed nearest neighbour method for mining spatio-temporal clusters in the presence of noise [J]. International Journal of Geographical Information Science, 2010, 24(6): 925-948.

[30] 唐建波, 邓敏, 刘启亮. 时空事件聚类分析方法研究 [J]. 地理信息世界, 2013(1): 38-45.

TANG J B, DENG M, LIU Q L. On spatio-temporal events clustering methods [J]. Journal of Spatio-temporal Information, 2013(1): 3-458.

[31] 周开乐, 杨善林, 丁帅, 等. 聚类有效性研究综述 [J]. 系统工程理论与实践, 2014, 34 (9): 2417-2431.

ZHOU K L, YANG S L, DING S, *et al.* On cluster validation [J]. Systems Engineering-Theory & Practice, 2014, 34(9): 2417-2431.

[32] 刘美霖. 基于深度时空网络的犯罪预测技术研究 [D]. 北京: 中国人民公安大学, 2020.

LIU M L. Research on crime prediction technology based on deep spatio-temporal networks [D]. Beijing: People's Public Security University of China, 2020.

# Spatiotemporal Spreading Characteristics and Influence Factors of COVID-19 in Urban Community Level: Evidence from Shanghai

*LI Zhouping<sup>1</sup>, GE Ruyi<sup>1\*</sup>, GUO Xiaoshuang<sup>2</sup>*

(1. Faculty of Business Information, Shanghai Business School, Shanghai 200235, China;  
2. Sino-British College, University of Shanghai for Science and Technology, Shanghai 200093, China)

**【Abstract】** The outbreak of the COVID-19 epidemic in Shanghai in March 2022 lasted for three months, with the cumulative number of infected people reported to be more than 620,000. At the urban scale, the epidemic in Shanghai exhibited high randomness and complexity, making it difficult to track and predict. However, in communities and surrounding living spaces, where residents are in close contact, the spread of COVID-19 showed significant characteristics of adjacent diffusion. Therefore, we first divide the entire urban pandemic area into sections, identify local transmission sources on a more microscopic community scale, define the spread boundaries, and explore the local spread patterns.

Based on the mechanisms of adjacent diffusion and relocation diffusion, we propose a spatiotemporal weighted density peak clustering model (STW-DPC), which effectively identifies local diffusion sources and delineates diffusion boundaries. Using the STW-DPC model, we review the spread of the COVID-19 epidemic in Shanghai from March to June 2022, analyzing regional spread characteristics and assessing the effectiveness of prevention and control measures. Finally, based on this review, we explore the correlation between the regional spread characteristics of the epidemic and the spatial distribution of urban facilities. This allows decision-makers to predict the local epidemic spread trends based on the STW-DPC model and the spatial distribution of urban infrastructure.

In the first part, we construct a new clustering model (STW-DPC) based on the density peak clustering (DPC) algorithm. The model uses the distance threshold parameter from the DPC algorithm to represent the community scale and introduces a time decay coefficient to account for the strength of the contagion relationship between new and prior infection events. From the perspective of epidemic prevention and control, we propose two indicators: the controllability of adjacent diffusion and the controllability of relocation diffusion, to evaluate the effectiveness of the STW-DPC model.

In the second part, we use the STW-DPC model to review the spread of the COVID-19 epidemic in Shanghai. The results show that developing regional prevention and control strategies at the scales of 0.8 km, 1.5 km, and 2.5 km can effectively control both adjacent diffusion and relocation diffusion. At the 2.5 km scale, we observe that the temporal and spatial characteristics of relocation diffusion are not entirely random but exhibit a hub-and-spoke spread pattern, moving from near to far around two transmission source communities.

In the third part, based on the review data, we apply a multiple regression model to explore the correlation between regional epidemic spread characteristics and the spatial distribution of urban facilities. Regarding adjacent diffusion risk, we find that the three spatial risk factors—bus stops, shopping spots, and large-scale hospitals—are significant at the 2.5 km scale. At the 1.5 km scale, bus stops and shopping spots show high significance. However, at the 0.8 km scale, no spatial risk factors show significance.



Regarding relocation diffusion risk, metro stations consistently show high significance across the three scales of 2.5 km, 1.5 km, and 0.8 km. Shopping malls show high significance at the 2.5 km scale, while supermarkets show high significance at the 1.5 km and 0.8 km scales.

In summary, this paper confirms that developing regional prevention and control strategies at the three scales of 0.8 km, 1.5 km, and 2.5 km can effectively control the spread of COVID-19 in Shanghai. The spatial risk factors of bus stops, shopping spots, and large-scale hospitals can assist in formulating regional prevention and control strategies, while the spatial risk factors of metro stations, shopping malls, and supermarkets can help predict the migration paths of the epidemic.

**Keywords:** pandemic prevention and control; density peak clustering; urban community; spatiotemporal diffusion