Spatial-temporal characteristics of epidemic spread in-out flow —Using SARS epidemic in Beijing as a case study

HU BiSong^{1,2}, GONG JianHua^{2,3*}, ZHOU JiePing^{2,3}, SUN Jun², YANG LiYang²,
XIA Yu¹ & Abdoul Nasser IBRAHIM^{2,3}

¹ *Geography and Environment Department, Jiangxi Normal University/Key Laboratory of Poyang Lake Wetland and Watershed Research, Ministry of Education, Nanchang 330022, China;*

² *Institute of Remote Sensing Applications, Chinese Academy of Sciences/State Key Laboratory of Remote Sensing Science, Beijing 100101, China;*

³ *Zhejiang-CAS Application Center for Geoinformatics, Jiaxing 314100, China*

Received February 27, 2012; accepted August 27, 2012; published online October 28, 2012

For better detecting the spatial-temporal change mode of individual susceptible-infected-symptomatic-treated-recovered epidemic progress and the characteristics of information/material flow in the epidemic spread network between regions, the epidemic spread mechanism of virus input and output was explored based on individuals and spatial regions. Three typical spatial information parameters including working unit/address, onset location and reporting unit were selected and SARS epidemic spread in-out flow in Beijing was defined based on the SARS epidemiological investigation data in China from 2002 to 2003 while its epidemiological characteristics were discussed. Furthermore, by the methods of spatial-temporal statistical analysis and network characteristic analysis, spatial-temporal high-risk hotspots and network structure characteristics of Beijing outer in-out flow were explored, and spatial autocorrelation/heterogeneity, spatial-temporal evolutive rules and structure characteristics of the spread network of Beijing inner in-out flow were comprehensively analyzed. The results show that (1) The outer input flow of SARS epidemic in Beijing concentrated on Shanxi and Guangdong provinces, but the outer output flow was disperse and mainly includes several north provinces such as Guangdong and Shandong. And the control measurement should focus on the early and interim progress of SARS breakout. (2) The inner output cases had significant positive autocorrelative characteristics in the whole studied region, and the high-risk population was young and middle-aged people with ages from 20 to 60 and occupations of medicine and civilian labourer. (3) The downtown districts were main high-risk hotspots of SARS epidemic in Beijing, the northwest suburban districts/counties were secondary high-risk hotspots, and northeast suburban areas were relatively safe. (4) The district/county nodes in inner spread network showed small-world characteristics and information/material flow had notable heterogeneity. The suburban Tongzhou and Changping districts were the underlying high-risk regions, and several suburban districts such as Shunyi and Huairou were the relatively low-risk safe regions as they carried out minority information/material flow. The exploration and analysis based on epidemic spread in-out flow help better detect and discover the potential spatial-temporal evolutive rules and characteristics of SARS epidemic, and provide a more effective theoretical basis for emergency/control measurements and decision-making.

in-out flow, SARS, Beijing, epidemic spread network, spatial-temporal characteristics, control measurement

Citation: 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

*Corresponding author (email: jhgong@irsa.ac.cn)

Various complicated infectious diseases come along with human evolution, and mankind has repeatedly confronted extremely serious threat of known or unknown infectious diseases in human development history. Recent serious epidemic outbreaks with worldwide impact include Severe Acute Respiratory Syndrome (SARS) during the spring of 2003 and influenza A H1N1 with worldwide spread in April 2009. SARS is a novel respiratory disease caused by a coronavirus infection, which has main symptoms of fever, dry cough, chest tight, respiratory failure, etc, and has an extremely strong infectious capacity with the paths including close droplet transmission, contact with respiratory secretions, and close spatial contact with infected cases. According to the statistics of the Ministry of Health of China, the count of provinces with SARS epidemic was 26 in mainland China, and there were significant regional differences in the distribution of number of SARS cases based on the situation of daily epidemic outbreak and cumulative count of suspected and confirmed cases. The suspected and confirmed SARS cases were concentrated in Beijing, Guangdong, Shanxi, Inner Mongolia, Tianjin, Hebei, etc., and Beijing and Guangdong were the most important clustering hotspot centers of SARS epidemic. The first SARS cases of most provinces were students, migrant workers, business travelers, and tourists from Guangdong, Beijing and Hong Kong. SARS epidemic was the first worldwide popular event of public health in the 21st century, and researchers have comprehensively and thoroughly studied the virus characteristics, transmission mechanism, spatial-temporal distribution, prevention and control measures of SARS epidemic according to the theories and methods of pathology, epidemiology, medical statistics, spatial information science, etc., The spread process of SARS epidemic was caused by the collaborative effect of novel infectious virus, resources, environment, society, economy and other complex factors, and the quantitative studies of spatial-temporal patterns of SARS epidemic help thoroughly recognize the potential spread mechanism in the spatial-temporal scale, explain the driving effects of various factors in SARS epidemic, evaluate the infectious risk of regions and populations, provide theoretical supports for the emergency measures of prevention and control in SARS epidemic, and provide scientific basis for the researches of other unknown novel infectious diseases.

Recent researches of spatial-temporal patterns of SARS epidemic could be divided into the following categories: (1) theoretical modeling for the mathematical epidemic model, including modeling for the SARS epidemic according to classic deterministic SIR model or stochastic model [1, 2], calculating of basic reproductive number R_0 and effective reproductive number R_t [3], and analysis of the dynamic mechanism of SARS epidemic process [4, 5]. (2) Mathematical statistical analysis of epidemiological data, such as quantitative analysis for temporal individual cases, close contacts and control measures using classical mathematical

statistical methods [6, 7]. (3) Spatial-temporal statistical analysis of SARS epidemic characteristics, including exploring the distribution patterns of SARS cases in spatial-temporal scales [8], analyzing the spatial autocorrelation and heterogeneity characteristics of SARS epidemic using spatial-temporal statistical methods [9–11], and exploring the spatial pattern, temporal process, and driving factors of SARS epidemic using meta models combining multiple spatial-temporal statistical methods [12]. (4) simulation and prediction of SARS epidemic process based on network dynamic models, such as simulating and analyzing the SARS epidemic process using system dynamics and multi-agent system [13, 14], and simulating the spread mechanisms based on models of small-world network and scale-free network [15, 16]. (5) Quantitative analysis and evaluation of SARS epidemic influencing factors such as close contact, demographic distribution, socioeconomic situation, etc. [17–21]. According to the research scale, the first three kinds of researches focus on the statistical epidemic characteristics and results based on macroscopic studies, the fourth kind of research studies the spread pattern of microcosmic individuals and regions and its corresponding reflecting spread results in macroscopic region, and the last kind of research studies the driving effectiveness of impact factors to SARS epidemic based on macroscopic studies.

Those above-mentioned methods of analyzing the spatial-temporal patterns of SARS epidemic have the following three inadequacies: (1) common methods analyze the spatial and temporal information of SARS epidemic separately and focus on the combination of analysis results in spatial and temporal dimensions, and the corresponding trend should be collaborative analysis of multiple datasets such as spatial data, temporal data, disease parameters, spread influencing factors, etc. [12]. (2) The epidemic data mainly include the confirmed infected cases and close contacts, and the analysis rarely focuses on the spatial-temporal changes of individuals during the process of susceptible-infected-symptomatic-treated-recovered. (3) The virus transmission mechanism between individuals and the interactive epidemic spread mode between regions will form a dynamic and directional network in the process of SARS epidemic, and the existed researches lack a quantitative analysis of SARS transmission network and explanation of its structural characteristics. Additionally, Beijing municipality was the most seriously affected region during SARS epidemic in mainland China, and its emergency measures of prevention and control were opportune and effective. Furthermore, Beijing as the capital of China has frequent interactions with other provinces and the rest of the world. Hence, it has remarkable demonstration effects and practical significance to take Beijing as the study area of analyzing the spatial-temporal patterns of SARS epidemic.

Based on the above considerations, this paper defined and explained the SARS epidemic spread in-out flow in

Beijing and its epidemiological characteristics based on the SARS epidemiological investigation data in China from 2002 to 2003, and by the methods of spatial-temporal statistical analysis and network characteristic analysis, explored the spatial-temporal high-risk hotspots and network structure characteristics of Beijing outer in-out flow. Furthermore, the spatial autocorrelation/heterogeneity, spatial-temporal evolutive rules and structure characteristics of the spread network of Beijing inner in-out flow were comprehensively analyzed. Compared with analytical methods based on infected cases or close contacts, the exploration and analysis based on epidemic spread in-out flow help better detect and discover the potential spatial-temporal evolutive rules and characteristics of SARS epidemic, and provide a more effective theoretical basis for emergency/control measures and decision-making.

1 Data processing and recovery

1.1 Processing spatial-temporal data

The original experimental data included attributes of confirmed infected individuals from November 16, 2002 to May 10, 2003. The individual attributes include gender, age, occupation, household registration, working unit/home address, onset location, reporting unit, onset time, treated time, confirmed time and etc. The proposed SARS epidemic spread in-out flow focuses on three sets of spatial information, i.e. working unit/home address, onset location, and reporting unit, and thus individual cases with at least one spatial information referred to “Beijing” were selected. The number of selected cases with those three sets of spatial

information had proportions of 15.88%, 35.87% and 34.09% respectively. Figure 1 shows the temporal changes of individual number with three sets of spatial information. Obviously, SARS individual number corresponding to the three sets of spatial information had similar changing trend, which had stable growth phase in early March, rapid growth to a peak value in mid to late April, and rapid decline to fadeaway in mid to late May. The changing trend of individual number corresponding to onset location and reporting unit was extremely consistent, which indicates the consistency of onset and treatment locations of infected cases during the process of SARS epidemic in Beijing. However, the trend of individual number corresponding to working unit/home address had some differences, which indicate that some individuals that were infected in Beijing had onset or treatment locations in other regions, and there were scattered new cases in early process of SARS epidemic.

The original SARS epidemiological data were stored in the form of data sheets, and the individual onset time which was selected as the temporal information could be directly applied for quantitative analysis in Time-Date format. Those three sets of spatial information were stored in the form of ordinary Text format and the description scales were not uniform. Therefore, the processing of spatial data had the following three rules: (1) the scales of three sets of spatial information were converted to consistent province or municipality, and the scale of spatial information referred to “Beijing” was converted to district or county. (2) The selected data were assured with at least two integrated sets of spatial information, which means those data with two or three sets of spatial information missing were eliminated. (3) The selected data had at least one set of spatial information

Figure 1 Changes of daily new cases of SARS epidemic in Beijing from 2002 to 2003. Corresponding to three sets of spatial information: working unit/home address, onset location and reporting unit.

referred to 'Beijing'. After the processing of spatial data, there were 14 eliminated individual cases in mainland China and an estimated 37.15% cases referred to 'Beijing'. For the final selected data, those three sets of spatial information should be matched artificially and manually with geographical administrative maps of nation and Beijing.

1.2 Recovery of missing data

Temporal information and onset location of the selected individual cases in mainland China were complete, but there were about 51.63% of cases with missing or fractional information of working unit/home address and about 1.26% of cases with missing information of reporting unit. Besides, there were little missing information of individual attributes such as gender, age, occupation, etc. Hence, recovery of missing data was necessary to keep the data's integrity before the next analysis. The recovery process was considered to keep the consistency of attribute distributions of regions and measure the relative influence of population distributions of regions.

For certain attributes with missing data such as working unit/home address, n_i and p_i are assumed as the amounts of existed effective data and population for some region i , and for all regions, the total corresponding amounts are $N = \sum_i n_i$ and $P = \sum_i p_i$. For some region i , its relative ratios of existed effective data and population could be calculated as

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

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

A parameter α is used to describe the relative adjustment between existing effective data and population. And the ratio of data recovery for region i and its accumulative ratio could be calculated as

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

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

A random number *rand* which is located from 0 to 1 is set for certain attributes with missing data; *rand* is determined to be located in some interval of $[\Theta_{i-1}, \Theta_i]$, and the certain information corresponding to Θ_i is used to recover the missing attribute data.

Given the characteristic consistency of spatial distribution of original data, the adjustment parameter α is set as 0.2. Based on the above recovery process, the recovery result for missing data of Beijing was satisfactory, and the similarity ratios of data recovery for the attributes of working unit/home address and reporting unit reached to values of

99.68% and 99.77%. In the recovery for missing data of whole nation, the similarity ratio for the attribute of reporting unit was 99.99% because of the extremely small proportion of missing data, and the similarity ratio for the attribute of working unit/home address was 98.22% due to the relatively large proportion of missing data.

2 Definition and interpretation of epidemic spread in-out flow

Only certain spatial information such as onset location is considered in many researches on epidemic transmission and spatial-temporal simulation. During the epidemic transmission period, however, susceptible individuals change the location information continuously after getting infected. Practically, susceptible individuals have infected, exposed and treatment periods corresponding to various spatial information such as infected, onset and treatment locations. Epidemic spread in-out flow focuses on the location transformation process of infected individuals during the epidemic transmission period, and explores the spread mechanism of epidemic transmission inputs and outputs between spatial regions in various scales. The definition and interpretation of epidemic spread in-out flow are introduced by using SARS epidemic in Beijing as an example.

Three sets of typical location information of individuals in epidemic period were selected for the definition of in-out flow, including working unit or home address (S1), onset location (S2), and reporting unit (S3). S1 could be considered as the main residence place of individuals because working unit and home address are the two most important aspects of location information. S2 is used as the infected location of individuals approximately. The exposed period of most infectious diseases is short and the infected spatial-temporal information of individuals is difficult to collect; therefore, the onset location is used instead of infected location. S3 commonly expresses the treatment location of infected individuals, which is collected by the last medical units. The SARS spread in-out flow could describe the transmission mechanism of infectious diseases for different regions in various scales. For infected individuals, the main residence place transforms to infected location after getting infected, and transforms to onset location when the symptoms appear, and finally transforms to reporting unit where individuals receive treatment. For spatial regions, there are spread process of virus inputs and outputs between regions because individuals transform the location information from regions to regions during the epidemic period.

For certain region of Beijing in the scale of province or municipality, the epidemic in-out flow has two types: outer flow and inner flow. Thereinto, outer flow describes the inputs or outputs of individuals between Beijing and other regions in the same scale, and inner flow describes the inputs or outputs between different districts or counties of

Beijing. For the outer flow of Beijing, inputs mean the cases which have the same residence place and treatment location corresponding to Beijing, and outputs are those cases whose treatment location is Beijing but residence place is not Beijing. Moreover, as far as the infected location of individuals is considered, inputs and outputs could be divided into primary and secondary ones. Primary inputs are cases whose infected location is Beijing and primary outputs are cases whose infected location is not Beijing; secondary inputs and outputs are defined oppositely to the corresponding primary ones. For the inner flow of Beijing, three sets of spatial information of individuals are districts or counties of Beijing such as Chaoyang and Haidian districts, and its inputs and outputs are defined similarly. The epidemic in-out flow of Beijing could be described in the following logical expressions:

Outer Input: $S1 \neq \text{'Beijing'}$ AND $S3$

$$= \text{'Beijing'} \begin{cases} \text{Primary: } S2 = \text{'Beijing'}, \\ \text{Secondary: } S2 \neq \text{'Beijing'}, \end{cases} \quad (5)$$

Outer Output: $S1 = \text{'Beijing'}$ AND $S3$

$$\neq \text{'Beijing'} \begin{cases} \text{Primary: } S2 \neq \text{'Beijing'}, \\ \text{Secondary: } S2 = \text{'Beijing'}, \end{cases} \quad (6)$$

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

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

Based on the logical definition of epidemic in-out flow, during the SARS spread process, Beijing had inputs and outputs of individuals from getting infected to receiving treatment with other provinces or municipalities, and there are also similar interaction processes between districts or

counties of Beijing. As illustrated in Figure 2, in time sequence, the outer flow describes the inputs and outputs between provinces centered in Beijing, and the inner flow expresses a dynamic network with direction and weight between districts in Beijing.

SARS spread in-out flow focuses on the location changes of individuals during the epidemic process and directly describes the virus transmission mechanism. In the perspective of epidemiology, outer flow reflects the interactive process of individuals from getting infected to receiving treatment between Beijing and other provinces, and inner flow describes the mechanism of inputs and outputs between districts in Beijing. Input flow reflects the transformation of infected individuals with residence place not in Beijing but getting treatment in Beijing, and output flow describes the transformation of infected individuals with residence place in Beijing but getting treatment in other places outside of Beijing. According to the definitions of inputs and outputs of inner and outer flows of Beijing, the epidemiological interpretation of SARS in-out flow could be shown in Table 1.

3 Epidemiological analysis

The input and output cases of Beijing began to appear respectively on January 14, 2003 and December 1, 2002. As of May 10, 2003, the total number of outer input cases was 1024, among which the primary and secondary input cases were 1012 and 12, and the total number of outer output cases was 237, among which the primary and secondary output cases were 191 and 46. The inner cases started at the beginning of SARS outbreak (in late March 2003) and fi-

Figure 2 Definition of SARS spread in-out flow in Beijing.

Table 1 Epidemiological interpretation of SARS in-out flow in Beijing

	Type	Characteristic	Epidemiological interpretation
Outer flow	Primary inputs	Population living in other provinces got infected in Beijing and received treatment in Beijing.	Individuals were from other provinces, got infected in Beijing while travelling, and received treatment in local Beijing. These cases were affected by economic and traffic developments between provinces and were main sources of outer input cases of Beijing.
	Secondary inputs	Population living in other provinces got infected in other provinces but received treatment in Beijing.	Individuals came from and got infected in other provinces, but received treatment in Beijing considering influences such as medical levels. Virus spread continued during the whole location transformation process. These cases were the focus of control measures of SARS epidemic in Beijing to outer floating population.
	Primary outputs	Population living in Beijing got infected in other provinces and received treatment in other provinces.	Individuals living in Beijing got infected and received treatment in other provinces while travelling. Virus spread was reduced in other local provinces, but for Beijing, these cases were main sources of output cases which control measures should focus on during the period of SARS period.
	Secondary outputs	Population living in Beijing got infected in Beijing but received treatment in other provinces.	Individuals living in Beijing got infected locally, but received treatment in other provinces for some reasons. Virus spread continued during the whole location transformation process started at Beijing. These cases were the focus of control measures of SARS epidemic to output cases from Beijing.
Inner flow	Inputs	Population living in Beijing received treatment in certain district of Beijing but got infected in another district of Beijing.	Epidemic spread continued in inner Beijing. These cases indicated the clustering of virus spread for some district/county, and were the focus of control measures to input cases in Beijing.
	Outputs	Population living in Beijing got infected in certain district of Beijing but received treatment in another district of Beijing.	Epidemic spread continued in inner Beijing. These cases indicated the dispersing of virus spread for some district/county, and were the focus of control measures to output cases in Beijing.

nally reached a total number of 853. Characteristics of spatial distribution of the various types of cases above are shown in Figure 3(a). From this point of view, outer input cases of Beijing were mainly from Guangdong (121 cases) and several provinces around Beijing including Hebei (52 cases), Shandong (69 cases), Shanxi (70 cases), Henan (68 cases) and etc, and the distribution of input cases from other provinces in central China were even more, whereas those from northeastern and northwestern provinces were relatively rare. Outer output cases of Beijing were more concentrated in Guangdong (90 cases), followed by Hebei (36 cases), Shanxi (32 cases), and Inner Mongolia (17 cases), whereas there were only sporadic cases in other provinces (less than 10). From the temporal changes of the various types of cases shown in Figure 3(b), outer input cases began to appear from early March, broke out in early April, continued increasing until late April when reaching the peak of daily increasing number of input cases, and then the cases began to decline until disappearing; the developing trend of self spread cases was consistent with that of outer input cases, indicating that the contributions of outer input cases and inner self spread cases were essentially identical to SARS epidemic of Beijing, and the number of inner cases plummeted for a short term in late April probably because isolation measures had a significant effect on controlling the spread of virus at the outset. Outer output cases appeared earlier with stable trend all the time, and even during the severe epidemic in April the number of daily increasing cases remained at about 10. Some conclusions could be drawn based on the above analysis: (1) SARS epidemic in Beijing was composed mainly of outer input cases and inner self spread cases, and the former was mainly from Guangdong and the provinces around Beijing; (2) outer output cases were fewer as a whole, and most concentrated in Guangdong, Hebei, Shanxi, and Inner Mongolia; (3) ratios

of outer secondary input and output cases were 1.17% and 19.4% respectively, indicating that outer input cases mostly got infected and received treatment in Beijing, and although the overall number of output cases was fewer than that of input cases, the output cases contributed higher risk of infection because some cases had location transformation during the process of getting infected and receiving treatment; and (4) prevention and control measures were certainly effective for inner self spread cases, but required more massive strengthening for controlling outer input and output cases.

Characteristics of attributes including gender, age and occupation of outer input and output cases of SARS epidemic in Beijing were analyzed and are shown in Figure 4. Male and female input/output cases were basically consistent, and corresponding ratios were respectively 53%/47% and 56%/44%, indicating that infection risk of male population was slightly higher. Ages of the input/output cases were concentrated respectively in the ranges of 20–50 and 20–40, and the average ages were respectively 38.09 and 35.47. From the occupation point of view, input cases were concentrated in domestic/housework/service personnel, retirees, civil service/business/staff and health care personnel who accounted for a ratio of 47.36%, and output cases were concentrated in migrant labor/peasants/workers, domestic/housework/service personnel and teacher/student/caregivers who accounted for a ratio of 54.43%. As for the characteristics of population movement, young migrant labors who are usually employed in construction/service/restaurant industries became the main source of input cases of Beijing, the minor source should be tourists composed mainly of the retired and medical staff scheduled to Beijing for the epidemic control measures, and finally middle-aged business people traveling to Beijing were also the source of input cases who resulted in high average age of input cases.

Figure 3 Spatiotemporal pattern of SARS input and output cases of Beijing. (a) Spatial distribution of outer input/output cases; (b) temporal changes of outer input/output cases and inner cases. Xinjiang: Xinjiang Uygur Autonomous Region; Tibet: Tibet Autonomous Region; Guangxi: Guangxi Zhuang Autonomous Region (the same below).

Output cases of Beijing were composed of common floating population as well as educational personnel among whom students are key members.

4 Spatial-temporal analysis of SARS spread in-out flow

4.1 Methods of spatial autocorrelation analysis and network characteristic analysis

Modern spatial statistical analysis is based on the first law of geography proposed by Tobler, which describes that

everything is related to everything else, but near things are more related than distant things [22]. Tobler's first law reveals the characteristics of spatial dependence, spatial association, and spatial autocorrelation of regionalized variables related with spatial location. Spatial statistical analysis mainly explains two key concepts in the first law of geography: relation and proximity. Generally, relation refers to similar spatial distribution of variables and proximity means there are adjacent boundaries between regions or the distance of region centroids is limited to a small critical value. And traditional spatial proximity could be extended to spatiotemporal proximity when the time dimension is consid-

Figure 4 Attribute characteristics of SARS input and output cases of Beijing. (a) Input cases/gender/age; (b) output cases/gender/age; (c) input cases/occupation; (d) output cases/occupation.

ered [23]. Usually, indicators to evaluate global spatial autocorrelation used in spatial statistical analysis mainly include Moran's I [24], Geary's C [25], Getis's G [26], etc., and indicators to evaluate local spatial autocorrelation include Local Moran's I , Local Geary's C , LISA (Local Indicator of Spatial Association) [27], Getis's G^* [28], Moran scatterplot [29], etc. Based on spatial autocorrelation analysis, hotspot detection analysis could help search abnormal spatiotemporal clustering in the study region and reveal high-risk areas of epidemic spatiotemporal distribution, and mainly includes Spatial Scan Statistics [35], Nearest Neighbor Hierarchical Clustering [30], Geographical detector [31], etc.

In different spatial scales, SARS spread in-out flow could describe the spread and evolutionary mechanism between districts in Beijing or between Beijing and other provinces. Outer flow shows a directed network centered in Beijing and composed of provinces or municipalities, and inner flow shows an interactive directed network between districts or counties in Beijing. The input and output flows indicate transformation directions and weights of infected cases, and form a complex directed and weighted network. Compared with regular and random networks, complex network is more accurate in describing the various network-characteristics of real world. Especially after models of small-world network [32] and scale-free network [33], researches of complex network reach climax. In general, characteristic indicators to describe network topological

structure mainly include degree distribution, characteristic path length, clustering coefficient, betweenness centrality, etc.

In this paper, indicators including Moran's I , Local Moran's I , Moran scatterplot, Scan Statistics, network degree distribution and clustering coefficient were selected to quantitatively analyze spatiotemporal characteristics of SARS spread in-out flow in Beijing.

4.1.1 Moran's I and Moran scatterplot

Global Moran's I is used for evaluation of spatial autocorrelation of regionalized variables in the whole studied area, and could be calculated as [24]:

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

where n is the amount of subregions in the study area, x_i and x_j are attribute values of subregion i and j , $\bar{x}$ is the average value of all subregions, and w_{ij} describes the spatial proximity between subregion i and j .

Generally w_{ij} is evaluated based on spatial topology or critical value of distance between region centroids. For instance, $w_{ij}=1$ indicating subregion i and j are adjacent when the two subregions have collective boundaries or the distance of centroids is less than certain critical value d , and

contrariwise $w_{ij}=0$. There are also other definitions of spatial proximity based on population size, medical standard, city structure level, etc. [12].

Local Indicator of Spatial Association could discover the characteristics of local abnormal clustering and instability of regionalized variables. As a typical LISA, Local Moran's I could be calculated as [27]:

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

where $z_i = x_i - \bar{x}$, $z_j = x_j - \bar{x}$, and $z^T = (z_1, z_2, \dots, z_n)$.

Spatial proximity matrix W is composed of w_{ij} , and Moran scatterplot could be plotted by the coordinates of (Wz, z) [29]. Horizontal and vertical axes divided the scatterplot into four quadrants, which indicate four types of local spatial autocorrelation between subregions: high-high clustering, low-high clustering, low-low clustering, and high-low clustering.

4.1.2 Spatial and spatiotemporal scan statistics

Spatial and spatiotemporal scan statistic methods were proposed by Kulldorff and could be applied to hotspot detecting based on epidemic spatiotemporal data. Log likelihood ratio (LLR) is used as the evaluation norm in spatial scan statistics and null hypothesis (e.g. the possibility of persons in the study regions is identical) is required as the basis. The principle of scan statistic is as follows: a series of scan clusters are firstly generated as circles whose central points are geographical centres of the study regions, and the radii of these scan circles vary from zero to a specified maximum value; secondly, LLR values of all scan circles are calculated, and several regions with maximal LLR values are selected as available hotspot regions of disease bursting; finally, Monte Carlo hypothesis testing is used for the statistical significance evaluation of those available hotspot regions, and regions which have passed the testing are the detected hotspot regions [35, 36]. Models used in spatial scan statistic include Bernoulli model, Poisson model, Ordinal model, Exponential model, etc. For instance, LLR value of Poisson model is calculated as [35, 36]:

$$\begin{aligned} \text{LLR} = \frac{\hat{L}(Z)}{L_0} = & \left\{ \left(\frac{n_Z}{\mu(Z)} \right)^{n_Z} \left(1 - \frac{n_Z}{\mu(Z)} \right)^{\mu(Z)-n_Z} \right. \\ & \left(\frac{n_G - n_Z}{\mu(G) - \mu(Z)} \right)^{n_G - n_Z} \\ & \left. \left(1 - \frac{n_G - n_Z}{\mu(G) - \mu(Z)} \right)^{(\mu(G) - \mu(Z)) - (n_G - n_Z)} \right\} / \\ & \left(\frac{n_G}{\mu(G)} \right)^{n_G} \left(\frac{\mu(G) - n_G}{\mu(G)} \right)^{\mu(G) - n_G}, \end{aligned} \quad (11)$$

where Z is a 3-dimensional vector including coordinates of the centre point and the radius of the scan circle, n_Z and n_G are case amounts of scan circle region and whole studied area, and $\mu(Z)$ and $\mu(G)$ are population amounts of scan circle region and the whole study area.

The main model of spatiotemporal scan statistic is space-time permutation model [37], and its principle is similar with spatial scan statistic. The differences are that scan cluster used in space-time permutation model is not circle but cylinder, bottom surface of scan cylinder has the same meaning as the scan circle in spatial scan statistic, and the height of scan cylinder stands for time value. The height and radius values of scan cylinder also have specified maximum values. Suppose C_{zd} is observed number of cases in region z during day d , total number of observed cases is C , and C_A is observed number of cases in a particular cylinder A . Under the null hypothesis, the expected case number of particular scan cylinder A could be expressed as

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

When the spatial summation $\sum_{z \in A} C_{zd}$ and temporal summation $\sum_{d \in A} C_{zd}$ are both small compared to C , C_A is approximately Poisson distributed with mean value μ_A . Poisson generalized likelihood ratio (GLR) is used as a measure of the evidence that cylinder A contains a disease outbreak, and GLR value could be calculated as [37]:

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

4.1.3 Network degree distribution and clustering coefficient

Node degree means the count of edges which have this node as one of their two nodes, and in directed networks, it has two types: in-degree and out-degree. $p(k)$ is used as the degree distribution of network, which means the count ratio of nodes which have the degree value of k to the total nodes, and furthermore it could describe the probability that the degree value of a random node in network equals to k . According to different degree distributions, the networks could be classified into different types [34]. C_i is defined as clustering coefficient of node i , which means the ratio of practical edge count to utmost probable edge count between the nodes connecting to node i . Assuming there are k_i nodes connecting to node i , and the practical edge count between those k_i nodes is A_i , the clustering coefficient of node i could be described as [32]:

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

Clustering coefficient of network describes the connection situation between nodes connecting to a same node, and reflects the trending of aggregation clustering of nodes in network. It could be implemented as the average value of

clustering coefficients of all nodes [32]:

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

In the transmission networks of SARS epidemic in-out flow, edges connecting nodes have both directions and weights, and the in-degree and out-degree distribution could help describe the direction of epidemic spread between regions. However, standard expression of clustering coefficient could only describe the information flow of epidemic spread between nodes. To reflect the material flow of epidemic spread, an improved expression of clustering coefficient is advanced as follows:

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

where A_N is the count of all edges in network, Cas_j is the material flow of edge j , which means the count of input or output cases in the transmission network of epidemic in-out flow.

The improved clustering coefficient C_i^* describes the clustering situation of material flow between nodes connecting to node i . Its corresponding network clustering coefficient C^* could directly reflect the small-world characteristic of epidemic in-out flow.

4.2 Spatiotemporal characteristic analysis of outer in-out flow

4.2.1 Spatiotemporal hotspot scan analysis

A spatial statistical analysis software named SatScan (<http://www.satscan.org/>) was applied to calculate the hotspots of

SARS epidemic in-out flow in Beijing in 2003, in which Bernoulli model and space-time permutation model were used for spatial and spatiotemporal analysis respectively. In order to unify time properties of input and output cases, the time range was set from January 20, 2003 to May 10, 2003, a total of 16 weeks, and the time range of input cases was from March 8, 2003 to May 10, 2003 (from the 9th to the 16th week). Time interval used in space-time permutation model was set to a natural week, and the largest radiuses of spatial and temporal scanning covered 50% of the total number of cases and 50% of the spread time range respectively.

According to Table 2, outer input cases had concentrated sources and the main hotspot regions were Shanxi and Guangdong, which had values of 2.73 and 2.02 for relative risk (RR) coefficients; primary hotspot region corresponding to outer output cases was Guangdong with an RR value of 9.23, and secondary hotspot regions included northern provinces centered in Inner Mongolia and Qinghai with RR values of 4.88 and 2.12 respectively. As shown in Figure 5, outer control measures for SARS epidemic in Beijing should focus on the isolation of input cases from Shanxi and Guangdong, and meanwhile prevent the inner cases to transmit to Guangdong and northern provinces. As for both spatial and temporal abnormal hotspots, conclusions were slightly different. From Table 2 and Figure 6, primary hotspots corresponding to outer input cases remained to be Shanxi with the time range from the 9th to the 12th week and a test statistic value of 3.8387, and time ranges of secondary hotspots including Fujian, Chongqing and Jilin maintained one week or so; primary hotspots corresponding to outer output cases remained to be Guangdong with the time range from the 3rd week to the 10th week and a test

Table 2 Results of hotspot detecting in outer in-out flow

SARS	Model	Hotspots								
		Hotspot Type	Centre	Included regions	Radius (km)	Time (week)	LLR	P	ODE	RR/Test statistic
Input cases	Bernoulli	Primary	Shanxi	—	0	—	25.2469	≈ 0	2.61	2.73
		Secondary	Guangdong	—	0	—	22.4257	≈ 0	1.90	2.02
	Space-time permutation	Primary	Shanxi	—	0	9–12		0.5700	2.12	3.8387
		Secondary	Fujian	—	0	16–16		0.7600	3.15	3.2718
		Secondary	Chongqing	—	0	9–9		0.9800	7.83	2.3732
		Secondary	Jilin	Liaoning	393.87	15–15		0.9800	1.80	2.3371
Output cases	Bernoulli	Primary	Guangdong	—	0	—	102.4074	≈ 0	6.10	9.23
		Secondary	Inner Mongolia	Hebei/Tianjin/Shanxi	737.08	—	58.9547	≈ 0	3.36	4.88
		Secondary	Qinghai	Gansu	494.22	—	2.5995	0.5790	2.06	2.12
	Space-time permutation	Primary	Guangdong	Hong Kong/Macao/Guangxi/Hunan	502.84	3–10		≈ 0	2.55	22.6499
		Secondary	Shandong	Tianjin/Jiangsu/Hebei/Henan, etc.	994.43	13–16		0.0001	1.46	10.2125
		Secondary	Tibet	Qinghai/Xinjiang/Gansu	1337.49	16–16		0.1090	8.70	5.1414

Figure 5 Spatial distributions of hotspots in outer in-out flow. (a) Input cases; (b) output cases. The others are the same as in Figure 3.

statistic value of 22.6499, secondary hotspots included northern provinces centered in Shandong and Tibet, which had test statistic values of 10.2125 and 5.1414, and the former had a time range from the 13th week to the 16th week whereas time range of the latter was only one week or so.

Based on the above analysis, some findings could include:

(1) outer in-out flow of Beijing epidemic was of typical difference, and output flow covered a wide range of areas including Guangdong and some northern provinces whereas input flow was concentrated; (2) high-risk regions of input flow were primarily in Shanxi with a time range of only about four weeks and secondarily in Guangdong which was

Figure 6 Temporal changes of hotspots in outer in-out flow.

one of most serious regions of SARS epidemic; (3) high-risk regions of output flow were primarily in Guangdong with a time range of about eight weeks and secondarily in northern provinces centered in Shandong and Inner Mongolia with a time range of only about four weeks; and (4) prevention and control measures of SARS epidemic between Beijing and other provinces should be treated differently according to location and time: measures should be focused on isolation of output cases from Beijing to Guangdong during the early period of SARS epidemic (before March), attention should be paid to controlling the input cases from Shanxi and Guangdong during the early outbreak (in March), and measures should be taken to prevent output cases from Beijing to Shandong and other provinces during the late period (after March). Overall, prevention and control measures for outer input cases should be taken at the early period and mid-period of the SARS outbreak and concentrated in provinces of Shanxi and Guangdong, and measures for outer output cases should be taken at the outbreak beginning and focus on Guangdong and northern provinces such as Shandong.

4.2.2 Analysis of network structural characteristics

According to outer in-out flow of SARS epidemic in Beijing from January 20, 2003 to May 10, 2003 (16 weeks), in-degree and out-degree corresponding to input flow and output flow were calculated respectively and the results were $D_{in}=29$ and $D_{out}=21$, indicating that SARS input cases of Beijing were from provinces all over the mainland of country but output cases of Beijing involved fewer provinces. However, risk ranges of input and output flows were different; for instance, input flow was of high-value degree but covered the concentrated risk areas (mainly in Shanxi and Guangdong), and the degree of output flow was not so high-value but the risk was more dispersed covering many areas including Guangdong and most northern provinces

such as Shandong. With time unit of one week, temporal changes of in-degree and out-degree of outer in-out flow in Beijing were calculated and are shown in Figure 7. Information can be concluded as follows: (1) out-degree appeared early and output cases from Beijing to other provinces began from late January, and in-degree appeared much later until the beginning of SARS outbreak in early March; (2) temporal change of out-degree was gentle, even though at the outbreak of SARS in late April, its value remained at the level from 5 to 15, and its value was mainly below 5 at early period of epidemic spread, indicating that out-flow from Beijing to other provinces affected a smaller range; (3) in-degree appeared in as late as the 9th week (early March) with an initial value of about 14, and soon expanded to 25–30, indicating that input flow covered the whole country in a relatively short time period; and (4) there were obvious turning points for temporal changes of in-degree, and it reduced from initial value of 14 to a value of 7 and then rapidly increased to the maximum in March, indicating that prevention and control measures of Beijing at the outbreak beginning were remarkably effective in about two weeks and played an important role in preventing

Figure 7 Temporal change of degree distribution in outer in-out flow of Beijing.

the spread of outer input flow, but the input flow affected nationwide rapidly during the outbreak period because of the increasing strength of SARS outbreak or the failure of control measures.

4.3 Spatiotemporal characteristic analysis of inner in-out flow

4.3.1 Global spatial autocorrelation analysis

Based on the data of inner input and output cases of SARS epidemic in Beijing, a spatial statistical analysis software named OpenGeoDa (<http://geodacenter.asu.edu/ogeoda/>) was applied to build the spatial proximity matrix between districts and counties in Beijing based on critical distances, and the values of global Moran's I of SARS epidemic inner in-out flow in Beijing in 2003 were calculated. Values of global Moran's I of inner input cases and self spread cases were calculated as 0.0477 and 0.0048 respectively, indicating insignificant spatial autocorrelation. However, global Moran's I of inner output cases was calculated as 0.2370, expected value of Moran's I equalled -0.0588 ($E(I) = -0.0588$) based on the 999 times of Monte Carlo simulation according to the null-hypothesis (H_0), standard deviation equalled 0.1387 ($S_d=0.1387$), test statistic value was 2.1327 ($Z\text{-Score}=(I-E(I))/S_d=2.1327$), and significance level was 0.031 and less than 0.05 ($P=0.031<0.05$). Therefore, inner output cases had significant characteristics of positive spatial autocorrelation in global area. Explanations could be proposed as follows for the above results: (1) inner self spread cases in Beijing mostly received treatment just in the same counties where he/she got infected, and would lead to insignificant spatial autocorrelation; (2) inner input cases lost the spatial autocorrelation due to the interference of inner self spread cases; and (3) inner output cases reflected the epidemic outbreak situation of districts and counties, and transformation pattern of individuals from getting infected to receiving treatment had relationship to some extent with spatial proximity and impact factors of districts, and therefore indicated significant positive spatial autocorrelation.

tion.

For further consideration of the influence of gender, age and occupation of inner output cases, global Moran's I was applied for evaluation, and temporal changes (from March 23, 2003 to May 10, 2003, a total of seven weeks) of spatial autocorrelation of inner output cases were analyzed. The results are shown in Table 3 and Figure 8, and some findings include: (1) spatial autocorrelation of female cases was slightly stronger than that of male cases, which was opposite to the overall gender distribution of SARS cases, but the difference according to gender was little; (2) there was weak significant spatial autocorrelation of cases younger than 20 and older than 60, but for cases of age from 20 to 60, the spatial autocorrelation was significant and increased with age; (3) there were obvious significant spatial autocorrelation of two types of populations including labor/peasants/workers and health care personnel; the latter was affected more by nosocomial infection due to the medical services, and the former was typical urban immigratory population with regular activity range and mode and therefore represented more significant spatial autocorrelation; and (4) spatial autocorrelation of inner output cases was not obvious in the early period of SARS epidemic because there were lots of infection sources bringing more new cases and the isolation control measures were mainly for hospitals; during the peak period of SARS epidemic from the 3rd to the 6th week (April 6 to May 3), there was a sharp increase of new cases and isolation control measures were mainly for hospitals and households, and thus inner output cases indicated a strong spatial autocorrelation and the strength was basically consistent; in the late period, there was a sharp reduction of new cases and infection sources mostly remained as hospitals and households, and therefore inner output cases continuously indicated a strong spatial autocorrelation but the significance was little due to the decrease of total cases.

Based on the above analysis, inner prevention and control measures for SARS epidemic in Beijing should focus on the isolations of inner output cases, minimize the virus

Table 3 Results of global Moran's I of various types of inner output cases^{a)}

Inner output cases		Moran's I	Monte Carlo simulation (999 times)		
			Standard deviation	Z value	P
Gender	Male	0.2253*	0.1450	1.9593	0.0380
	Female	0.2264*	0.1467	1.9441	0.0340
Age	(, 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
	Health care personnel	0.2084*	0.1492	1.7909	0.0460
Occupation	Domestic/housework/service personnel	0.0766	0.1298	1.0431	0.1170
	Labor/peasants/workers	0.2258*	0.1401	2.0314	0.0360
	Retiree	0.0486	0.1413	0.7601	0.1870

a) * indicates the remarkable spatial autocorrelation ($P<0.05$); $E(I) = -0.0588$.

Figure 8 Temporal change of global Moran's I of inner output cases.

spread caused by spatial transformation process of individuals for treatment, control the young and middle-aged with high-risk infection and with no gender differences, concentrate on surveillance of two types of populations including labor/peasants/workers and health care personnel, and at the peak period of SARS epidemic, pay more attention to control measures for the two main infection sources: hospitals and households.

4.3.2 Local spatial autocorrelation analysis

Similar with global spatial autocorrelation analysis, local Moran's I and Moran scatterplot were applied to detect the spatiotemporal characteristics of inner in-out flow based on inner output cases of SARS epidemic in Beijing in 2003. Risk analysis of inner output cases was implemented by software OpenGeoDa, and local spatial autocorrelation and risk levels were explored respectively according to typical attributes (gender/age/occupation) corresponding to significant global spatial autocorrelation. From Moran scatterplot of inner output cases in various districts of Beijing (Figure 9(a)) and distribution of risk types (Figure 9(b)), high-risk regions of inner output cases were concentrated primarily in urban centers including Haidian, Chaoyang, Dongcheng, Xicheng and Fengtai districts, secondary high-risk regions were located in such urban areas as Chongwen, Xuanwu, Shijingshan districts as well as Changping in the northwest, and relatively safe region with low-risk was Miyun in the suburb northeast whereas other regions had no significant local spatial autocorrelation due to too small number of cases. The primary and secondary high-risk regions were located in the city center and formed a significant cluster, and only few suburbs in the northwest suffered from the virus high-risk spread while the northeast suburbs were low-risk and relatively safe. Above results showed that the overall prevention and control measures were remarkably effective. Under the assurance of that high-risk areas located in urban centers, control measures should be strengthened in suburbs in the northwest, and given the limited resources, measures for suburbs in the northeast could be lowered moderately.

Figure 9(c) and (d) shows that risk regional distribution of female cases was slightly stronger than that of male cases, which was consistent with gender distribution of global spatial autocorrelation, and distributions of high-risk areas of female and male cases were unanimous while that of male cases was relatively more concentrated. Figure 9(e) and (f) illustrates that high-risk areas of cases aged from 20 to 40 were concentrated in two or three districts in urban centers and were without spread trend, and high-risk areas of cases aged from 40 to 60 covered most of the administrative districts in urban centers and were with apparent tendency of spreading high-risk to the suburbs in the northwest and southeast, indicating that isolation and control measures for young population were relatively effective and middle-aged population which had the consistent hotspot distribution with the overall distribution was the main spread source of SARS epidemic. Figure 9(g) and (h) indicates that high risk areas of health care personnel were concentrated in a few regions in urban centers and were with only a weak spread trend to the southeast which showed isolation measures for nosocomial infection was significantly effective, and high risk areas of labor/peasants/workers which were high-risk spread source were located mainly in urban centers and were with a fixed pattern of spread trend.

Considering the evolvement mechanism of risk types of districts or counties with significant local spatial autocorrelation in SARS epidemic in Beijing, we obtained some findings according to Table 4: (1) urban central districts including Haidian, Chaoyang and Dongcheng were high-risk hotspots in SARS epidemic, and after being relatively safe only during the early two or three weeks, continued to be under high-risk; (2) Dongcheng and Fengtai districts were overall high-risk areas with transformation of low-high-low in risk levels, and the corresponding high-risk levels remained continuously during the mid-late epidemic period; (3) Chongwen and Xuanwu in urban centers were L-H regions continuously and the secondary hotspots surrounded by high-risk areas; (4) Shijingshan as an urban central district was considered as a secondary hotspot with L-H level only at the early epidemic period, and Changping located in suburb northwest continued to be a secondary hotspot during the mid-late period; and (5) Miyun as the only cold spot continued its low-risk level in only about two weeks during the mid-late period, and other insignificant regions only seemed to be secondary hotspots shortly in time sequence. Overall, control measures during the whole epidemic period should focus primarily on urban central districts including Haidian, Chaoyang and Dongcheng and secondarily on another two central districts including Chongwen and Xuanwu, whereas attention should also be paid to Shijingshan district in the early epidemic period, and during the mid-late period, concentrated on urban central districts including Dongcheng and Fengtai as well as northwest suburbs including Changping, while simultaneously the strength of control measures to northeast suburbs

Figure 9 Spatial characteristics and risk types of inner output cases in various districts of Beijing. (a) Moran scatterplot (all cases). Risk types: (b) all cases; (c) male; (d) female; (e) 20<age≤40; (f) 40<age≤60; (g) health care personnel; (h) labor/peasants/workers.

Table 4 Temporal evolution process of risk types of inner output cases in various districts^{a)}

District/county	Week 1		Week 2		Week 3		Week 4		Week 5		Week 6		Week 7		Entire risk
Miyun	0	→	0	→	0	→	0	→	4	→	4	→	0	→	4
Changping	0	→	0	→	2	→	2	→	2	→	2	→	2	→	2
Shunyi	2	→	0	→	0	→	0	→	0	→	0	→	0	→	0
Haidian	0	→	0	→	1	→	1	→	1	→	1	→	1	→	1
Chaoyang	1	→	0	→	1	→	1	→	1	→	1	→	1	→	1
Shijingshan	0	→	2	→	2	→	0	→	0	→	0	→	0	→	2
Fengtai	0	→	2	→	2	→	0	→	0	→	1	→	0	→	1
Daxing	0	→	0	→	0	→	0	→	2	→	2	→	2	→	0
Chongwen	2	→	2	→	2	→	2	→	2	→	2	→	2	→	2
Dongcheng	0	→	0	→	2	→	1	→	1	→	1	→	2	→	1
Xicheng	2	→	2	→	1	→	1	→	1	→	1	→	1	→	1
Xuanwu	2	→	0	→	2	→	2	→	2	→	2	→	2	→	2

a) 0: Not significant; 1, H-H (Hotspots); 2, L-H; 3, H-L; 4, L-L; (b) regions of consistent insignificant autocorrelation: Yanqing/Huairou/Pinggu/Mentougou/Tongzhou/Fangshan.

including Miyun could be reduced appropriately.

4.3.3 Characteristic analysis of inner transmission network

Based on data of inner input, output and self spread cases of SARS epidemic in Beijing in 2003, the transmission network of inner SARS epidemic could be built and shown in Figure 10. The proposed network was a double-directed and weighted network with 18 administrative districts of Beijing as nodes and with edges indicating the input/output cases between nodes, and there were two directed edges between two nodes whereas the direction shows the input or output of SARS cases and the weight value indicates the number of input or output cases. In addition, there were rings in the network structure because some nodes were connected to themselves indicating self spread cases of districts. Due to the close characteristic of transmission network, either of total numbers of in-degree and out-degree was 149, and average degree of the network was 8.2778, which indicated that the average number of districts with input/output interaction was about 8. Values of overall clustering coefficients C and C^* of the network were 0.4288 and 0.3143 respectively, indicating that information flow was more aggregated than material flow in inner transmission network of SARS epidemic. In other words, compared with the amount of districts with input/output interaction, input/output cases were more concentrated in few districts.

Degrees and clustering coefficients of nodes representing districts were calculated by program coding, and the results are shown in Table 5. Some important findings could be obtained: (1) besides several districts in urban centers connecting a great number of other nodes, suburbs including Tongzhou and Changping had high-value in-degree and out-degree as well, indicating that such nodes were potentially high-risk areas, and moreover, values of clustering coefficients C_i and C_i^* of Tongzhou and Changping were slightly higher, indicating that their connecting nodes might easily form local small-world for disease outbreak and should be strengthened in control measures; (2) there was certain heterogeneity for the information flow of input/output interaction, and the numbers of edges for input and output cases to a certain node were not identical; for instance, suburb nodes including Shunyi, Huairou, Mentougou and Miyun had only 0–2 input nodes but 3–10 output nodes, and suburb nodes including Tongzhou and Fangshan had 10–15 input nodes but only 6–10 output nodes, indicating that various districts had heterogeneity in information flow and their corresponding control measures should be different, which means different levels of control measures to various districts should be implemented according to different information flows in order to reduce the resource costs and improve the control efficiency; (3) values of C_i and C_i^* of most districts were consistent, which showed that there were significant small-world clustering characteristics

Figure 10 Demonstration of inner transmission network of SARS epidemic in Beijing.

Table 5 Node degrees and clustering coefficients of various districts/counties

District/county	In-degree	Out-degree	C_i	C_i^*
Changping	17	11	0.3860	0.6389
Chaoyang	16	13	0.4167	0.5522
Chongwen	10	9	0.6556	0.5885
Daxing	5	9	0.6000	0.0903
Dongcheng	12	10	0.5606	0.5862
Fangshan	10	6	0.5778	0.4338
Fengtai	14	12	0.4176	0.4924
Haidian	13	12	0.4103	0.3623
Huairou	0	6	0.0000	0.0000
Mentougou	1	4	0.0000	0.0000
Miyun	1	3	0.0000	0.0000
Pinggu	2	2	0.5000	0.0012
Shijingshan	7	8	0.6429	0.3927
Shunyi	2	10	0.0000	0.0000
Tongzhou	15	10	0.4190	0.5428
Xicheng	9	12	0.5417	0.2696
Xuanwu	12	9	0.5909	0.6249
Yanqing	3	3	1.0000	0.0821

in the transmission network, and a small number of districts carried majority of information and material flows in SARS epidemic; (4) there was certain heterogeneity for values of C_i and C_i^* of few districts; for instance, suburbs including Yanqing, Daxing and Pinggu had high C_i values of 1.0000, 0.6000 and 0.5000 but had low C_i^* values of 0.0821, 0.0903 and 0.0012, indicating that although there was strong interaction of information flow between the connecting nodes around these three districts, the corresponding material flow was relatively weak, which made the small world around them a relatively cold area; and (5) values of C_i and C_i^* of Shunyi, Huairou, Mentougou and Miyun were zero, indicating that no small-world characteristics existed in local transmission network centered in them and there was no material flow carried by them, and therefore they could be considered as absolute cold regions of SARS epidemic and the strength of their control measures could be reduced to effectively save resource costs.

5 Conclusions

Researches of spatiotemporal characteristics and pattern of epidemic spread refer to many fields such as epidemiology, spatial information science, spatial statistical analysis, sociology, and anthropology. Numerous researches have provided detailed explanations to epidemic spread mechanism, spatiotemporal distribution, prevention/control measures and etc. However, there are still many unsolved mysteries about epidemic spread. Retrospective analysis based on existed epidemic spread could help analyze and interpret its spread mechanism and spatiotemporal characteristics, and provide good theoretical and practical supports for the prevention and control of novel infectious diseases in future.

SARS epidemic in China in 2003 indicated the shortages of our epidemiological studies and the limitation of prevention/control measures with public health emergency, and meanwhile attracted worldwide interests and attentions for researchers to epidemiology. Recent studies on SARS epidemic spread have been extensive and effective, but there is still a lack of understanding and explanation of the spatial-temporal change mode of individual susceptible-infected-symptomatic-treated-recovered epidemic progress and the characteristics of information/material flow in the epidemic spread network between regions. SARS spread in-out flow was presented in this paper, which could effectively explain the transformation mechanism of individual location and the characteristics of information/material flow in the dynamic spread network, based on the SARS epidemiological investigation data in China from 2002 to 2003 and by the methods of spatial-temporal statistical analysis and network characteristic analysis, we explored and analyzed comprehensively epidemiological characteristics, spatial-temporal spread mechanism and high-risk hotspots in outer in-out flow, and spatial-temporal evolutive rules and structure characteristics in inner in-out flow, as well as evaluated and suggested prevention and control measures of SARS epidemic in Beijing. Compared with analysis methods based on infected cases or close contacts, the exploration and analysis in this paper from a new perspective help better detect and discover the potential spatial-temporal evolutive rules and characteristics of SARS epidemic and provide a more effective theoretical basis for emergency/control measures to other novel infectious diseases.

This work was supported by National Natural Science Foundation of China (Grant Nos. 40871181 and 41101369), Key Knowledge Innovative Program of Chinese Academy of Sciences (Grant No. KZCX2-EW-318), Jiangxi Provincial Natural Science Foundation (Grant No. 20114BAB215024), Natural Science Youth Foundation of Jiangxi Provincial Office of Education (Grant No. GJJ11073) and Open Foundation of Key Laboratory of Poyang Lake Wetland and Watershed Research, Ministry of Education (Grant No. PK2010001).

- 1 Dye C, Gay N. Epidemiology modeling the SARS epidemic. *Science*, 2003, 300: 1884–1885
- 2 Shi Y L. Stochastic dynamic model of SARS spreading. *Chin Sci Bull*, 2003, 48: 1287–1292
- 3 Jia N, Tsui L. Epidemic modelling using SARS as a case study. *North Amer Actuar 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 Pang X H, Liu D L, Gong X H, et al. Study on risk factors related to severe acute respiratory syndrome among close contactors in Beijing (in Chinese). *Chin J Epidemiol*, 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. *Publ Health*, 2005, 119: 1080–1087
- 11 Cao Z D, Zeng D J, Zheng X L, et al. Spatio-temporal evolution of Beijing 2003 SARS epidemic. *Sci China Earth Sci*, 2010, 53: 1017–1028
- 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 Gong J H, Sun Z L, Li X W, et al. Simulation and analysis of control of severe acute respiratory syndrome (in Chinese). *J Remote Sensing*, 2003, 7: 260–265
- 14 Gong J H, Zhou J P, Xu S, et al. Dynamics model and multi-agent based simulation of SARS transmission (in Chinese). *J Remote Sensing*, 2006, 10: 829–835
- 15 Lin G J, Jia X, Ouyang Q. Predict SARS infection with the small world network model (in Chinese). *J Peking Univ (Health Sci)*, 2003, 35(Suppl): 66–69
- 16 Small M, Tse C K. Small world and scale free model of transmission of SARS. *Int J Bifurcation Chaos*, 2005, 15: 1745–1755
- 17 Wang J F, Meng B, Zheng X Y, et al. Analysis on the multi-distribution and the major influencing factors on severe acute respiratory syndrome in Beijing (in Chinese). *Chin J Epidemiol*, 2005, 26: 16–20
- 18 Liu Y L, Yan S Y, Li X W, et al. Study on population migration characteristics in mainland China and its applications to decision-making for SARS control (in Chinese). *J Remote Sensing*, 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 Li X W, Cao C X, Chang C Y. The first law of geography and spatial-temporal proximity (in Chinese). *Chin J Nat*, 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 Stat: 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