



论文

疾病传播输入输出流的时空特征分析——以北京 SARS 流行为例

胡碧松^{①②*}, 龚建华^{②③†}, 周洁萍^{②③}, 孙麋^②, 杨荔阳^②, 夏宇^①,
IBRAHIM Abdoul Nasser^{②③}

① 江西师范大学地理与环境学院鄱阳湖湿地与流域研究教育部重点实验室, 南昌 330022;

② 中国科学院遥感应用研究所遥感科学国家重点实验室, 北京 100101;

③ 浙江中科空间信息技术应用研发中心, 嘉兴 314100

* 联系人, E-mail: hubisong624@126.com

† 同等贡献, E-mail: jhgong@irsa.ac.cn

收稿日期: 2012-02-27; 接受日期: 2012-08-27; 网络版发表日期: 2013-08-09

国家自然科学基金(批准号: 40871181)、国家科技重大专项课题(编号: 2012ZX10003-005)、中国科学院知识创新工程重要方向项目(编号: KZCX2-EW-318)、国家自然科学基金青年基金(批准号: 41101369)、江西省自然科学基金(编号: 20114BAB215024)、江西省教育厅青年科学基金项目(编号: GJJ11073)和鄱阳湖湿地与流域研究教育部重点实验室开放基金(编号: PK2010001)资助

摘要 为更好揭示传染病流行的个体受感染-发病-就诊-康复的时空过程变化模式以及区域之间的信息流与物质流的传播网络结构特征, 基于 2002~2003 年中国 SARS 病例数据, 选取 SARS 流行的三个典型空间位置信息: 工作单位或住址、发病地点以及报告单位, 从感染个体和空间区域两个角度来探索病毒感染扩展的输入与输出传播机制, 定义并探讨了北京市 SARS 传播输入-输出流(In-Out Flow)的概念与流行病学特征. 采用空间统计分析方法和网络特征分析方法, 探索了北京市外部输入-输出流的高风险时空热点以及网络结构特征, 并对北京市内部输入-输出流传播网络的空间自相关性与异质性、时空演化规律以及网络结构特征进行了全面分析. 结果表明: (1) 外部输入流集中在山西和广东, 而外部输出流较为分散, 主要是广东以及山东等北部省份, 且防控措施重点应分别是 SARS 爆发的早中期与初期; (2) 内部输出病例在整体区域上存在显著正自相关性特征, 高危人群集中在 20~60 岁的医务与民工等中青年人群; (3) 市中心区的若干区县是 SARS 传播高风险热点区域, 西北方位市郊是次级高风险区域, 而东北方位市郊则相对安全; (4) 在内部传播网络中各区县节点存在显著小世界特征且具有信息流和物质流的异质性特征, 市郊通州区与昌平区是潜在的高风险热点区域, 顺义区与怀柔区等节点承载了极低的物质流和信息流信息, 是相对的低风险冷点区域. 基于输入-输出流的探索与分析更有助于揭示 SARS 流行过程中的潜在时空演化特征与规律, 可为应急决策与防控措施提供更有理论依据.

关键词输入-输出流
SARS
北京市
传播网络
时空特征
控制措施

中文引用格式: 胡碧松, 龚建华, 周洁萍, 等. 疾病传播输入输出流的时空特征分析——以北京 SARS 流行为例. 中国科学: 地球科学, 2013, 43: 1499–1517
英文引用格式: Hu B S, Gong J H, Zhou J P, et al. Spatial-temporal characteristics of epidemic spread in-out flow—Using SARS epidemic in Beijing as a case study. Science China: Earth Sciences, 2013, 56: 1380–1397, doi: 10.1007/s11430-012-4479-z

各类纷繁复杂的传染病伴随着人类的发展和演化, 在发展历史上人类曾多次面对过各种危害极其严重的已知或未知传染病的威胁, 近年来影响比较大的传染病爆发包括 2003 年春天的非典型性肺炎 (Severe Acute Respiratory Syndrome, SARS) 及 2009 年 4 月在全球范围内蔓延的甲型 H1N1 流感. SARS 全称严重急性呼吸系统综合症, 是一种因感染冠状病毒而导致的新型呼吸道传染病, 以发热、干咳、胸闷、呼吸系统衰竭等为主要症状, 具有极强的感染传播性, 其传播途径主要是近距离飞沫传播、接触患者的呼吸道分泌物及密切空间接触传播. 据国家卫生部统计, 中国大陆出现 SARS 疫情的省份共有 26 个, 就每日疫情与累计报告临床诊断与疑似病例情况来看, SARS 病例的数量分布存在着显著的地区差异. 确诊病例与疑似病例都主要集中在北京、广东、山西、内蒙古、天津、河北等省份, 其中尤以广东和北京为重. 多数省份的首次病例是由广东、北京和香港等地区输入的, 主要是在上述几个地区学习、务工、出差、旅游的返乡人员. SARS 疫情的突发流行是 21 世纪人类面临的第一次公共卫生全球性流行事件, 近年来大量研究学者从病理学、流行病学、卫生统计学和空间信息科学等各个领域对 SARS 的病毒特征、传播机制、时空分布和防控措施等方面进行了深入全面的研究. SARS 的传播过程是新型传染性病毒与资源、环境、社会、经济等多因素复杂综合作用的结果, 对 SARS 时空传播规律的定量研究可以深入认识 SARS 在时空尺度上的潜在传播机制, 解释各类影响因素对 SARS 流行的驱动作用, 评价 SARS 疫情在区域和人群上的感染风险, 为 SARS 传播的防控应急措施提供理论支持, 同时也是未来未知新型突发传染性疾病的应急措施和分析研究的科学依据.

目前针对 SARS 疫情时空传播规律的研究主要可以分为以下几类: (1) 传播过程数学模型的理论建模, 主要是基于经典确定性 SIR 模型或随机性模型对 SARS 过程进行理论建模^[1,2], 计算基本再生数 R_0 和有效再生数 R_t ^[3], 分析疾病传播过程的动力学机制^[4,5]等. (2) 流行病学数据的数理统计分析, 例如采用经典数理统计方法对 SARS 疫情的时序病例、密切接触者与控制措施等进行定量分析^[6,7]. (3) 流行病学数据的时空分布特征与空间统计分析, 包括对 SARS 病例在时空尺度上分布特征的探索分析^[8], 利用空间统计方法对 SARS 疫情的空间自相关性和异质性特征进

行分析^[9-11], 以及采用多种时空统计分析方法相结合的元模型来探索 SARS 疫情的空间模式、时序变化与驱动因素^[12]. (4) 基于网络动力学模型的疾病传播过程的模拟与预测, 例如利用系统动力学 (System Dynamics) 和多智能体系统 (Multi-Agent System) 对 SARS 传播过程进行模拟分析^[13,14], 基于小世界网络 (Small-World Network) 或无标度网络 (Scale-Free Network) 来模拟 SARS 病毒的传播机制^[15,16]. (5) 影响因素的定量分析与评价, 主要是针对 SARS 疫情传播的密切接触、人口、社会经济等各类影响因素进行定量分析与等级评价^[17-21]. 从研究尺度来看, 上述 (1)~(3) 类研究都属于从宏观角度研究统计意义上的疾病传播规律与结果; (4) 类研究是从微观角度研究个体与区域之间的传播特性, 反映疾病在宏观区域内的传播结果; (5) 类研究是从宏观角度探索影响因素对疾病传播的驱动作用.

上述 SARS 疫情时空传播规律的各类研究方法有着以下三点不足之处: (1) 一般时空分析方法是将 SARS 疫情的时间信息和空间信息各自分离进行分析, 侧重点仅是时间与空间两个维度的叠加, 其研究趋势应是对时间、空间、疾病因子与传播影响因素等多维数据集的协同分析^[12]; (2) 疫情数据主要以确诊感染者或密切接触者为主, 缺乏对个体在受感染-发病-就诊-康复传播过程中的空间位置时序变化的关注与解读; (3) 在 SARS 传播过程中, 个体之间的病毒扩散机制以及区域之间的疾病传播交互模式, 将形成一个动态的有向扩散网络, 已有研究缺少对 SARS 传播网络的结构特征解释与定量分析. 另外, 北京市是中国 SARS 流行期间疫情最严重的省级区域, 且其应急处置和防控措施具有及时性和高效性特点, 再者北京作为中国的首都, 与世界各地和全国各省均有着密切交互, 以北京作为 SARS 时空传播规律的研究区域具有显著的示范特征与现实意义.

基于以上考虑, 本文以 2002~2003 年期间中国 SARS 流行的感染病例数据为基础, 定义并解释了北京市 SARS 传播输入-输出流 (In-Out Flow) 的概念与流行病学特征, 采用空间统计分析方法和网络特征分析方法, 对北京市 SARS 传播的输入-输出流的流行病学特征进行解释, 探讨了北京市与其他省级区域之间的疾病传播的空间自相关性与异质性以及时空演化规律, 并对北京市内部的疾病传播网络进行了全面的时空演化分析与网络特征分析. 与基于单

一感染病例或密切接触者的时空分析方法相比, 针对输入-输出流的探索与分析, 更有助于揭示 SARS 流行过程中的潜在时空演化特征与规律, 可为应急决策与防控措施提供更有理论依据。

1 数据处理与修复

1.1 时空数据处理原则

原始实验数据来源于军事医学科学院微生物流行病研究所, 主要依据 SARS 感染者的个体病例属性, 发病时间范围从 2002 年 11 月 16 日至 2003 年 5 月 10 日, 共计 5497 例确诊感染者数据(注: 此为 SARS 流行末期的流行病学调查数据, 与最终精确统计数字有着极少量误差)。流行病学病例属性包括性别、年龄、职业、户籍、工作单位或住址、发病地点、报告单位、发病时间、入院时间、确诊时间等信息内容。本文研究的 SARS 传播输入-输出流主要针对感染病例的常住、发病与就诊三个空间位置信息进行特征分析, 因而选取了至少有一个空间位置信息包含了“北京市”的病例数据, 其中工作单位或住址、发病地点、报告单位分别涉及“北京市”的原始数据共计 873 例、1972 例与 1874 例。图 1 分别表示三类空间位置信息涉及“北京市”的原始病例数据的时序变化情况, 由此可知, 三类位置信息对应 SARS 感染人数的整体变化趋

势较为一致, 均在 3 月初进入稳定增长阶段, 中期快速增长至 4 月中下旬达到峰值, 然后后期至 5 月上中旬迅速下降直至完全消失。其中发病地点和报告单位对应的变化趋势极其一致, 表明 SARS 传播期间北京市内部的感染病例基本保持发病地与就诊地的一致性, 而工作单位或住址对应的变化趋势有所差异, 表征的是常住地为北京的感染者人数在全国各地的变化, 故总体少于前两者的新增病例人数变化, 且在 SARS 流行前期就已经有零散的新增病例。

原始 SARS 流行病学病例数据是以数据表形式进行存储的, 其中时间信息是以 TimeDate 格式进行描述, 可以直接用于定量分析, 而三个空间位置信息均是以普通 Text 文本格式存储, 并且存在记录尺度不统一等问题。因而, 本文在进行 SARS 传播时空数据处理时, 遵循了以下三个原则: (1) 将三类空间位置信息转换为统一空间尺度(省/直辖市/自治区), 同时将涉及“北京市”的空间位置信息转换为统一区/县尺度; (2) 剔除原始三类空间位置信息中至少有两项是缺失值或无法解译值的病例数据, 也即保证选用的数据至少包含完整的两项空间位置信息; (3) 选取三类空间位置信息涉及“北京市”的病例数据。采取感染病例的发病时间作为 SARS 时空数据的时间信息, 其空间信息涉及三类空间位置信息, 依据上述三条空间化处理原则, 采用人工校正匹配方法将三类

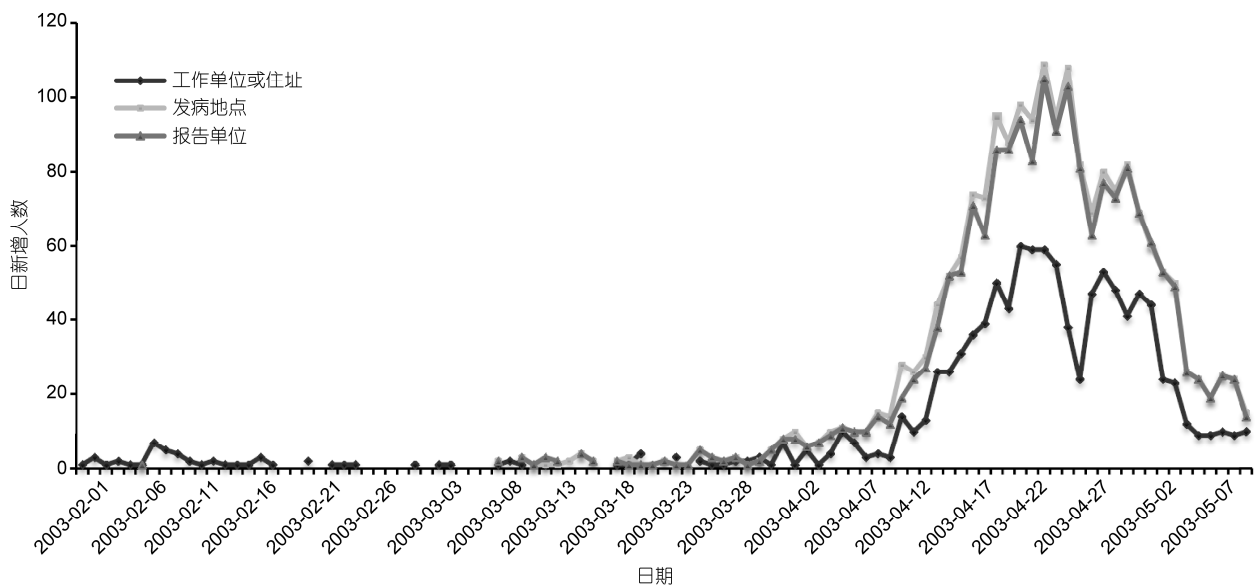


图 1 2003 年北京市 SARS 感染病例日新增人数变化曲线(居住地/发病地/就诊地)

空间位置信息与全国行政区划地图和北京市行政区划地图进行地名匹配,剔除至少有两项无效空间信息之后,得到全国范围内 5483 例数据,进一步选取涉及“北京市”的病例数据共计 2037 例。

1.2 缺失数据修复方案

全国范围内初步有效的 5483 例 SARS 病例数据中,时间信息是完整有效的,而三类空间信息中,除完备的发病地点信息之外,工作单位或住址以及报告单位信息分别有着 2831 例和 69 例缺失或语义不详情况,此外其他性别、年龄、职业等属性也存在少量数据缺失情况。因而在后续分析之前,先对缺失数据进行修复以保证数据完整性。数据修复方案主要基于两方面考虑,一是首要保证各行政单元各类属性数据的分布一致性,二是附加考虑各行政单元人口分布的相对影响作用。

针对存在缺失情况的某一属性数据(如工作单位或住址),假设某地区 i 的已有属性数据总数为 n_i , 其人口数量为 p_i , 则所有地区的已有属性数据以及人口数量的总和分别为 $N = \sum_i n_i$ 与 $P = \sum_i p_i$ 。那么对于地区 i , 其已有属性数据和人口数量相对比率分别为

$$\rho_i = n_i / N = n_i / \sum_i n_i, \quad (1)$$

$$\rho_{i, pop} = p_i / P = p_i / \sum_i p_i. \quad (2)$$

设置一个调节参数 α , 用以表示已有属性数据以及人口数量之间的相对调整系数,则对应于地区 i 的数据修复概率与累计修复概率分别为

$$\theta_i = (1 - \alpha) \times \rho_i + \alpha \times \rho_{i, pop}, \quad (3)$$

$$\Theta_i = \sum_1^i \theta_i. \quad (4)$$

对于某个具体缺失属性数据,给定一个 $[0, 1]$ 之间的随机数 $rand$, 判定 $rand$ 位于的具体区间 $[\Theta_{i-1}, \Theta_i]$, 则将 Θ_i 对应的位置信息赋予该缺失属性数据。

基于保持原始数据的空间分布特征为主的考虑,设置调节参数 $\alpha = 0.2$, 采用上述修复方案,对原始缺失属性数据进行修复。北京数据的修复结果比较理想,对于工作单位或住址与报告单位的缺失数据修复相似率达到 99.68% 和 99.77%, 主要由于整体缺失数据比例较少且各地区之间的数据分布差异较少,而对于全国数据,报告单位的缺失数据修复相似率达到 99.99%, 因为缺失数据比例极小,而工作单位

或住址的缺失数据比例较大,但修复相似率仍达到 98.22%。

2 输入-输出流的定义与解释

在传染病传播特征规律与时空模拟的研究中,研究人员往往关注的是传染病的某一特定空间位置信息(如感染者发病地点),然而健康人群从受到病毒感染到出现症状再到就诊的过程中,会出现潜伏期和就诊期的阶段,也将会对应于不同的空间信息(如受感染地点 $\rightarrow$ 发病地点 $\rightarrow$ 就诊地点)。输入-输出流(In-Out Flow)主要针对传染病受感染者在疾病流行期间的空间信息变化过程,探索不同尺度下的区域单元之间的疾病传播输入-输出的扩散蔓延机制。下面以北京市 SARS 传播为例,介绍 In-Out Flow 的定义与解释。

选取 SARS 流行的三个典型空间位置信息: 工作单位或住址(S1)、发病地点(S2)以及报告单位(S3)。其中, S1 可以认为是受感染者的主要常驻地点,因为人群的两个最重要地点信息即为家庭居住地和工作单位; S2 近似表示人群的感染地点,因为一般传染病潜伏期较短且被感染的时间和地点信息较难获取,因而采用发病地点来近似代替受感染地点; S3 表示受感染者的就诊地点,一般也即由医疗单位进行向上报告。SARS 传播 In-Out Flow 可以针对不同区域在不同空间尺度上描述病毒感染扩展的传播机制。从感染个体角度来看,病毒扩展源自健康者与感染者之间的空间接触(飞沫等呼吸道分泌物接触),个体的常驻活动地点延伸至受病毒感染地点,再到出现症状的发病地点,再到就诊治疗的报告地点;从空间区域角度来看,某区域与其他外部区域之间以及其内部各子区域之间,均存在着病毒输入和输出的扩散过程。

针对北京市 SARS 传播过程,定义其 In-Out Flow 包括外部流与内部流两种类型。其中,外部流是指北京市与其他同等尺度的区域(省/直辖市/自治区)之间的输入-输出,而内部流是指北京市各区县之间的输入-输出。在外部流中,输入(Input)是指非北京市常住人口选择在北京市进行就诊治疗的病例,进而依据感染者的发病地点是否在北京,又可细分为一类输入(Primary Input)和二类输入(Secondary Input); 输出(Output)是指北京市常住人口选择在非

北京市区域进行就诊医疗的病例, 类似依据感染者发病地点是否在北京, 也可细分为一类输出(Primary Output)和二类输出(Secondary Output). 在内部流中, 感染病例的工作单位或住址、发病地点与报告地点均位于北京市, 其输入与输出定义方式保持一致, 只是区域对象是北京市内的各区县(如朝阳区). 北京市 SARS 传播的 In-Out Flow 可用下列逻辑表达式进行描述:

$$\begin{aligned} \text{Outer Input: } S1 \neq \text{'Beijing'} \text{ AND } S3 \\ = \text{'Beijing'} \left\{ \begin{array}{l} \text{Primary: } S2 = \text{'Beijing'} \\ \text{Secondary: } S2 \neq \text{'Beijing'} \end{array} \right\} \quad (5) \end{aligned}$$

$$\begin{aligned} \text{Outer Output: } S1 = \text{'Beijing'} \text{ AND } S3 \\ \neq \text{'Beijing'} \left\{ \begin{array}{l} \text{Primary: } S2 \neq \text{'Beijing'} \\ \text{Secondary: } S2 = \text{'Beijing'} \end{array} \right\} \quad (6) \end{aligned}$$

$$\text{Inner Input: } S1 \neq \text{'Chaoyang'} \text{ AND } S3 = \text{'Chaoyang'}, \quad (7)$$

$$\text{Inner Output: } S1 = \text{'Chaoyang'} \text{ AND } S3 \neq \text{'Chaoyang'}. \quad (8)$$

在上述 In-Out Flow 的逻辑定义下, 北京市在 SARS 传播过程中, 一方面与其他各省直辖市之间发生病例感染-就诊的输入与输出, 另一方面内部各区

县之间也存在病例感染-就诊的交互过程. 如图 2 所示, 在时间序列上, 外部流是以北京市为中心的输入-输出流, 而内部流是北京市各区县之间形成了输入-输出流的有向加权网络.

SARS 传播 In-Out Flow 侧重于疾病扩散过程中的人群常住、发病与就诊等空间位置信息, 直接反映了病毒传播过程的特征机制. 从流行病学角度来看, 外部流反映的是北京市与其他省直辖市之间的病例感染-就诊的交互过程, 内部流则是北京市各区县之间的输入-输出机制; 输入流反映了非北京常住人口受感染后在北京就诊的情况, 输出流则表征了北京常住人口在其他区域就诊的情况. 根据各类型内部和外部的输入与输出的定义, 可进行相应流行病学解释如表 1 所示.

3 流行病学分析

北京市于 2003 年 1 月 14 日和 2002 年 12 月 1 日分别出现第一例外部输入和输出病例, 截至 5 月 10 日, 总外部输入病例共计 1024 例, 一类和二类

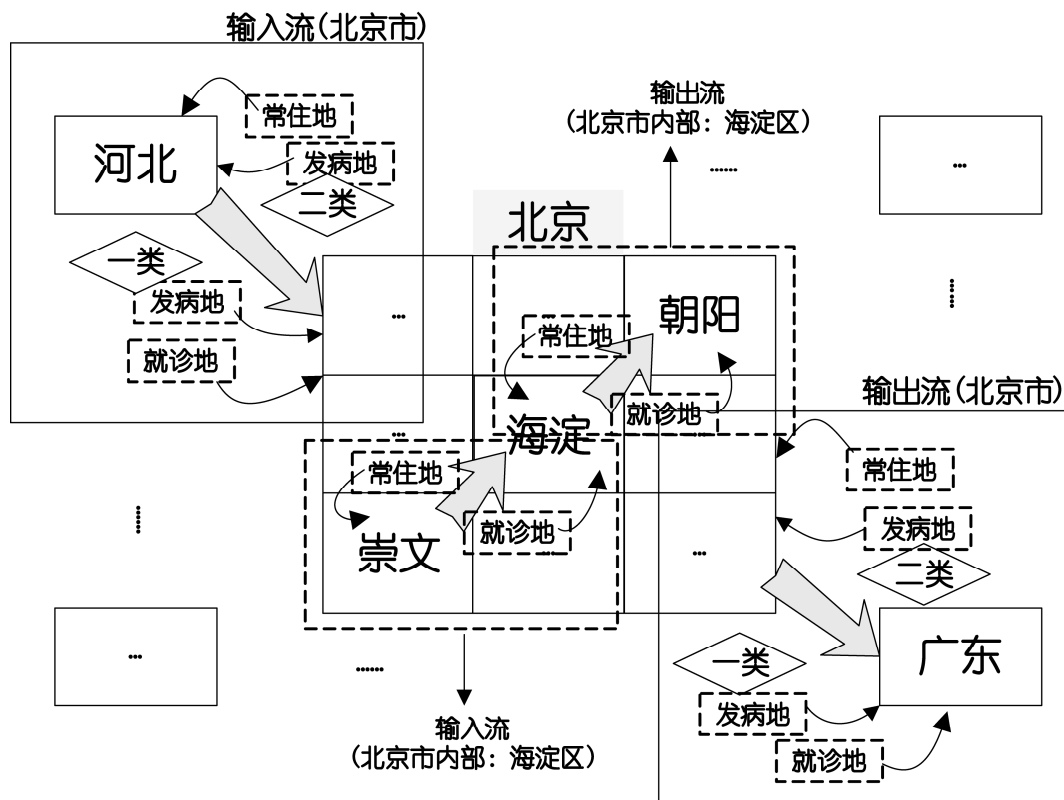


图 2 北京市 SARS 传播输入-输出流的定义框架

表1 北京市 SARS 传播输入-输出流的流行病学解释

类型		特征	流行病学解释
外部	一类输入	非北京常住人口, 在北京地区受到感染, 并在北京进行就诊治疗	受 SARS 病毒感染的人群来自其他省市, 但由于出差旅游等原因在北京受到感染, 并且就近选择在北京市进行治疗. 此类输入病例直接受到地区发达程度与交通便利行等因素干扰, 是从外部控制北京 SARS 流行的主要对象
	二类输入	非北京常住人口, 在非北京地区受到感染, 但在北京进行就诊治疗	受感染者常住地与受感染地均位于其他省市, 但基于医疗设施等因素的考虑, 选择在北京进行治疗. 病毒传播可能会持续整个空间转移过程, 此类输入病例是从外部流动输入人群进行北京 SARS 预防控制的重点对象
	一类输出	北京常住人口, 在非北京地区受到感染, 并在非北京地区进行就诊治疗	受感染者是北京常住人口, 但由于出差旅游等在外地受到感染并选择在外地进行就诊, 病毒扩散得到及时遏止. 但对于北京, 应作为在 SARS 流行期间外出人群的主要控制对象, 是病毒向外部扩散的控制措施的主要依据
	二类输出	北京常住人口, 在北京地区受到感染, 但在非北京地区进行就诊治疗	受感染者是北京常住人口, 在北京受到感染, 但由于某些原因选择在外地进行就诊. 病毒传播将持续整个以北京为起始的空间转移过程, 此类输出病例是从内部流动输出人群进行北京 SARS 预防控制的重点对象
内部	输入	北京常住人口在北京发病并接受治疗. 在某区县进行就诊治疗, 但是在其他区县受到感染	疾病扩散过程完全位于北京内部. 输入病例表征了某区县的病毒传播汇聚程度, 是从流动输入人群进行北京内部 SARS 预防控制的重点对象
	输出	北京常住人口在北京发病并接受治疗. 在某区县受到感染, 但是在其他区县进行就诊治疗	疾病扩散过程完全位于北京内部. 输出病例表征了某区县的病毒传播扩散程度, 是从流动输出人群进行北京内部 SARS 预防控制的重点对象

输入病例分别为 1012 例和 12 例. 总外部输出病例共计 237 例, 一类和二类输出病例分别为 191 例和 46 例. 内部病例共计 853 例, 首例出现在 SARS 流行开发爆发的 3 月下旬. 从各类病例的空间分布特征(图 3(a))来看, 北京外部输入病例主要来自广东(121 例)与北京周围的若干省份, 主要是河北(52 例)、山东(69 例)、山西(70 例)、河南(68 例)等省份, 其他中部省份较为均匀, 而东北与西北省份较少; 北京外部输出病例较多集中在广东(90 例), 其次是河北(36 例)、山西(32 例)与内蒙古(17 例), 而其他省份均仅有零星输出病例(小于 10 例). 从各类病例的时序变化(图 3(b))来看, 北京外部输入病例从 3 月初开始出现, 在 4 月初进入爆发阶段直至 4 月下旬达到日增病例数的高峰, 随之进入衰减直至消失; 内部病例与外部输入病例基本保持一致的变化趋势, 说明北京 SARS 流行期间外部输入和内部传播病例的贡献度基本持平, 而内部病例在 4 月下旬有着短期的骤减趋势, 可能由于隔离控制措施在初期对于病毒内部扩散有着显著效果; 外部输出病例出现较早, 且变化趋势基本保持平稳, 即使在 4 月的严重流行期间, 输出病例的日增人数也基本保持在 10 例左右. 基于以上分析可得出几点启示: (1) 北京 SARS 流行主要由外部输入病例和内部病例所组成, 前者主要来自与北京周围省份以及广东; (2) 北京外部输出病例整体较少, 且大多集中在广东、河北、山西与内蒙古; (3) 外部输入输出的二类病例所占比例分别为 1.17%和 19.4%, 表明北京输

入病例大多是在北京当地发病并就诊治疗, 而输出病例的部分感染者在发病至就诊过程中发生了空间位置转移, 整体病例人数虽较少, 但在空间传播过程中却比输入病例带来的感染风险要更高; (4) 北京内部的隔离控制措施对于内部病例的感染扩散有着一定效果, 但针对外部输入与输出病例的防控还需更为大规模的严厉措施.

通过对北京 SARS 的外部输入输出病例的性别、年龄段与职业的构成特征分析(图 4), 男性与女性输入输出病例基本持平, 构成比分别为 53%/47%和 56%/44%, 男性人群的受感染风险略高; 输入与输出病例的年龄集中在 20~50 岁和 20~40 岁, 平均年龄分别为 38.09 与 35.47 岁; 输入病例的职业属性较多集中在家政/家务/服务人员、离退休人员、公务/商务/职员以及医疗卫生行业人员, 占比例达 47.36%; 输出病例的职业属性主要集中在民工/农民/工人、家政/家务/服务人员以及教师/学生/保育人员, 占比例达 54.43%. 从人口流动特点来看, 以青壮年为主的外来务工人员成为北京输入病例的主要来源, 其主要就业范围集中在建筑业/服务业/餐饮业等, 其次由离退休人员所组成的外来旅游人员以及因防控措施而调度的医务人员也占一定比例, 再者以中年为主的外来商务人员也是输入病例的主要构成因素, 造成输入病例的平均年龄偏高; 输出病例的组成除常规的流动人群之外, 还包括了一定程度的以学生为主的教育行业人员.

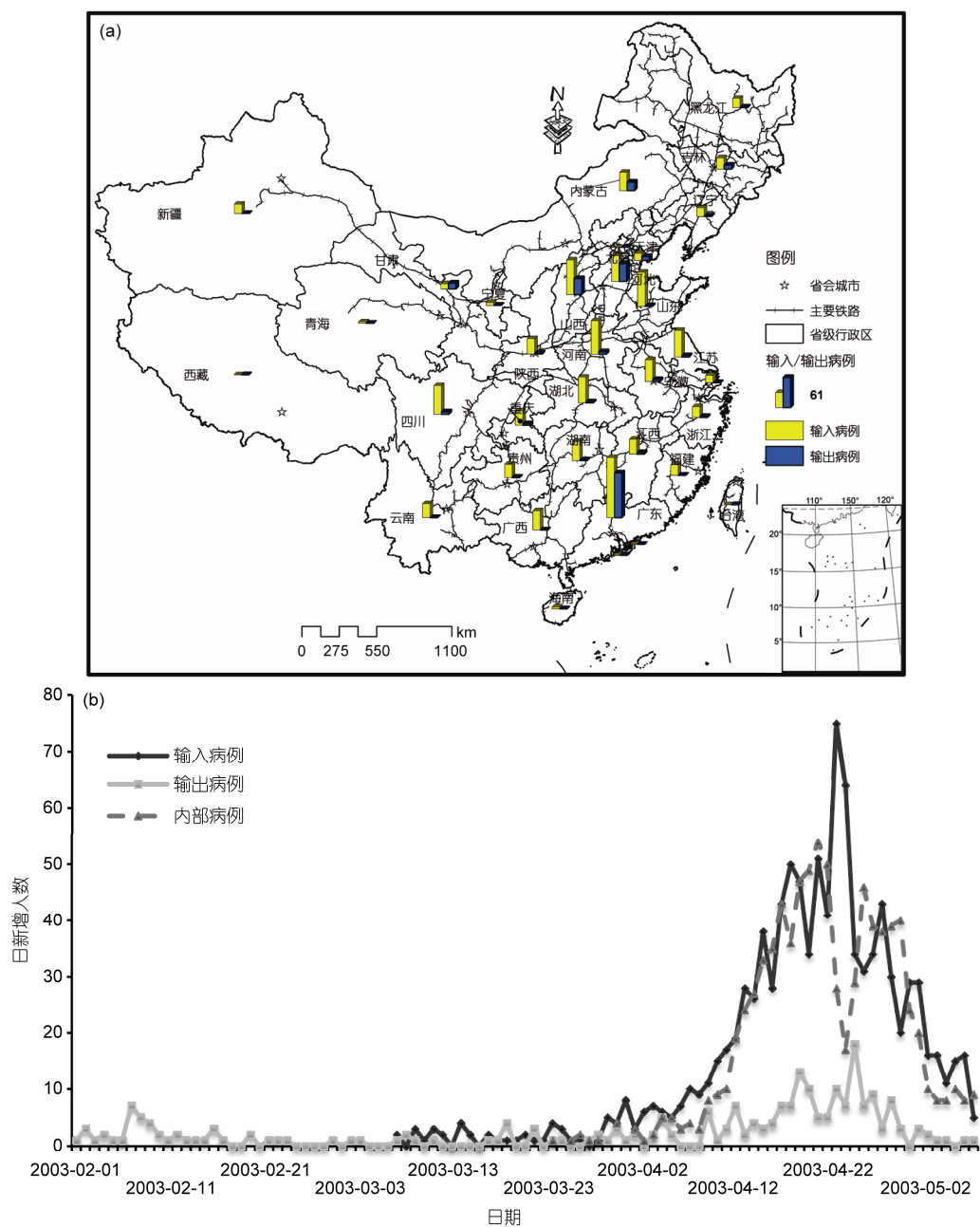


图 3 北京 SARS 输入输出病例的时空模式
(a) 外部输入输出病例的空间分布; (b) 外部输入输出病例与内部病例的时序变化

4 SARS 传播 In-Out Flow 的时空特征分析

4.1 空间自相关性分析方法与网络特征分析方法

现代空间统计分析方法源自于 Tobler 地理学第一定律, 即“地理空间中要素或现象都是关联相似的, 且邻近的要比远距离的区域或现象更相似”^[22], 主要

核心是揭示与空间位置相关的区域化变量的空间依赖、空间关联与空间自相关特征. 空间统计分析的关键是解释地理学第一定律中的两个核心概念, 即“关联”与“邻近度”, 一般意义上的“关联”是指变量在空间上的分布具备相似性, 而“邻近度”是指区域具备邻接边或依据单元中心的距离, 若考虑增加时间维

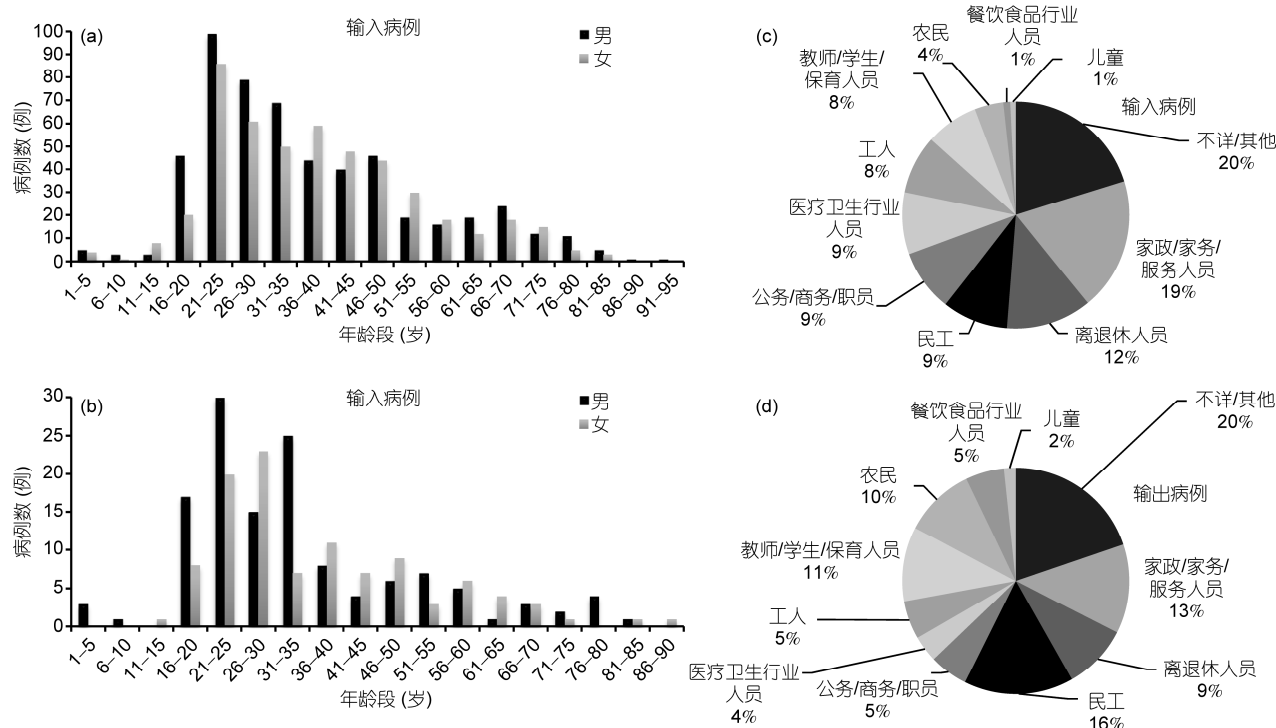


图4 北京 SARS 输入输出病例的属性构成特征

(a) 输入病例/性别/年龄; (b) 输出病例/性别/年龄; (c) 输入病例/职业; (d) 输出病例/职业

度, 还可将传统的“空间邻近度”扩展为“时空邻近度”^[23]. 目前空间统计分析方法中用以评价地理要素全局空间自相关性的指标主要包括 Moran's I^[24], Geary's C^[25]与 Getis's G^[26]等, 评价局部空间自相关性的指标主要包括 Local Moran's I, Local Geary's C, LISA(Local Indicator of Spatial Association)^[27], 以及 Getis's G^[28]与 Moran Scatterplot^[29]等. 在空间自相关性分析基础上, 热点探测分析可以寻找研究区内时空尺度上的异常聚集子区域, 揭示疾病时空分布的高风险区, 主要有 Spatial Scan Statistics^[35], Nearest Neighbor Hierarchical Clustering^[30], **Geographical Detector**^[31]等方法.

北京市 SARS 传播输入-输出流在不同空间尺度下可以描述北京与各区域之间的疾病传播演化机制. 在外部流中, 各省直辖市之间形成了以北京为中心的有向集聚网络; 在内部流中, 北京各区县之间形成了交互式有向网络. 输入流与输出流表征了感染病例的流向以及病例数的转移量, 承载了信息流与物质流的双重转移信息, 形成了一个复杂的有向加权网络. 与传统规则网络和随机网络相比, 复杂网络能

更为准确地描述真实世界中的各类网络特征, 尤其自小世界网络模型^[32]和无标度网络模型^[33]的理论提出, 复杂网络的研究更进入了白热化的阶段. 一般而言, 描述复杂网络的拓扑结构特征变量包括度分布 (Degree Distribution)、特征路径长度 (Characteristic Path Length)、聚类系数 (Clustering Coefficient)、介数 (Betweenness Centrality) 等.

本文选取全局与局部 Moran's I 指数、Moran 散点图、扫描统计方法 (Scan Statistics) 以及网络度分布与聚类系数指标, 定量分析北京市 SARS 传播输入-输出流的时空特征.

4.1.1 Moran's I 指数与 Moran 散点图

全局 Moran's I 指标用于评价区域化变量在整体地理空间上的自相关性特征, 其计算公式如下^[24]:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad (9)$$

其中, n 是研究区域所有子区域的数量, x_i 与 x_j 分别为

区域 i 和 j 的变量属性值, $\bar{x}$ 为整体研究区域变量属性值的均值, w_{ij} 表示区域 i 和 j 的空间临近关系, 一般基于空间拓扑关系或区域中心点距离权值来衡量, 如区域 i 和 j 存在共同的边界线或其中心点距离小于某阈值 d 时, 则认为区域 i 和 j 是邻接的, 其 $w_{ij}=1$, 反之不邻接, 其 $w_{ij}=0$. 也有其他临近关系的定义规则, 如依据人口规模、医疗规模、城市结构等属性排序进行临近关系的定义^[12].

空间关联的局部指标 LISA 可以揭示区域化变量的局部异常聚集状况与局部不稳定性特征, 局部 Moran's I 指标是典型的 LISA, 其计算公式为^[27]

$$I_i = \frac{n(x_i - \bar{x}) \sum_{j=1}^n w_{ij}(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{nz_i \sum_{j=1}^n w_{ij} z_j}{z^T z}, \quad (10)$$

其中, x_i , x_j 与 w_{ij} 的定义与式(9)相同, $z_i = x_i - \bar{x}$, $z_j = x_j - \bar{x}$, $z^T = (z_1, z_2, \dots, z_n)$.

w_{ij} 组成空间邻接矩阵 W , 以 (Wz, z) 为坐标点即可绘出 Moran 散点图^[29]. 横纵坐标轴将散点图分割成

四个象限, 代表了局部区域的不同空间关联形式: (1) 高-高聚集(H-H); (2) 低-高聚集(L-H); (3) 低-低聚集(L-L); (4) 高-低聚集(H-L).

4.1.2 空间/时空扫描统计方法

空间/时空扫描统计是由 Kulldorff 提出的热点 (Hotspot) 探测方法, 可针对疾病流行等时空数据进行快速有效的统计分析与热点探测. 其基本原理是采用对数似然比 LLR(Log Likelihood Ratio) 作为疾病暴发的评价指标, 建立零假设和备择假设, 依次以每个区域质心为圆心, 生成一系列半径由零到规定上限逐渐变化的同心圆, 计算每个圆形区域内的患病概率的 LLR, 取最大若干值所在的圆形区域为暴发热点备选区域, 采用蒙特卡罗方法对热点备选区域依次进行显著性检验, 通过检验的便是探测到的热点区域^[35,36]. 空间扫描统计的 LLR 计算方法一般依据泊松模型(Poisson)、贝努利模型(Bernoulli)、序数模型(Ordinal)、指数模型(Exponential)等, 例如在贝努利模型中 LLR 的计算方法如下^[35,36]:

$$LLR = \frac{L(\hat{Z})}{L_0} = \frac{\left(\frac{n_Z}{\mu(Z)}\right)^{n_Z} \left(1 - \frac{n_Z}{\mu(Z)}\right)^{\mu(Z)-n_Z} \left(\frac{n_G - n_Z}{\mu(G) - \mu(Z)}\right)^{n_G - n_Z} \left(1 - \frac{n_G - n_Z}{\mu(G) - \mu(Z)}\right)^{(\mu(G) - \mu(Z)) - (n_G - n_Z)}}{\left(\frac{n_G}{\mu(G)}\right)^{n_G} \left(\frac{\mu(G) - n_G}{\mu(G)}\right)^{\mu(G) - n_G}}, \quad (11)$$

其中, Z 是一个由圆心坐标和圆半径坐标组成的三维向量, n_Z 与 n_G 分别是扫描圆形区域和总研究区域的实测发病数, $\mu(Z)$ 与 $\mu(G)$ 是扫描圆形区域和总研究区域的人口数.

时空扫描统计的基本方法是时空重排(Space-time permutation)模型^[37], 其基本原理与空间扫描统计类似, 所不同的是, 时空重排模型中采用的是扫描圆柱, 圆柱的底面是扫描圆, 圆柱的高代表时间序列. 对于整个研究区域, 假定区域 z 在第 d 天的病例数观测值是 C_{zd} , 总病例数观测值为 C , 扫描圆柱 A 的病例数观测值是 C_A . 圆柱 A 基于无效假设下的病例数期望值为

$$\mu_A = \sum_{(z,d) \in A} \mu_{zd}. \quad (12)$$

当病例数在空间 z 上的总和 $\sum_{z \in A} C_{zd}$ 和与在时间 d 上的总和 $\sum_{d \in A} C_{zd}$ 均小于总病例数的观测值 C 时, 扫描圆柱 A 的病例数观测值 C_A 近似服从均值为 μ_A 的 Poisson 分布. 采用 Poisson 广义似然比 GLR(Generalized

Likelihood Ratio) 作为圆柱 A 疾病暴发的评价指标. GLR 的计算如下^[37]:

$$GLR = \left(\frac{C_A}{\mu_A}\right)^{C_A} \left(\frac{C - C_A}{C - \mu_A}\right)^{(C - C_A)}. \quad (13)$$

4.1.3 网络度分布与聚类系数

节点的度描述的是与其相连接的边数, 有向网络中节点的度又可分为出度(out-degree)和入度(in-degree). 网络度分布 $p(k)$ 表示度数为 k 的节点数目占总节点数目的比例, 或表示在网络中任意选出一个度数为 k 的节点的概率. 度分布是对网络进行分类的重要依据指标, 根据特定的度分布可以将网络进行不同的分类^[34]. 节点 i 的聚类系数 C_i 定义为与其相连接的 k_i 个节点之间实际的边数 A_i 与最多边数之比^[32]:

$$C_i = \frac{A_i}{k_i(k_i - 1)/2}. \quad (14)$$

网络的聚类系数 C 用以描述网络中与同一节点直接相连的节点之间的连接关系, 也即网络中点与点之间集结成簇的趋势, 一般用所有节点聚类系数的平均值来衡量^[32]:

$$C = \frac{1}{N} \sum_i C_i. \quad (15)$$

对于 SARS 传播输入-输出流形成的有向加权网络, 各区域之间的输入与输出流描述的是病毒扩散过程的空间连接信息, 可用出度和入度的分布函数来表征各区域的病毒扩散范围的风险因素. 在标准聚类系数的计算公式中, 节点之间的连接边仅描述了区域之间的 SARS 传播信息流, 为描述其物质流信息, 对节点聚类系数 C_i 的计算公式进行了扩充调整:

$$C_i^* = \sum_j^{A_i} \text{Cas}_j / \sum_j^{A_N} \text{Cas}_j, \quad (16)$$

其中, A_N 是网络边数总数, Cas_j 为边 j 所承载的物质流量 (输入或输出病例数). 调整后节点聚类系数 C_i^* 描述了与节点 i 相连接的其他节点之间的物质流汇聚程度, 也直接表示了 SARS 传播输入-输出流的小世界聚集性质.

4.2 外部 In-Out Flow 的时空特征分析

4.2.1 时空热点探测分析

采用空间统计分析软件 SatScan(<http://www.satscan.org/>) 计算得出 2003 年北京市 SARS 传播输入输出流的热点区域情况, 分别采用贝努利模型和时空重排模型进行纯空间与时空探测. 为统一输入与输出病例的时间属性, 将分析的时间范围定为 2003 年 1 月 20 日至 2003 年 5 月 10 日, 共计 16 周, 其中输入病例的作用范围仅是 2003 年 3 月 8 日至 2003 年 5 月 10 日 (第 9~16 周). 时空重排模型的时间步长为自然周, 最大空间与时间扫描半径分别为包含 50% 的病例总数以及包含 50% 的传播时长.

由表 2 知, 外部输入病例来源较为集中, 主要热点风险区域依次为山西和广东, 相对风险系数 RR 值分别是 2.73 和 2.02; 外部输出病例的首要热点风险区域是广东, RR 值为 9.23, 次要热点区域包括以内蒙古和青海为中心的北部区域, RR 值分别为 4.88 和 2.12. 如图 5 所示, 北京 SARS 传播的外部防控重点应一方面集中排除隔离由山西和广东迁移的输入病例, 另一方面要控制内部病例与广东以及北部区域的迁移输出. 考虑时空尺度上的异常热点状况, 其结

论稍有不同, 由表 2 和图 6 知, 外部输入病例的主要热点区域仍然是山西, 时间范围是第 9~12 周, 检验统计量达到 3.8387, 而福建、重庆、吉林等次要热点区域的时间范围基本都保持在一周左右; 外部输出病例的主要热点区域仍是广东, 时间范围是第 3~10 周, 检验统计量为 22.6499, 次要热点区域包括以山东与西藏为中心的北部区域, 其检验统计量分别为 10.2125 和 5.1414, 前者时间范围是第 13~16 周, 后者时间范围仅维持一周左右.

由以上分析可得出如下启示: (1) 北京的外部输入-输出流具有典型的差异性, 输入流较为集中, 而输出流的辐射范围较广, 包括广东以及北部若干省份; (2) 输入流的主要高风险区是山西, 但时间范围仅持续约 4 周, 其次是另一个 SARS 疫情重区广东; (3) 输出流的主要高风险区是广东, 时间范围持续达 8 周左右, 其次是以山东和内蒙古为中心的北部区域, 但其时间范围仅持续约 4 周; (4) 北京与外部省份之间的疾病传播防控措施应有所区分与侧重, 在 SARS 病毒扩散初期 (3 月前), 应侧重北京至广东输出病例的隔离控制, 而在病毒爆发的早期 (3 月), 应重点关注山西和广东的输入病例, 而在病毒爆发的中后期 (3 月后), 应侧重北京至山东等省份输出病例的控制. 整体而言, 针对外部输入病例的防控措施重点在 SARS 爆发的早中期, 且集中在山西和广东两地, 而针对外部输出病例的防控措施重点应在 SARS 爆发的初期, 且集中在广东以及山东等北部省份.

4.2.2 网络结构特征分析

针对 2003 年 1 月 20 日至 2003 年 5 月 10 日 (16 周) 期间北京市 SARS 传播外部输入-输出流, 分别计算其输入流与输出流对应的入度与出度分别为 $D_{in}=29$ 和 $D_{out}=21$, 表明北京的 SARS 输入病例涉及省份基本已经覆盖全国, 而输出病例涉及范围较少, 但其各自的风险范围却有所区别, 输入流的度数较高但风险较为集中, 主要是山西和广东两地, 而输出流的度数不高但风险却较为分散, 包括广东以及山东等大部分北部省份. 以周为单位, 分别计算北京的入度与出度的时序变化趋势 (图 7), 可发现: (1) 出度出现较早, 基本在 1 月下旬就存在北京至外地的输入病例, 而入度在 SARS 爆发的初期即 3 月上旬才开始出现; (2) 出度的时序变化较为平缓, 在 SARS 爆发的中后期即 4 月份, 其值基本保持在 5~15, 而在早初

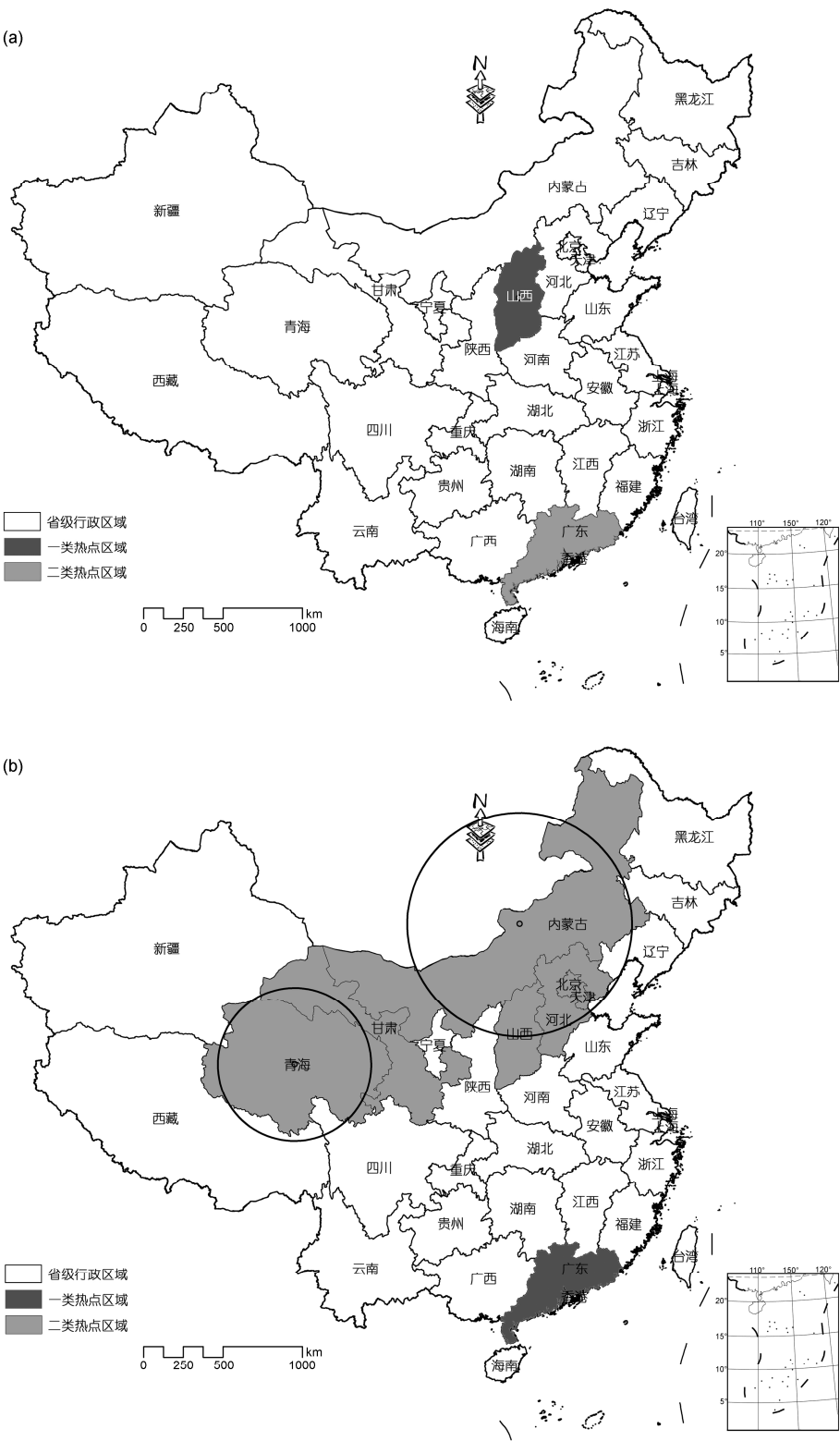


图 5 外部输入-输出流的热点区域空间分布
(a) 输入病例; (b) 输出病例

表 2 外部输入-输出流的热点探测结果

SARS	模型	热点区域								
		热点类型	中心点	包含区域	半径(km)	时间(周)	LLR	P	ODE	RR/检验统计量
输入病例	贝努利模型	一类	山西	-	0	-	25.2469	≈ 0	2.61	2.73
		二类	广东	-	0	-	22.4257	≈ 0	1.90	2.02
	时空重排模型	一类	山西	-	0	9~12		0.5700	2.12	3.8387
		二类	福建	-	0	16~16		0.7600	3.15	3.2718
		二类	重庆	-	0	9~9		0.9800	7.83	2.3732
		二类	吉林	辽宁	393.87	15~15		0.9800	1.80	2.3371
输出病例	贝努利模型	一类	广东	-	0	-	102.4074	≈ 0	6.10	9.23
		二类	内蒙古	河北/天津/山西	737.08	-	58.9547	≈ 0	3.36	4.88
		二类	青海	甘肃	494.22	-	2.5995	0.5790	2.06	2.12
	时空重排模型	一类	广东	香港/澳门/广西/湖南	502.84	3~10		≈ 0	2.55	22.6499
		二类	山东	天津/江苏/河北/河南等	994.43	13~16		0.0001	1.46	10.2125
		二类	西藏	青海/新疆/甘肃	1337.49	16~16		0.1090	8.70	5.1414

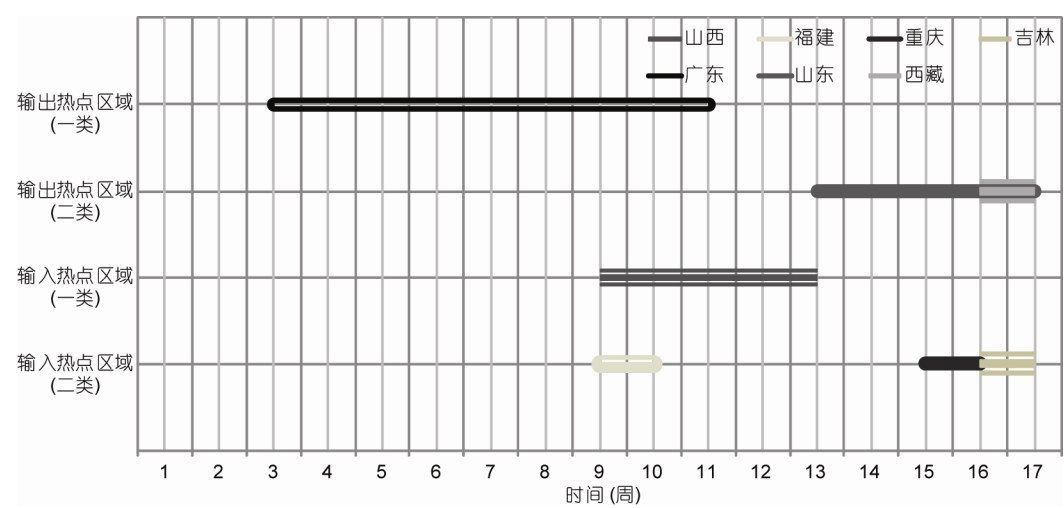


图 6 外部输入-输出流的热点区域时序变化

期其值基本在 5 以下, 表明北京至外地的输出流辐射范围较小; (3) 入度在第 9 周即 3 月初才出现, 但其初值即达到 14 左右, 并且很快扩大到 25~30, 即输入流辐射范围在较短时间内即覆盖全国; (4) 入度的时序变化有着明显的转折点, 在 3 月期间由初值 14 减少为 7 而后又迅速上升至最大值, 表明在 SARS 爆发初期北京针对外部输入病例的防控措施较为有效, 在两周左右的时间有着显著效果, 控制了外部输入流的扩散, 但由于疾病爆发强烈或防控措施失效等可能原因, 输入流迅速扩散至全国范围。

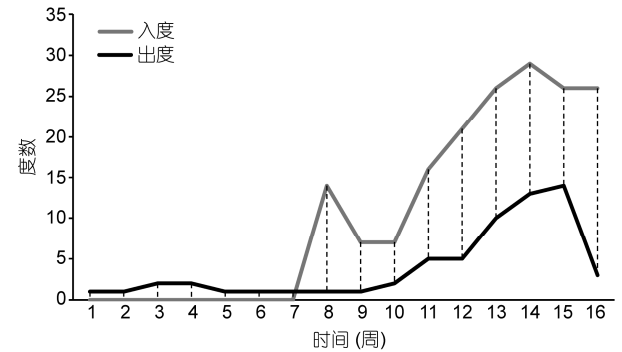


图 7 北京市外部输入-输出流的度分布时序变化

4.3 内部 In-Out Flow 的时空特征分析

4.3.1 全局空间自相关特征分析

针对北京市 SARS 传播的内部输入输出病例数据, 利用空间统计分析软件 OpenGeoDa(<http://geodacenter.asu.edu/ogeoda/>), 基于距离阈值构建北京市各区县的空间邻接矩阵, 计算得出 2003 年北京市内部输入病例与自传播病例(当地发病且当地就诊)数据的全局 Moran's I 指数分别为 0.0477 和 0.0048, 不具备空间自相关性特征, 而内部输出病例数据的全局 Moran's I 指数为 0.2370, 依据零假设 H_0 进行 999 次 Monte Carlo 随机模拟得到 Moran's I 的期望值 $E(I) = -0.0588$, 标准差 $S_d = 0.1387$, 检验统计量 $Z\text{-Score} = (I - E(I)) / S_d = 2.1327$, $P = 0.031 < 0.05$, 表明输出病例在整体区域上存在显著正自相关性特征. 以上可有如下解释: (1) 北京市内部自传播病例较多在当前发病区县里就近进行就诊治疗, 因而不具备显著空间相关性; (2) 内部输入病例受到自传播病例的干扰影响, 丢失了本应具备的空间自相关性; (3) 内部输出病例反映了各区县自身疾病爆发与流行的状况以及发病-就诊的转移模式, 在一定程度上与空间邻近度和各区县的影响因素有关联, 因而表现出显著正相关性.

再进一步考虑内部输出病例的性别、年龄段与职业的影响作用, 以及内部输出病例空间相关性的时序变化(2003 年 3 月 23 日至 2003 年 5 月 10 日, 共 7 周), 同样以全局 Moran's I 指数进行评价计算, 结果

如表 3 与图 8 所示, 可知: (1) 女性病例的空间自相关性特征略强于男性病例, 与整体 SARS 病例的性别分布情况恰好相反, 但都基本保持持平; (2) 年龄处于 20 岁以下或 60 岁以上的病例的空间自相关性特征较弱且不具备显著性, 在 20~60 岁范围的病例中, 随着年龄的增大, 其空间自相关性特征增强; (3) 工人/农民/民工与医疗卫生行业人员两类人群的空间自相关性特征都较强, 后者由于从事医疗服务行业, 受院内感染的影响较多, 而前者属于典型城市流动人群, 其活动范围与模式较为规律, 因而其空间自相关性表现更为显著; (4) 内部输出病例的空间自相关性特征在 SARS 流行的初期表现不明显, 因为此期间新增病例的感染途径较多且隔离控制措施的主要是医院; 而在第 3~6 周(4 月 6 日至 5 月 3 日)的 SARS 流行高峰期, 新增病例急剧增多, 且由于隔离控制措施而主要集中在院内感染和家庭感染, 因而表现出较强的空间自相关性特征且程度基本保持一致; 在 SARS 流行后期, 新增病例急剧减少, 感染途径仍然主要是院内和家庭感染, 因而具备较高的空间自相关性, 但由于病例总数较少导致其特征不具备显著性.

基于上述分析, 北京市 SARS 传播内部防控措施重点应放在发病-就诊模式的内部输出病例上, 尽量减少由于治疗产生的空间转移过程而导致的病毒扩散程度, 无性别差异地对待中青年高风险感染人群的隔离与控制, 同时重点监测并控制工人/农民/民工与医疗卫生行业人员两类人群, 在流行高峰期应对

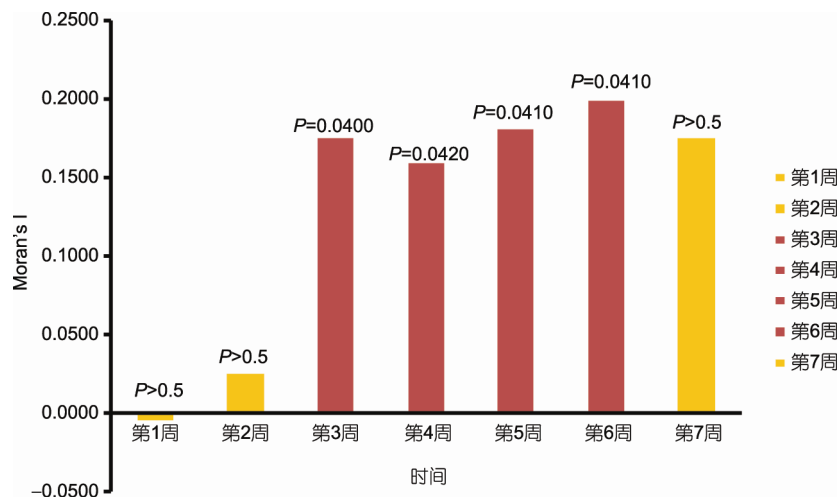


图 8 内部输出病例全局 Moran's I 的时序计算结果

表 3 不同类型内部输出病例的全局 Moran's I 计算结果^{a)}

内部输出病例		Moran's I	Monte Carlo 模拟结果(999 次)		
			标准差	Z 统计量	P
性别	男	0.2253*	0.1450	1.9593	0.0380
	女	0.2264*	0.1467	1.9441	0.0340
年龄	(, 20]	0.1450	0.1273	1.6009	0.0610
	(20, 40]	0.1758*	0.1399	1.6769	0.0450
	(40, 60]	0.2172*	0.1460	1.8904	0.0340
	(60,)	0.1891	0.1433	1.7299	0.0560
职业	医疗卫生行业人员	0.2084*	0.1492	1.7909	0.0460
	家政/家务/服务人员	0.0766	0.1298	1.0431	0.1170
	工人/农民/民工	0.2258*	0.1401	2.0314	0.0360
	离退休人员	0.0486	0.1413	0.7601	0.1870

a) *表示空间自相关性是显著的($P<0.05$), $E(I)=-0.0588$

医院和家庭两类主要感染途径采取重点防控措施.

4.3.2 局部空间自相关特征分析

与全局空间自相关特征分析相对应, 针对 2003 年北京市 SARS 传播内部输出病例, 采用局部 Moran's I 这一典型 LISA 指标以及 Moran 散点图来探测其内部传播的时空演变特征. 利用空间统计分析软件 OpenGeoDa 进行内部传播病例的风险区域分析, 再根据具备显著全局空间自相关性的若干属性(性别/年龄/职业), 分别探索其局部自相关性特征与风险等级. 从所有内部输出病例的 Moran 散点图(图 9(a))和风险类型分布图(图 9(b))可知, 北京内部输出病例的高风险区域主要集中在城市中心区域, 包括海淀区、朝阳区、东城区、西城区与丰台区等, 而位于城市中心的崇文区、宣武区、石景山区以及位于西北方位昌平区等属于次风险区域, 相对安全的低风险区是位于市郊东北方位的密云县, 其他区域由于病例数较少等可能原因而不具备局部空间自相关性. 内部输出病例的高风险与次风险区域皆在城市中心区形成了显著集聚状况, 仅向西北方位的少部分市郊区域产生的病毒高危扩散, 而东北方位的市郊区域属于相对安全的低风险区, 说明北京 SARS 传播整体防控措施较为显著, 在保证控制高风险传播局限于城市中心区域的同时, 应加强对西北方位若干市郊区域的防控, 而考虑资源有限的前提下可适度降低对东北方位市郊区域的防控措施程度.

从图 9(c)和(d)可知, 女性病例的风险区域分布要强于男性病例, 这与性别分布的全局空间自相关性特征保持一致, 两者的高风险区域分布基本一致, 而男性病例的风险区域分布相对更为集中. 从图 9(e)和(f)可知, 20~40 岁人群的高风险区域集中在市中心两三个行政区且基本没有风险扩散的趋势, 而 40~60 岁人群的高风险区域基本覆盖了市中心大部分行政区, 且对西北方位和东南方位的市郊区域有着明显的高危扩散趋势, 说明针对青年人群的隔离控制措施效果相对较为显著, 而中年人群的空间热点分布基本与整体保持一致, 是 SARS 流行病毒传播的主体. 从图 9(g)和(h)可知, 医疗卫生行业人员的高风险区域集中在市中心少数区域且仅向东南方位有较弱程度的扩散趋势, 说明针对院内感染的控制措施非常显著, 而工人/农民/民工人群的热点区域主要集中在市中心大部分区域, 且有着固定模式的扩散趋势, 属于 SARS 传播的高风险人群.

考虑北京市具备显著局部空间自相关性特征的若干区县在 SARS 流行期间的风险类型演化机制, 由表 4 可知: (1) 市中心区的海淀区、朝阳区与东城区是 SARS 传播高风险热点区域, 且在整个流行期间, 仅在最初 2~3 周内表现相对安全, 之后持续保持高风险传播状况; (2) 整体表现为高风险的东城区与丰台区, 在风险等级上经历了低-高-低的变化过程, 且其高风险持续阶段位于流行中后期; (3) 市中心区的崇文区与宣武区基本持续表现为低-高区域, 表明其受

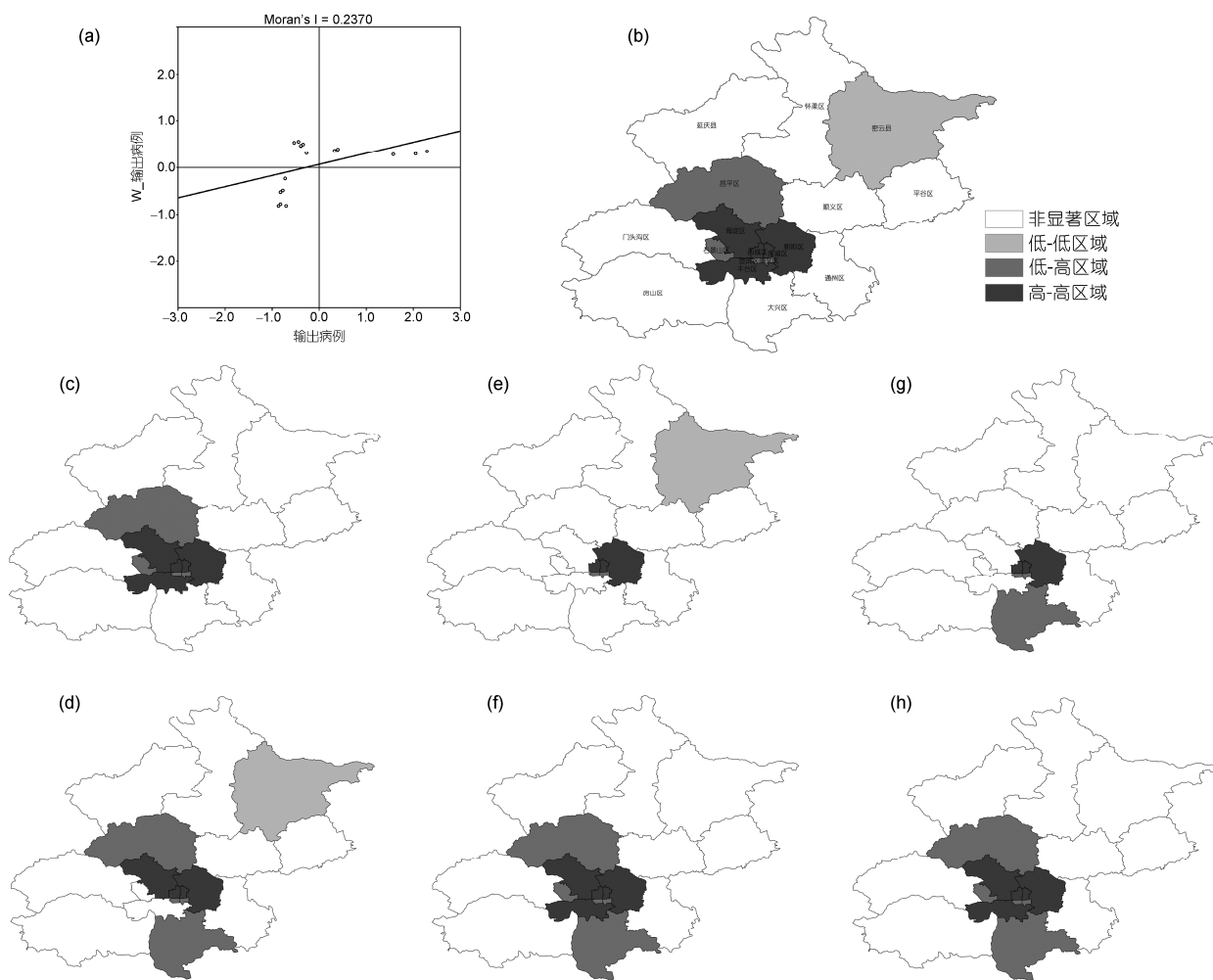


图 9 各区县内部输出病例风险类型的空间特征

(a) Moran 散点图(所有内部输出病例). 风险类型图: (b) 所有内部输出病例; (c) 男性病例; (d) 女性病例; (e) 20<年龄≤40 的病例; (f) 40<年龄≤60 的病例; (g) 医疗卫生行业的病例; (h) 工人/农民/民工职业的病例

表 4 各区县内部输出病例风险类型的时序演化过程^{a)}

区县	时间										整体风险
	第 1 周	第 2 周	第 3 周	第 4 周	第 5 周	第 6 周	第 7 周				
密云县	0	→	0	→	0	→	4	→	4	→	4
昌平区	0	→	0	→	2	→	2	→	2	→	2
顺义区	2	→	0	→	0	→	0	→	0	→	0
海淀区	0	→	0	→	1	→	1	→	1	→	1
朝阳区	1	→	0	→	1	→	1	→	1	→	1
石景山区	0	→	2	→	2	→	0	→	0	→	2
丰台区	0	→	2	→	2	→	0	→	1	→	1
大兴区	0	→	0	→	0	→	2	→	2	→	0
崇文区	2	→	2	→	2	→	2	→	2	→	2
东城区	0	→	0	→	2	→	1	→	1	→	1
西城区	2	→	2	→	1	→	1	→	1	→	1
宣武区	2	→	0	→	2	→	2	→	2	→	2

a) 0-非显著区域; 1-高-高区域(热点); 2-低-高区域; 3-高-低区域; 4-低-低区域. 始终为非显著自相关性的区域: 延庆县/怀柔区/平谷区/门头沟区/通州区/房山区

高风险区所包围, 为次级热点区域; (4) 市中心区的石景山区仅在流行初期表现为低-高区域, 西北方位的市郊昌平区在流行中后期皆表现为次级热点区域; (5) 唯一表现为冷点区域的密云县仅位于流行中后期的两周范围内, 其他非显著区域在整个时序上仅有短暂的次级热点区域的表现形式。综上北京市 SARS 传播防控重点应在整个流行期内首要关注市中心的海淀区、朝阳区与东城区的病毒扩散情况, 其次是市中心区的崇文区与宣武区; 在流行初期同时侧重市中心的石景山区; 在流行中后期侧重东城区与丰台区以及西北方位的市郊昌平区, 同时可适当减轻对东北方位市郊的密云县的防控程度。

4.3.3 内部传播网络特征分析

基于 2003 年北京市 SARS 传播内部输入、输出与自传播病例数据, 构建北京市 SARS 内部传播网络(图 10)。该网络是一个双向加权网络, 节点是北京市 18 个行政区县, 连接边代表某节点的 SARS 病例的输入与输出, 两节点之间最多有着两条有向连接边, 其

方向代表 SARS 病例的输入或输出方向, 其权重值为两节点之间输入或输出病例的数量。此外该网络中还存在着环结构, 即某些节点有着与自身相连接的边, 表示自传播病例的数量与方向信息。由于内部传播网络的封闭性, 该网络的出度与入度总数均为 149, 平均度为 8.2778, 表明存在 SARS 病例输入输出交互的区县个数平均值约为 8 个。整体网络的聚类系数 C 和 C^* 值分别为 0.4288 和 0.3143, 表明北京市 SARS 内部传播网络的信息流比物质流的聚集程度要更高, 即相对于存在病例输入输出交互的区县数量而言, 输入输出病例数量更多地存在与少数区县之间。

自行编码计算每个区县节点的度数与聚类系数, 如表 5 所示, 可得出以下几点重要启示: (1) 除了市中心区的若干区县存在高值的入度与出度之外, 市郊通州区与昌平区也有着大量的输入输出节点, 是潜在的高风险区域, 且其聚类系数 C_i 和 C_i^* 值均不低, 表明其邻接节点之间较易形成局部小世界疾病爆发区域, 应整体予以重点防控; (2) 内部传播的输入输出交互的信息流存在一定异质性特征, 即某区县病

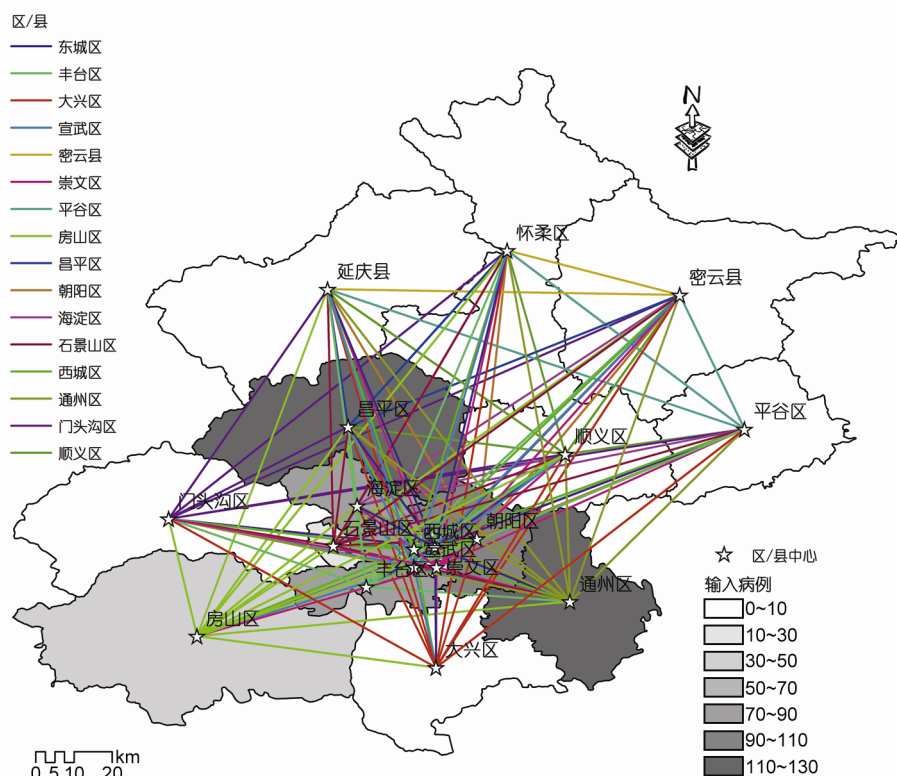


图 10 北京市 SARS 内部传播网络示意图

表 5 各区县节点的度分布与聚类系数

区/县	入度	出度	C_i	C_i^*
昌平区	17	11	0.3860	0.6389
朝阳区	16	13	0.4167	0.5522
崇文区	10	9	0.6556	0.5885
大兴区	5	9	0.6000	0.0903
东城区	12	10	0.5606	0.5862
房山区	10	6	0.5778	0.4338
丰台区	14	12	0.4176	0.4924
海淀区	13	12	0.4103	0.3623
怀柔区	0	6	0.0000	0.0000
门头沟区	1	4	0.0000	0.0000
密云县	1	3	0.0000	0.0000
平谷区	2	2	0.5000	0.0012
石景山区	7	8	0.6429	0.3927
顺义区	2	10	0.0000	0.0000
通州区	15	10	0.4190	0.5428
西城区	9	12	0.5417	0.2696
宣武区	12	9	0.5909	0.6249
延庆县	3	3	1.0000	0.0821

例输入与输出节点个数不一致,如市郊顺义区、怀柔区、门头沟区和密云县的输入节点个数仅为 0~2 个,而输出节点个数却为 3~10 个,又如市郊通州区和房山区的输入与输出节点个数分别为 10~15 个以及 6~10 个,这表明各区县在风险程度上有着信息流异质差别,在隔离防控措施上应予以区别对待,分别对区县病例输入和输出给予不同程度的防控措施,将有效减少资源成本提高防控效率;(3) 大部分区县的 C_i 和 C_i^* 值均保持一致,表明在北京市 SARS 内部传播网络中存在较程度上的小世界聚集特征,即少数区县承载了大部分 SARS 传播输入输出的信息流与物质流信息;(4) 少数区县的 C_i 和 C_i^* 值存在一定异质性特征,如延庆县、大兴区和平谷区的 C_i 值分别高达 1.0000, 0.6000 和 0.5000, 而其 C_i^* 值仅分别为 0.0821, 0.0903 和 0.0012, 表明此三个区县所邻接的其他节点之间虽有很高的信息流交互,但承担的物质流却极低,说明它们周围的小世界应是相对冷点区域;(5) 顺义区、怀柔区、门头沟区和密云县的 C_i

和 C_i^* 值均为零,表明以其为中心的局部网络中没有形成小世界特征,同时基本没有承载物质流信息,可视为 SARS 传播的绝对冷点区域,降低其防控措施强度可有效减少资源成本.

5 讨论与总结

传染病时空传播特征与规律的研究涉及了流行病学、空间信息科学、空间统计分析、社会学、人类学等多个学科领域,大量研究成果已对传染病的传播机制、时空分布、防控措施等做出了详尽的解释与建议,但仍然有很多未解谜团,以已有的传染病流行为基础的回顾性分析可以较全面地分析解译该传染病的病毒机理与时空传播规律,也为未来新型传染病爆发流行的预防与控制提供良好的理论支持与措施建议. 以 2003 年的中国 SARS 流行为例,充分暴露了我国流行病学研究的不足以及突发公共卫生事件应急处置措施的局限,同时也引发了国内外学者对于流行病学研究的大量关注与投入. 时至今日,关于 SARS 传播规律的研究已初具一定规模与成效,但对于个体在受感染-发病-就诊-康复传播过程的时空演化规律以及个体与区域之间形成的动态有向传播网络仍缺乏充分的认识 and 解释. 本文提出的 SARS 传播输入-输出流的概念,可有效解释 SARS 传播个体位置的时空演化机制,以及动态有向传播网络的信息流与物质流特征信息,进而基于 2003 年中国 SARS 感染者的个案病例时空属性数据,利用空间统计分析方法与网络特征分析方法,对北京市 SARS 传播的流行病学特征、外部流的时空传播规律与热点区域特征、以及内部流的时空传播特征与网络结构特征进行了全面的分析与研究,并依据相应的分析结果对北京市 SARS 流行的防控措施提出相应的建议与评价. 本文的研究思路是从一个全新的角度去解释传染病的时空传播规律,与关于传染病流行的一般研究中的单一空间位置属性相比,可以更好揭示 SARS 流行的时空传播机制,为传染病时空传播规律的研究提供较好的新型理论思路与实践手段,研究成果更深入地揭示了北京市 SARS 流行的特征与规律,也为未来其他突发传染病的防控措施提供了更有利的理论支持.

致谢 感谢审稿人提出的修改意见.

参考文献

- 1 Dye C, Gay N. Epidemiology modeling the SARS epidemic. *Science*, 2003, 300: 1884–1885
- 2 石耀霖. SARS 传染扩散的动力学随机模型. *科学通报*, 2003, 48: 1373–1377
- 3 Jia N, Tsui L. Epidemic modelling using SARS as a case study. *North Am Actuarial J*, 2005, 9: 28–42
- 4 Anderson R M, Fraser C, Ghani A C, et al. Epidemiology transmission dynamics and control of SARS: The 2002–2003 epidemic. *Philos Trans R Soc Lond B Biol Sci*, 2004, 359: 1091–1105
- 5 Lipsitch M, Cohen T, Cooper B, et al. Transmission dynamics and control of severe acute respiratory syndrome. *Science*, 2003, 300: 1966–1970
- 6 Pang X, Zhu Z, Xu F, et al. Evaluation of control measures implemented in the severe acute respiratory syndrome outbreak in Beijing, 2003. *JAMA*, 2003, 290: 3215–3221
- 7 庞星火, 刘东磊, 龚晓红, 等. 北京市 SARS 患者密切接触者发病及其危险因素分析. *中华流行病学杂志*, 2004, 25: 35–37
- 8 Fang L Q, de Vlas S J, Feng D, et al. Geographical spread of SARS in mainland China. *Trop Med Int Health*, 2009, 14(Suppl 1): 14–20
- 9 Wang J F, McMichael A J, Meng B, et al. Spatial dynamics of an epidemic of severe acute respiratory syndrome in an urban area. *Bull World Health Organ*, 2006, 84: 965–968
- 10 Meng B, Wang J F, Liu J, et al. Understanding the spatial diffusion process of severe acute respiratory syndrome in Beijing. *Public Health*, 2005, 119: 1080–1087
- 11 曹志冬, 曾大军, 郑晓龙, 等. 北京市 SARS 流行的特征与时空传播规律. *中国科学: 地球科学*, 2010, 40: 776–788
- 12 Wang J F, Christakos G, Han W G, et al. Data-driven exploration of ‘spatial pattern-time process-driving forces’ associations of SARS epidemic in Beijing, China. *J Public Health (Oxf)*, 2008, 30: 234–244
- 13 龚建华, 孙战利, 李小文, 等. SARS 疫情控制的模拟分析. *遥感学报*, 2003, 7: 260–265
- 14 龚建华, 周洁萍, 徐珊, 等. SARS 传播动力学模型及其多智能体模拟研究. *遥感学报*, 2006, 10: 829–835
- 15 林国基, 贾珣, 欧阳颀. 用小世界网络模型研究 SARS 病毒的传播. *北京大学学报(医学版)*, 2003, 35(增刊): 66–69
- 16 Small M, Tse C K. Small world and scale free model of transmission of SARS. *Int J Bifurcat Chaos*, 2005, 15: 1745–1755
- 17 王劲峰, 孟斌, 郑晓瑛, 等. 北京市 2003 年 SARS 疫情的多维分布及其影响因素分析. *中华流行病学杂志*, 2005, 26: 16–20
- 18 刘亚岚, 阎守邕, 李小文, 等. 中国内地人口流动空间规律研究及其在 SARS 控制宏观决策中的应用. *遥感学报*, 2003, 7: 273–276
- 19 Jia N, Feng D, Fang L Q, et al. Case fatality of SARS in mainland China and associated risk factors. *Trop Med Int Health*, 2009, 14(Suppl 1): 21–27
- 20 Wu J, Xu F, Zhou W, et al. Risk factors for SARS among persons without known contact with SARS patients, Beijing, China. *Emerg Infect Dis*, 2004, 10: 210–216
- 21 Lau J T, Tsui H, Lau M, et al. SARS transmission, risk factors, and prevention in Hong Kong. *Emerg Infect Dis*, 2004, 10: 587–592
- 22 Tobler W. A computer movie simulating urban growth in the detroit region. *Econ Geogr*, 1970, 46: 234–240
- 23 李小文, 曹春香, 常超一. 地理学第一定律与时空邻近度的提出. *自然杂志*, 2007, 29: 69–71
- 24 Moran P A. Notes on continuous stochastic phenomena. *Biometrika*, 1950, 37: 17–23
- 25 Geary R C. The contiguity ratio and statistical mapping. *Incorp Statistician*, 1954, 5: 115–145
- 26 Getis A, Ord J K. The analysis of spatial association by use of distance statistics. *Geogr Anal*, 1992, 24: 189–206
- 27 Anselin L. Local indicators of spatial association-LISA. *Geogr Anal*, 1995, 27: 93–115
- 28 Ord J K, Getis A. Local spatial autocorrelation statistics: Distributional issues and application. *Geogr Anal*, 1995, 27: 286–306
- 29 Anselin L. *Spatial Analytical Perspectives on GIS*. London: Taylor & Francis Ltd, 1996. 111–125
- 30 Levine N. *CrimeStat: A spatial statistics program for the analysis of crime incident locations (version 3.3)*. Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC. 2010
- 31 Wang J F, Li X H, Christakos G, et al. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. *Int J Geogr Inf Sci*, 2010, 24: 107–127
- 32 Watts D J, Strogatz S H. Collective dynamics of ‘small-world’ networks. *Nature*, 1998, 393: 440–442

- 33 Barabasi A L, Albert R. Emergence of scaling in random networks. *Science*, 1999, 286: 509–512
- 34 Amaral L A, Scala A, Barthelemy M, et al. Classes of small-world networks. *Proc Natl Acad Sci USA*, 2000, 97: 11149–11152
- 35 Kulldorff M, Nagarwalla N. Spatial disease clusters: Detection and inference. *Stat Med*, 1995, 14: 799–810
- 36 Kulldorff M. A spatial scan statistic. *Commun Statist Theory Methods*, 1997, 26: 1481–1496
- 37 Kulldorff M, Heffernan R, Hartman J, et al. A space-time permutation scan statistic for disease outbreak detection. *PLoS Med*, 2005, 2: e59