scientific reports

Check for updates

OPEN

A stochastic model explains the periodicity phenomenon of influenza on network

Hong Yang[1,2] & Zhen Jin[1,2]✉

Influenza is an infectious disease with obvious periodic changes over time. It is of great practical significance to explore the non-environment-related factors that cause this regularity for influenza control and individual protection. In this paper, based on the randomness of population number and the heterogeneity of population contact, we have established a stochastic infectious disease model about influenza based on the degree of the network, and obtained the power spectral density function by using the van Kampen expansion method of the master equation. The relevant parameters are obtained by fitting the influenza data of sentinel hospitals. The results of the numerical analysis show that: (1) for the infected, the infection period of patients who go to the sentinel hospitals is particularly different from the others who do not; (2) for all the infected, there is an obvious nonlinear relationship between their infection period and the visiting rate of the influenza sentinel hospitals, the infection rate and the degree. Among them, only the infection period of patients who do not go to the sentinel hospitals decreased monotonously with the infection rate (increased monotonously with the visiting rate), while the rest had a non-monotonic relationship.

Influenza is an respiratory infection caused by a virus. Since the outbreak of the Spanish flu in 1918, there has been a large amount of literature on influenza, among which mathematical model is a very important tool for studying influenza[1–4]. The traditional deterministic mean-field model[5–8] regards different populations as uniformly mixed and ignores the differences in the contact process and behavior among individuals. In general, this contact process can be viewed as a network in which nodes represent individuals and edges represent contacts between individuals. Network has its unique topology structure, such as degree k (i.e. the number of edges connecting to one node), degree distribution (i.e. $P(k)$ denotes the probability that a randomly chosen node in the network has degree k), clustering coefficient (i.e. the average probability that two neighbors of a node are themselves neighbors) and so on[9,10]. The topological structure of complex network may lead to some results of overthrowing the traditional, for instance, there is no epidemic threshold in a scale-free network which follows a degree distribution of power-law ($P(k) \propto k^{-\gamma}$ when $1 < \gamma \leq 3$), for the SIS model[11] raised by Pastor. So far, there are many epidemic disease modelling methods based on different features of the network, such as pairwise model[12], edge-based compartmental modelling[13], degree-based modelling[14] and effective degree modelling[15].

The recurrent of epidemic brings a big challenge for people to treat the disease[16]. Many diseases exhibit the phenomenon of annual, biennial, multi-annual and irregular oscillations, such as measles, whooping cough, influenza and so on. The prediction of these diseases by using deterministic dynamic model satisfactorily both the internal mechanism and persistence properties exhibited by case report remains elusive, especially the multi-year periodicity phenomenon, which can be regarded as the comprehensive results of nonlinear dynamics, random factors and seasonal forcing[17]. Many works of literature related to seasonal cases using the deterministic dynamic model where the contact rate is temporal[18,19], and the period is artificial, yet, the endogenous period is not always fixed at 1 year in reality. By using some methods, such as Fourier transform and wavelet analysis, which are used in the following part, we can get the endogenous frequency of these case data, which better conforms to the actual situation and helpful in controlling a pandemic.

In addition, noise is usually considered a nuisance, yet it is unavoidable in reality. However, many researches show that it may play a constructive role. Noise interacting with the deterministic dynamic system may lead to some different results: (1) extinction[20], the disease may die out even though the basic reproduction number R_0 is larger than one in the deterministic system; (2) fluctuation[21], the autonomous system is allowed to remain

[1]Complex System Research Center, Shanxi University, Taiyuan 030006, Shanxi, China. [2]Shanxi Key Laboratory of Mathematical Techniques and Big Data Analysis on Disease Control and Prevention, Shanxi University, Taiyuan 030006, Shanxi, China. ✉email: jinzhn@263.net

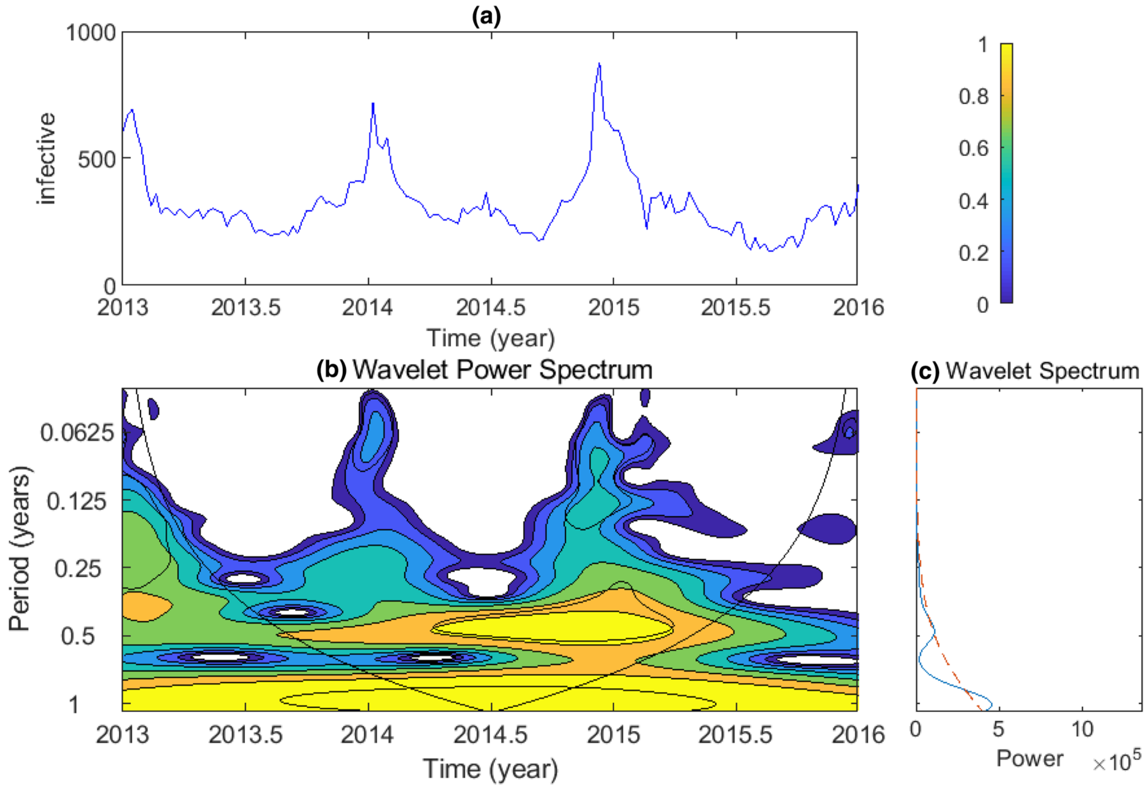

Figure 1. Temporal periodicity analysis of influenza in Taiyuan using the wavelet method. (**a**) Weekly infective number of influenza from 2013 to 2016 of Taiyuan sentinel hospitals. (**b**) The wavelet spectrum analysis corresponds to time series of the data of (**a**). High power values are colored in yellow, orange and cyan denote intermediate power, blue denotes low one. Note the black line is the 95% confidence level. Only patterns within these lines which can neglect the edge effects are considered reliable. (**c**) The average wavelet spectrum (blue line) and the corresponding 95% confidence contour(red).

oscillation around the unique nonzero stable equilibrium, which is one of the main themes of this paper; (3) stochastic phase transition[22], the system is caused to shift between different attractors.

Distinction from the general Susceptible-Infectious-Recovered-Susceptible model for influenza, which is abbreviated as *SIRS*, we divide the infected people into two part: ones who go to the sentinel hospitals (H) and the others (I) who do not. In this paper, we consider the *SHIRS* model based on the degree to describe analytically the fluctuations produced by demographic stochasticity, in which, the method called van Kampen's system-size expansion[23] is used. The power spectral densities (PSD) of the number of infected and susceptible individuals can be derived to describe the effect of stochastic amplification and heterogeneity by using this method, which based on individual-based formulation has been applied to epidemiology[24,25]. However, the researches which involve in the relevant work on network are still relatively few.

The influenza week data from the sentinel hospitals of Taiyuan since 2013–2016 have a period of semi-annual to annual through wavelet method[26,27] illustrated by Fig. 1. To investigate the relationship between the multi-span phenomenon of period and network structure, we first develop a fully stochastic heterogeneous *SHIRS* model. Then, the PSD is obtained from the Langevin equations by Fourier transformation which provides a prediction for the dominant period of the flu, and the sensitivity of parameters is analyzed, in which the values of the corresponding parameters are obtained from the part of data mentioned above by using least square estimation. Finally, we discuss the limitations and extensions of this method. Our analysis sheds new light on the importance of heterogeneity for influenza outbreaks and persistence.

The stochastic SHIRS model on network

Stochastic modelling. In this paper, we divide the population N into four classes, in which S denotes the number of individuals who are susceptible to disease, but are not yet infected at the moment; H or I represents the size of those who are infected with the disease and are able to transmit disease to the susceptible people by contacting with them, the only difference between them is whether to go to the sentinel hospitals, in which H denotes the infected people who have gone to the sentinel hospitals, yet I not; R is the size of people who have recovered from the infected and immune to the disease.

The number of contacts (i.e. the degree) is extraordinary different because of the diversity of factors between people like age, profession and so on. For example: people in school due to the factors of enormous population density has a higher degree. So degree is a important indicator to depict the heterogeneity. Similar to the modelling method proposed by Pastor[11,14], we mark off the whole population into n groups according to the degree,

Event	Transition	Rate
Infection	$S_k \to S_k - 1, H_k \to H_k + 1$	$\rho\beta k S_k \theta$
	$S_k \to S_k - 1, I_k \to I_k + 1$	$(1-\rho)\beta k S_k \theta$
Recovery	$H_k \to H_k - 1$	$\gamma_1 H_k$
	$I_k \to I_k - 1$	$\gamma_2 I_k$
Lose immunity	$S_k \to S_k + 1$	$\alpha(N_k - S_k - H_k - I_k)$

Table 1. List of transition rates.

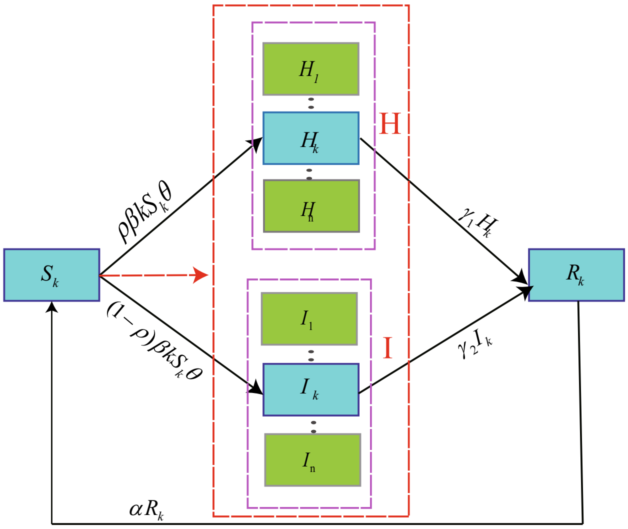

Figure 2. Schematic diagram of the *SHIRS* based on degree, the parameters meaning are listed in Table 2.

i.e. $N = N_1 + ... + N_n$, in which, N_k is the size of population with degree k, $(k = 1, 2, ..., n)$. $P(k) = \frac{N_k}{N}$ is the degree distribution of the network, $\langle k \rangle$ is the mean degree, i.e. $\langle k \rangle = \sum_{k=1}^{n} kP(k)$, similarly, $\langle k^2 \rangle = \sum_{k=1}^{n} k^2 P(k)$ is the secondary moment of the degree distribution.

Based on the classification of disease we mentioned above, people in one group can be divided to four classes: S_k, H_k, I_k, R_k, in which S_k denotes the number of the susceptible people with degree k, and H_k, I_k, R_k have the similar meanings. We assume the total population size and degree distribution do not change with time, which implies that the N_k is a constant. By using the relationship expression: $R_k = N_k - S_k - H_k - I_k$, we can obtain the recovered individual R_k under the new (S_k, H_k, I_k) framework, so the state of the system can be defined by the three integers (S_k, H_k, I_k).

Though the deterministic ordinary differential equation can be viewed as the approximate solution of the stochastic differential equation when the number of population is large, the stochastic modelling method is more accurate, in which continuous time Markov modelling is used in our paper.

At first, we should understand the whole transfer processes of the disease and the corresponding transfer rate, which has been listed in Table 1. Figure 2 has shown the schematic diagram of our model, from which we can know people in one compartment may enter into another due to the three processes:

Infection S_k may become infected by contacting with a infected individual in any group. i.e. $S_k \xrightarrow{\rho\beta k S_k \theta} H_k$, $S_k \xrightarrow{(1-\rho)\beta k S_k \theta} I_k$, where ρ is the proportion of infected people who have go to the sentinel hospitals, β is the probability of infection per contact, if the correlation of the states of the diseases and the degree have been neglected, under this condition, θ which means the probability of any chosen edge being connected to infected has the following form:

$$\theta_A \triangleq \frac{\sum_{m=1}^{n} mH_m}{\sum_{m=1}^{n} mN_m}, \theta_I \triangleq \frac{\sum_{m=1}^{n} mI_m}{\sum_{m=1}^{n} mN_m}, \theta \triangleq \theta_H + \theta_I, \tag{1}$$

Because of the Markov property, the current system state just depends on the one of last time, so we use $T(\delta'|\delta)$ to denote the transfer rate of the system state changing from δ to δ', then the transition rate of the infection process can be written as:

$$\begin{aligned} &T(S_k-1,H_k+1,I_k|S_k,H_k,I_k)=\rho\beta kS_k\theta,\\ &T(S_k-1,H_k,I_k+1|S_k,H_k,I_k)=(1-\rho)\beta kS_k\theta. \end{aligned} \tag{2}$$

Recovery Infective people can move out this compartment because of recovery over time, $H_k \xrightarrow{\gamma_1} R_k, I_k \xrightarrow{\gamma_2} R_k$, where γ_1 and γ_2 are recovery rates. This transition rate can be denoted as:

$$\begin{aligned} &T(S_k,H_k-1,I_k|S_k,H_k,I_k)=\gamma_1 H_k,\\ &T(S_k,H_k,I_k-1|S_k,H_k,I_k)=\gamma_2 I_k. \end{aligned} \tag{3}$$

Lose immunity The recovered people will lose immunity after some time, $R_k \xrightarrow{\alpha} S_k$, in which, α is the rate of lose immunity. The corresponding transition rate of this process is:

$$T(S_k+1,H_k,I_k|S_k,H_k,I_k)=\alpha(N_k-S_k-H_k-I_k). \tag{4}$$

Now, this continuous time Markov process can be modeled using a master equation, which has the general form:

$$\frac{dp(\delta,t)}{dt}=\sum_{\delta'\neq\delta}T(\delta|\delta')p(\delta',t)-\sum_{\delta'\neq\delta}T(\delta'|\delta)p(\delta,t). \tag{5}$$

This equation describes the process of state transition of the system: some system states shift from δ' to δ, meanwhile δ changes to the other system state δ', $p(\delta,t)$ is the probability of the system state in the δ at time t. The explicit form of the master equation involving the three processes is given below:

$$\begin{aligned} \frac{dp(S_k,H_k,I_k,t)}{dt}=\ &T(S_k,H_k,I_k|S_k+1,H_k-1,I_k)p(S_k+1,H_k-1,I_k,t)\\ &+T(S_k,H_k,I_k|S_k+1,H_k,I_k-1)p(S_k+1,H_k,I_k-1,t)\\ &+T(S_k,H_k,I_k|S_k,H_k+1,I_k)p(S_k,H_k+1,I_k,t)\\ &+T(S_k,H_k,I_k|S_k,H_k,I_k+1)p(S_k,H_k,I_k+1,t)\\ &+T(S_k,H_k,I_k|S_k-1,H_k,I_k)p(S_k-1,H_k,I_k,t)\\ &-[T(S_k-1,H_k+1,I_k|S_k,H_k,I_k)+T(S_k,H_k,I_k-1|S_k,H_k,I_k)\\ &+T(S_k,H_k-1,I_k|S_k,H_k,I_k)+T(S_k-1,H_k,I_k+1|S_k,H_k,I_k)\\ &+T(S_k+1,H_k,I_k|S_k,H_k,I_k)]p(S_k,H_k,I_k,t). \end{aligned} \tag{6}$$

The deterministic limit system. From the last section, the master equation has been obtained, which contains all information of the system, in this section, we will study the deterministic limit system corresponding to the (6).

For the mean values

$$\begin{aligned} &\langle S_k\rangle=\sum_{S_k,H_k,I_k}S_kp(S_k,H_k,I_k,t),\ \langle H_k\rangle=\sum_{S_k,H_k,I_k}H_kp(S_k,H_k,I_k,t),\\ &\langle I_k\rangle=\sum_{S_k,H_k,I_k}I_kp(S_k,H_k,I_k,t) \end{aligned}$$

may be obtained by multiplying Eq. (6) by S_k, H_k, I_k in turn, and then summing over all the states of the system, the mean field theory takes the explicit form:

$$\begin{aligned} \frac{d\langle S_k\rangle}{dt}=\sum_{S_k,H_k,I_k}\{&-T(S_k-1,H_k,I_k+1|S_k,H_k,I_k)p(S_k,H_k,I_k,t)\\ &-T(S_k-1,H_k+1,I_k|S_k,H_k,I_k)p(S_k,H_k,I_k,t)\\ &+T(S_k+1,H_k,I_k|S_k,H_k,I_k)p(S_k,H_k,I_k,t)\}, \end{aligned} \tag{7a}$$

$$\begin{aligned} \frac{d\langle H_k\rangle}{dt}=\sum_{S_k,H_k,I_k}\{&T(S_k-1,H_k+1,I_k|S_k,H_k,I_k)p(S_k,H_k,I_k,t)\\ &-T(S_k,H_k-1,I_k|S_k,H_k,I_k)p(S_k,H_k,I_k,t)\}, \end{aligned} \tag{7b}$$

$$\begin{aligned} \frac{d\langle I_k\rangle}{dt}=\sum_{S_k,H_k,I_k}\{&T(S_k-1,H_k,I_k+1|S_k,H_k,I_k)p(S_k,H_k,I_k,t)\\ &-T(S_k,H_k,I_k-1|S_k,H_k,I_k)p(S_k,H_k,I_k,t)\}. \end{aligned} \tag{7c}$$

When we substitute these transition rates of Eqs. (2)–(4) into the above and introduce the corresponding density variables in the limit $N_k\to\infty$,

$$s_k = \lim_{N_k\to\infty}\frac{\langle S_k\rangle}{N_k}, h_k = \lim_{N_k\to\infty}\frac{\langle H_k\rangle}{N_k}, i_k = \lim_{N_k\to\infty}\frac{\langle I_k\rangle}{N_k}. \tag{8}$$

The corresponding deterministic equations of s_k, h_k and i_k can be obtained:

$$\begin{aligned}\frac{ds_k}{dt} =& -\beta k s_k\theta + \alpha(1-s_k-h_k-i_k) = f_1(s_k,h_k,i_k),\\ \frac{dh_k}{dt} =& \rho\beta k s_k\theta - \gamma_1 h_k = f_2(s_k,h_k,i_k),\\ \frac{di_k}{dt} =& (1-\rho)\beta k s_k\theta - \gamma_2 i_k = f_3(s_k,h_k,i_k),\end{aligned} \tag{9}$$

in which θ has the following equivalent form:

$$\theta_h = \theta_H = \sum_{m=1}^{n}\frac{mP(m)h_m}{\langle k\rangle}, \theta_i = \theta_I = \sum_{m=1}^{n}\frac{mP(m)i_m}{\langle k\rangle}, \theta = \theta_h + \theta_i. \tag{10}$$

It is easy to obtain that for the system (9), the basic reproduction number $R_0 = [\frac{\rho\beta}{\gamma_1} + \frac{(1-\rho)\beta}{\gamma_2}]\frac{\langle k^2\rangle}{\langle k\rangle}$ is the critical threshold, which means the average number of secondary infections caused by one infected individual in a susceptible population.

Theorem 2.1 *For the system (9), the following situations hold*

(1) *when $R_0 < 1$, the disease-free equilibrium $E^0(1,0,0)$ is locally asymptotically stable*;
(2) *when $R_0 > 1$, the system has a unique epidemic equilibrium $E^*(s_k^*, h_k^*, i_k^*)$ in which*,

$$s_k^* = \frac{\gamma_1\gamma_2}{\beta k\theta_{h_k^*}(\rho\gamma_2 + (1-\rho)\gamma_1)}h_k^*, i_k^* = \frac{(1-\rho)\gamma_1}{\rho\gamma_2}h_k^*,$$

h_k^* has no concrete expression, but it satisfies the conditions below:

$$1 = h_k^*[1 + \frac{\gamma_1}{\rho\alpha} + \frac{(1-\rho)\gamma_1}{\rho\gamma_2} + \frac{\gamma_1\gamma_2}{\beta k\theta_{h_k^*}(\rho\gamma_2 + (1-\rho)\gamma_1)}] \tag{11}$$

and the disease is permanent.

The proof of the theorem is given in the supplementary, and in which

$$\theta_h^* = \sum_{m=1}^{n}\frac{mP(m)h_m^*}{\langle k\rangle}, \theta_i^* = \sum_{m=1}^{n}\frac{mP(m)i_m^*}{\langle k\rangle}, \theta^* = \theta_h^* + \theta_i^*.$$

The periodicity in different groups. Through the above analysis, we know that for the deterministic system (9), when $R_0 > 1$, the disease ultimately fixed in the point of E^*, yet, the influenza exhibits oscillating phenomenon in reality by the influence of noise even neglecting the seasonal factor, which can be seen from the Fig. 3. Hence we use the van Kampen's system-size expansion method to obtain the higher-order terms which can be used to investigate the perturbations around the steady-state solution of the deterministic system.

Firstly, the new continuous random variables x_k, y_k, z_k are brought in , which have the following relationship between the discrete variables S_k, H_k, I_k and the corresponding density variables s_k, h_k, i_k

$$\begin{aligned}S_k/N_k &= s_k + x_k/\sqrt{N_k},\\ H_k/N_k &= h_k + y_k/\sqrt{N_k},\\ I_k/N_k &= i_k + z_k/\sqrt{N_k}.\end{aligned} \tag{12}$$

The variables x_k, y_k, z_k are corrections of s_k, h_k, i_k, which are in ${N_k}^{-\frac{1}{2}}$ terms and can be viewed as the fluctuations around the epidemic equilibrium to be the order of ${N_k}^{-1}$, from the aspect of central-limit theorem.

Second, in order to study the property of the fluctuation, we define a new probability distribution function $\pi(x_k, y_k, z_k, t) = p(S_k, H_k, I_k, t)$, the following equation can be obtained by using the chain rule:

$$\frac{dp}{dt} = \frac{\partial\pi}{\partial t} - \sqrt{N_k}\frac{\partial s_k}{\partial t}\times\frac{\partial\pi}{\partial x_k} - \sqrt{N_k}\frac{\partial h_k}{\partial t}\times\frac{\partial\pi}{\partial y_k} - \sqrt{N_k}\frac{\partial i_k}{\partial t}\times\frac{\partial\pi}{\partial z_k}. \tag{13}$$

The detailed derivation process of (13) is given in the supplementary.

By introducing the step operators below,

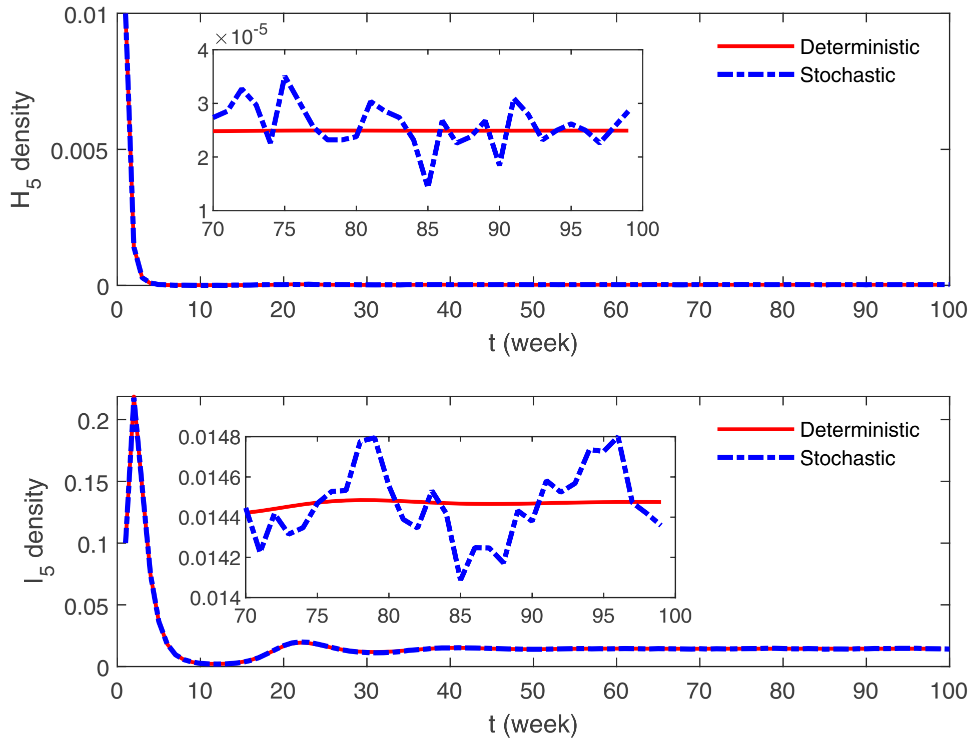

Figure 3. A realization of a stochastic *SHIRS* model with the degree of 5 and its deterministic counterpart. The values of parameters used in this simulation are listed in Table 2. The stochastic simulation is implemented by the event algorithm of Gillespie with transition rates listed in Table 1. The deterministic curve is generated according to the Eq. (9).

$$
\begin{aligned}
\varepsilon_{S_k}^{\pm 1} f(S_k, H_k, I_k) &= f(S_k \pm 1, H_k, I_k),\\
\varepsilon_{H_k}^{\pm 1} f(S_k, H_k, I_k) &= f(S_k, H_k \pm 1, I_k),\\
\varepsilon_{I_k}^{\pm 1} f(S_k, H_k, I_k) &= f(S_k, H_k, I_k \pm 1),
\end{aligned}
\tag{14}
$$

the Eq. (6) with transition rates can be rewritten as

$$
\begin{aligned}
\frac{dp(S_k, H_k, I_k, t)}{dt} =& [(\varepsilon_{S_k}^{+1}\varepsilon_{H_k}^{-1} - 1)T(S_k - 1, H_k + 1, I_k | S_k, H_k, I_k)\\
&+ (\varepsilon_{S_k}^{+1}\varepsilon_{I_k}^{-1} - 1)T(S_k - 1, H_k, I_k + 1 | S_k, H_k, I_k)\\
&+ (\varepsilon_{H_k}^{+1} - 1)T(S_k, H_k - 1, I_k | S_k, H_k, I_k)\\
&+ (\varepsilon_{I_k}^{+1} - 1)T(S_k, H_k, I_k - 1 | S_k, H_k, I_k)\\
&+ (\varepsilon_{S_k}^{-1} - 1)T(S_k + 1, H_k, I_k | S_k, H_k, I_k)]p(S_k, H_k, I_k, t).
\end{aligned}
\tag{15}
$$

Expanding the step operators $\varepsilon_{S_k}^{\pm 1}$, $\varepsilon_{H_k}^{\pm 1}$ and $\varepsilon_{I_k}^{\pm 1}$ in a power series in $N_k^{-\frac{1}{2}}$

$$
\begin{aligned}
\varepsilon_{S_k}^{\pm 1} =& 1 \pm \frac{1}{\sqrt{N_k}}\frac{\partial}{\partial x_k} + \frac{1}{2N_k}\frac{\partial^2}{\partial x_k^2},\\
\varepsilon_{H_k}^{\pm 1} =& 1 \pm \frac{1}{\sqrt{N_k}}\frac{\partial}{\partial y_k} + \frac{1}{2N_k}\frac{\partial^2}{\partial y_k^2},\\
\varepsilon_{I_k}^{\pm 1} =& 1 \pm \frac{1}{\sqrt{N_k}}\frac{\partial}{\partial z_k} + \frac{1}{2N_k}\frac{\partial^2}{\partial z_k^2},
\end{aligned}
\tag{16}
$$

and then substituting the (2)–(4), (12) and (16) into (15), making a comparison with (13) order by order yields the so-called macroscopic Eq. (9) to leading order and the following Fokker-Plance equation (FPE):

$$
\begin{aligned}
\frac{\partial \pi}{\partial t} =& -A_{11}\frac{\partial (x_k\pi)}{\partial x_k} - A_{12}\frac{\partial (y_k\pi)}{\partial x_k} - A_{13}\frac{\partial (z_k\pi)}{\partial x_k} - A_{21}\frac{\partial (x_k\pi)}{\partial y_k} - A_{22}\frac{\partial (y_k\pi)}{\partial y_k} \\
& - A_{23}\frac{\partial (z_k\pi)}{\partial y_k} - A_{31}\frac{\partial (x_k\pi)}{\partial z_k} - A_{32}\frac{\partial (y_k\pi)}{\partial z_k} - A_{33}\frac{\partial (z_k\pi)}{\partial z_k} + \frac{1}{2}B_{11}\frac{\partial^2 \pi}{\partial x_k^2} \\
& + \frac{1}{2}B_{12}\frac{\partial^2 \pi}{\partial x_k y_k} + \frac{1}{2}B_{13}\frac{\partial^2 \pi}{\partial x_k z_k} + \frac{1}{2}B_{21}\frac{\partial^2 \pi}{\partial y_k x_k} + \frac{1}{2}B_{22}\frac{\partial^2 \pi}{\partial y_k^2} + \frac{1}{2}B_{23}\frac{\partial^2 \pi}{\partial y_k z_k} \\
& + \frac{1}{2}B_{31}\frac{\partial^2 \pi}{\partial z_k x_k} + \frac{1}{2}B_{32}\frac{\partial^2 \pi}{\partial z_k y_k} + \frac{1}{2}B_{33}\frac{\partial^2 \pi}{\partial z_k^2},
\end{aligned}
\tag{17}
$$

in which, the matrix $A = (A_{uv})_{3\times 3}$ and $B = (B_{uv})_{3\times 3}$ are evaluated at the epidemic equilibrium E^*, whose concrete values are:

$$
A = (A_{uv})_{3\times 3} = \begin{pmatrix} \frac{\partial f_1}{\partial s_k} & \frac{\partial f_1}{\partial h_k} & \frac{\partial f_1}{\partial i_k} \\ \frac{\partial f_2}{\partial s_k} & \frac{\partial f_2}{\partial h_k} & \frac{\partial f_2}{\partial i_k} \\ \frac{\partial f_3}{\partial s_k} & \frac{\partial f_3}{\partial h_k} & \frac{\partial f_3}{\partial i_k} \end{pmatrix}_{(E^*)},
$$

in which

$$
\begin{array}{lll}
A_{11} = -\beta k\theta^* - \alpha, & A_{12} = -\beta s_k^* \frac{k^2 P(k)}{\langle k\rangle} - \alpha, & A_{13} = -\beta s_k^* \frac{k^2 P(k)}{\langle k\rangle} - \alpha, \\
A_{21} = \rho\beta k\theta^*, & A_{22} = \rho\beta s_k^* \frac{k^2 P(k)}{\langle k\rangle} - \gamma_1, & A_{23} = \rho\beta s_k^* \frac{k^2 P(k)}{\langle k\rangle}, \\
A_{31} = (1-\rho)\beta k\theta^*, & A_{32} = (1-\rho)\beta s_k^* \frac{k^2 P(k)}{\langle k\rangle}, & A_{33} = (1-\rho)\beta s_k^* \frac{k^2 P(k)}{\langle k\rangle} - \gamma_2.
\end{array}
$$

and

$$
B = (B_{uv})_{3\times 3} = \begin{pmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{pmatrix}_{(E^*)},
$$

in which B is a symmetric matrix,

$$
\begin{array}{ll}
B_{11} = \beta k s_k^*\theta^* + \alpha(1 - s_k^* - h_k^* - i_k^*), & B_{12} = -\rho\beta k s_k^*\theta^*, \\
B_{13} = -(1-\rho)\beta k s_k^*\theta^*, & B_{22} = \rho\beta k s_k^*\theta^* + \gamma_1 h_k^*, \\
B_{23} = 0, & B_{33} = (1-\rho)\beta k s_k^*\theta^* + \gamma_2 i_k^*.
\end{array}
$$

which is the microscopic equation to the next-to-leading order. It is obvious that using the system-size expansion methods one can obtain the deterministic equation, which is just the leading order. For the microscopic equation, we want to describe the fluctuations about the stochastic model by using Fourier analyze. For this purpose, the following equivalent Langevin equations (LE) is fairly intuitive than the FPE (17),

$$
\begin{aligned}
\frac{dx_k}{dt} &= A_{11}x_k + A_{12}y_k + A_{13}z_k + \eta_1, \\
\frac{dy_k}{dt} &= A_{21}x_k + A_{22}y_k + A_{23}z_k + \eta_2, \\
\frac{dz_k}{dt} &= A_{31}x_k + A_{32}y_k + A_{33}z_k + \eta_3,
\end{aligned}
\tag{18}
$$

in which the stochastic variables x_k, y_k, z_k are corrections of s_k, h_k, i_k, and η_1, η_2, η_3 are Gaussian white noises whose mean are zero, correlation function are

$$
\langle \eta_u(t)\eta_v(t')\rangle = B_{uv}\delta(t - t')(u, v = 1, 2, 3),
$$

in which $\delta(t - t')$ denotes the Dirac delta function. The equivalence prove of LE and FPE can be found in[28]. Taking the Fourier transform of (18) gets the following results:

$$
\begin{aligned}
-iw\tilde{x}_k &= A_{11}\tilde{x}_k + A_{12}\tilde{y}_k + A_{13}\tilde{z}_k + \tilde{\eta}_1, \\
-iw\tilde{y}_k &= A_{21}\tilde{x}_k + A_{22}\tilde{y}_k + A_{23}\tilde{z}_k + \tilde{\eta}_2, \\
-iw\tilde{z}_k &= A_{31}\tilde{x}_k + A_{32}\tilde{y}_k + A_{33}\tilde{z}_k + \tilde{\eta}_3, \\
< \tilde{\eta}_u(w)\tilde{\eta}_v(w') >&= B_{uv}\delta(w - w')(u, v = 1, 2, 3).
\end{aligned}
\tag{19}
$$

in which, $\tilde{x}_k = \int_{-\infty}^{+\infty} x_k(t)e^{iwt}dt$ and i is imaginary unit, similar to the $\tilde{y}_k$ and $\tilde{z}_k$. Because (18) is a OU process, the corresponding limits of the mean depend on the eigenvalues of A, when $R_0 > 1$, by means of the Fig. 3, we can see the endemic equilibrium E^* is stable, i.e. the eigenvalues of A are negative, so when $t \to \infty, \langle x_k, y_k, z_k\rangle \to \mathbf{0}$.

Solving the equation of (19), we can obtain:

$$\begin{aligned}\tilde{x}_k &= \frac{-(iw)^2\tilde{\eta}_1 + (iw)C_1 + D_{11}\tilde{\eta}_1 + D_{21}\tilde{\eta}_2 + D_{31}\tilde{\eta}_3}{D(w)},\\ \tilde{y}_k &= \frac{-(iw)^2\tilde{\eta}_2 + (iw)C_2 + D_{12}\tilde{\eta}_1 + D_{22}\tilde{\eta}_2 + D_{32}\tilde{\eta}_3}{D(w)},\\ \tilde{z}_k &= \frac{-(iw)^2\tilde{\eta}_3 + (iw)C_3 + D_{13}\tilde{\eta}_1 + D_{23}\tilde{\eta}_2 + D_{33}\tilde{\eta}_3}{D(w)},\end{aligned} \tag{20}$$

where

$$\begin{aligned}C_1 &= -(A_{22} + A_{33})\tilde{\eta}_1 + A_{12}\tilde{\eta}_2 + A_{13}\tilde{\eta}_3,\\ C_2 &= A_{21}\tilde{\eta}_1 - (A_{11} + A_{33})\tilde{\eta}_2 + A_{23}\tilde{\eta}_3,\\ C_3 &= A_{31}\tilde{\eta}_1 + A_{32}\tilde{\eta}_2 - (A_{11} + A_{22})\tilde{\eta}_3,\end{aligned}$$

D_{uv} is the inverse of algebraic complement of the corresponding element A_{uv} of the matrix A, and the expression of denominator is given by

$$\begin{aligned}D(w) &= (iw)^3 + \mathrm{tr}A(\mathrm{iw})^2 + \Omega(\mathrm{iw}) + \det A,\\ \Omega &= A_{11}A_{22} + A_{11}A_{33} + A_{22}A_{33} - A_{13}A_{31} - A_{12}A_{21} - A_{23}A_{32},\end{aligned}$$

where $\mathrm{tr}A$ denotes the trace of A and $\det A$ means the value of the determinant of A.

Averaging the squared moduli of $\tilde{x}_k$, $\tilde{y}_k$ and $\tilde{z}_k$ gives the power-spectrum:

$$\begin{aligned}P_{S_k}(w) &= \langle|\tilde{x}_k|^2\rangle = \frac{B_{11}w^4 + \Lambda_1 w^2 + \Gamma_1}{|D(w)|^2},\\ P_{H_k}(w) &= \langle|\tilde{y}_k|^2\rangle = \frac{B_{22}w^4 + \Lambda_2 w^2 + \Gamma_2}{|D(w)|^2},\\ P_{I_k}(w) &= \langle|\tilde{z}_k|^2\rangle = \frac{B_{33}w^4 + \Lambda_3 w^2 + \Gamma_3}{|D(w)|^2},\end{aligned} \tag{21}$$

$$\begin{aligned}\Lambda_1 &= B_{11}(A_{22}^2 + A_{33}^2) + B_{22}A_{12}^2 + B_{33}A_{13}^2 + 2[B_{11}A_{22}A_{33} - B_{12}A_{12}(A_{22} + A_{33})\\ &\quad - B_{13}A_{13}(A_{22} + A_{33}) + B_{11}D_{11} + B_{12}D_{21} + B_{13}D_{31}],\\ \Lambda_2 &= B_{11}A_{21}^2 + B_{22}(A_{11}^2 + A_{33}^2) + B_{33}A_{23}^2 + 2[B_{13}A_{21}A_{23}\\ &\quad + B_{22}A_{11}A_{33} - B_{12}A_{21}(A_{11} + A_{33}) + B_{21}D_{12} + B_{22}D_{22}],\\ \Lambda_3 &= B_{11}A_{31}^2 + B_{22}A_{32}^2 + B_{33}(A_{11}^2 + A_{22}^2) + 2[B_{33}A_{11}A_{22}\\ &\quad + B_{12}A_{31}A_{32} - B_{13}A_{31}(A_{11} + A_{22}) + B_{31}D_{13} + B_{33}D_{33}],\\ \Gamma_1 &= B_{11}D_{11}^2 + B_{22}D_{21}^2 + B_{33}D_{31}^2 + 2B_{12}D_{11}D_{21} + 2B_{13}D_{11}D_{31},\\ \Gamma_2 &= B_{11}D_{12}^2 + B_{22}D_{22}^2 + B_{33}D_{32}^2 + 2B_{12}D_{12}D_{22} + 2B_{13}D_{12}D_{32},\\ \Gamma_3 &= B_{11}D_{13}^2 + B_{22}D_{23}^2 + B_{33}D_{33}^2 + 2B_{12}D_{13}D_{23} + 2B_{13}D_{13}D_{33},\\ |D(w)|^2 &= (w^3 - \Omega w)^2 + (\det A - w^2\mathrm{tr}A)^2.\end{aligned}$$

Up to this point, the analytical expression of PSD have been derived, because of the difficulty of the calculation in high-dimension coupled system, the concrete expressions of the epidemic equilibrium can't be obtained, so we can analysis the frequency of different groups just depending on the numerical simulation.

Results

In this section, we will use the influenza data of Taiyuan sentinel hospitals in the first quarter of 2013 to fit about our model in order to avoid the influence of environment changing, and then discuss the influence of heterogeneity to the periodicity of influenza.

Estimation of parameters. Based on the statistical bulletion on national economic and social development of Taiyuan, the number of population in 2013 is 4.3 million. Removal rate can be set to $\gamma_1 = 7/3, \gamma_2 = 1$ and the rate of lose immunity is supposed to be $\alpha = 1/52$ according to the mean disease course of influenza. The infection probability β is an estimated parameter. As for the degree distribution, the power-law distribution usually conforms to the features of real world in most cases, hence, $P(k) = 2m^2k^{-\nu}(m = 3, \nu = 3.5)$[29], is used in this paper.

As far as we know, Chinese National Influenza Center (CNIC) has set up 554 sentinel hospitals all over the country. The city of Taiyuan has two of which: Taiyuan central hospital and the first affiliated hospital of Shanxi medical university. We have the influenza data from the two sentinel hospitals, yet, a small enough number compared with those in the whole city of Taiyuan, so ρ as the proportion of Taiyuan influenza patients in sentinel hospitals, is a parameter to be evaluated.

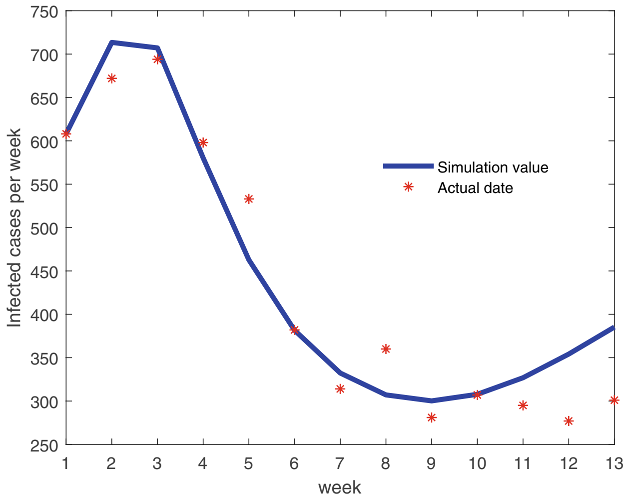

Figure 4. Sentinel hospitals data in the first quarter of 2013 and fitting data by using least-square are showed. The minimum (maximum) degree is 5 (500) and the mean degree is 7.5.

Parameters	Meaning	Value
N	The total number of people	4.3×10^6
γ_1	The recovery rate of H	7/3
γ_2	The recovery rate of I	1
α	The rate of loss immunity	1/52
β	The infection probability per contact	0.8
ρ	The proportion of H in all infected people	0.004

Table 2. Notation for model formulation and parameters value.

We adopt the least-square estimation to obtain the parameters sets $\Lambda = \{\beta, \rho\}$, which minimizes the objective function: $Y(\Lambda) = [Y(t) - \rho N \sum_{k=\min}^{\max} P(k)\beta k s_k \theta]^2$, in which $Y(t)$ denotes the actually influenza data from sentinel hospitals at week t. The Fig. 4 shows that the model data agrees well with the actual data, and the corresponding value of parameters are summarized in Table 2.

Prevalence of influenza in Taiyuan. CNIC has a history of more than 60 years since its foundation in 1957, especially the recently decade, the wholesome national influenza surveillance network system has made the influenza data more transparent. Our influenza data from Taiyuan sentinel hospitals are just a very small part of this system. We use the week data from 2013 to 2016 to detect whether there exists a periodic phenomenon. The data reveals a clear periodicity in the outbreaks of influenza, which can be confirmed through the wavelet power spectrum of Fig. 1. The wavelet analysis can reveal the periodic changing of a time-series, which performs the spectral characteristics as a function of time. The oscillations of influenza have a obvious annual variation, yet, the wavelet analysis based on the data shows another significant periodicity: semi-annual.

Influence of stochasticity and network structure on influenza outbreaks. The difference between the deterministic and stochastic simulation is revealed in Fig. 3, which is a particular case of fluctuation who has been mentioned in introduction. It indicates that the demographic noise and network struction can induce rich periodic phenomenon. We can understand the influence of network struction on the expected fluctuations of influenza via our analysis.

The accuracy of the theoretical analysis consequence of PSD via (21) has been verified with the simulated results in Fig. 5, using the parameters listed in Table 2. We can go further to detect how the period of influenza varies with the changes of degree k by using the formula (21). Based on this, the x-coordinate (i.e. ω) for the maximum value of the PSD can be obtained from the analytical expression, then the corresponding period (i.e. $1/\omega$) for different parameters are plotted in Figs. 6, 7, 8, and 9. In order to have a better visual effect, we just present the degree from 5 to 25 in these figures, on the whole, it does not affect our conclusions, because this group accounts for 99 percent of the total population.

When the $\rho = 0.004$ is fixed, how the period is influenced by the degree and β is showed in Figs. 6 and 7. In Fig. 6, for the H whose degree is less than 16, the period goes up as the β decrease, yet, when the degree is large

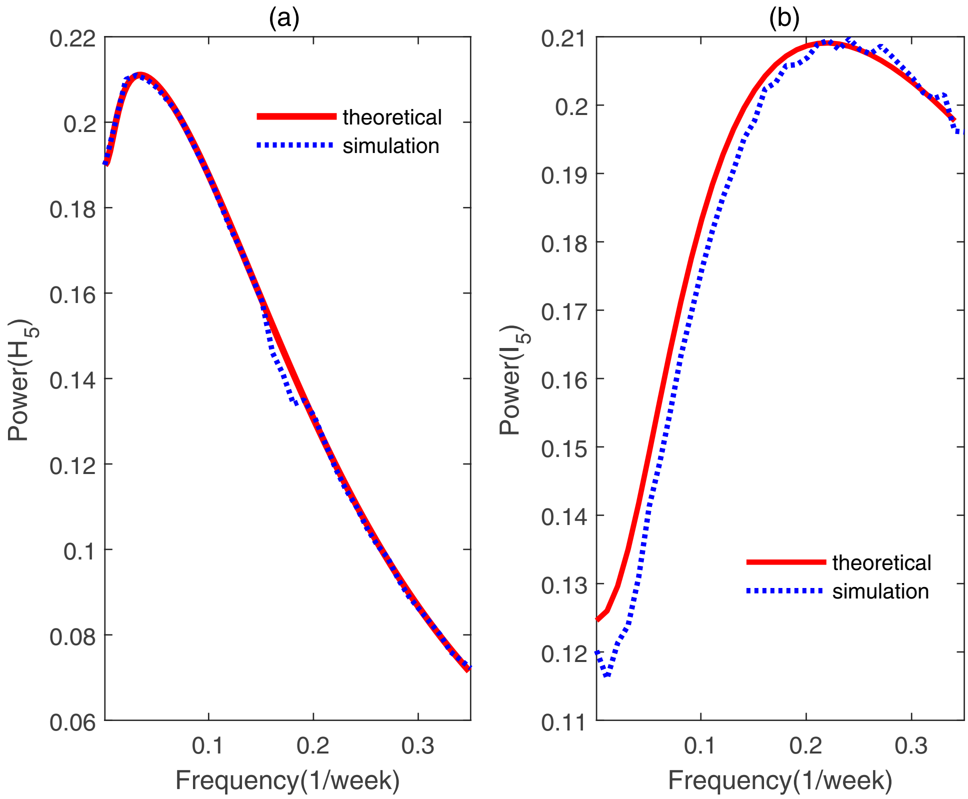

Figure 5. Comparisons between the theoretical prediction of PSD (21) and the average PSD obtained from the stochastic simulation, for the fluctuations of the infected with degree 5.

than 16, the relationship between β and period is nonmonotonic, as β increases, the period decreases and then increases. On the other hand, when $\beta \geqslant 0.4$, as the degree increases, so does the H's period, while $\beta < 0.4$, the period increases at the beginning and then decreases with the augment of the degree. There exists obvious non-linear relationship between the I's period, degree and β. For the fixed degree, increasing the β, the value of period becomes small, yet, when the β is fixed, the relationship between the degree and I's period is non-monotonous. When the degree is small than $\langle k \rangle$, the period becomes larger as the degree increases, yet, the trend is opposite on the other side of $\langle k \rangle$, which can be obtained from the Fig. 7. This means that, for the infected people I whose degree more close to $\langle k \rangle$, the more safer.

When $\beta = 0.8$ is fixed, the relationship between the period, ρ and degree is showed in Figs. 8 and 9. From Fig. 8 we can observe that when $\rho \leqslant 0.4$, for the infected people of H, the degree is more bigger, the period is more larger, which means that people are more safer, yet, when $\rho > 0.4$, with the degree increasing, the period increases firstly and decreases later, which means the people of H whose degree located intermediate is more safer. This is quite opposite for I in Fig. 9. On the other hand, when H's degree is fixed, the period decreases at the beginning and then increases as the ρ increasing. This phenomenon from a side explains that the importance of the medical resources. For the infected people of I, the relationship is different in Fig. 9. With the increase of ρ, which means more and more infected people go to the sentinel hospitals, leading the I becomes more and more safer, i.e. the period is more and more longer, this change is even more pronounced for I with higher degree.

Discussion

In this paper, we have investigated the effect of network structure on the frequency of the influenza outbreaks, focusing on the power-law networks, by adapting the van Kampen expansion approach from *SHIRS* disease to develop a fully stochastic model. This leads to a reasonable explanation for the periodic phenomena of influenza, and helps us to understand the complex fluctuations of influenza in Taiyuan with a dominant period of half years without the external factors of seasonal.

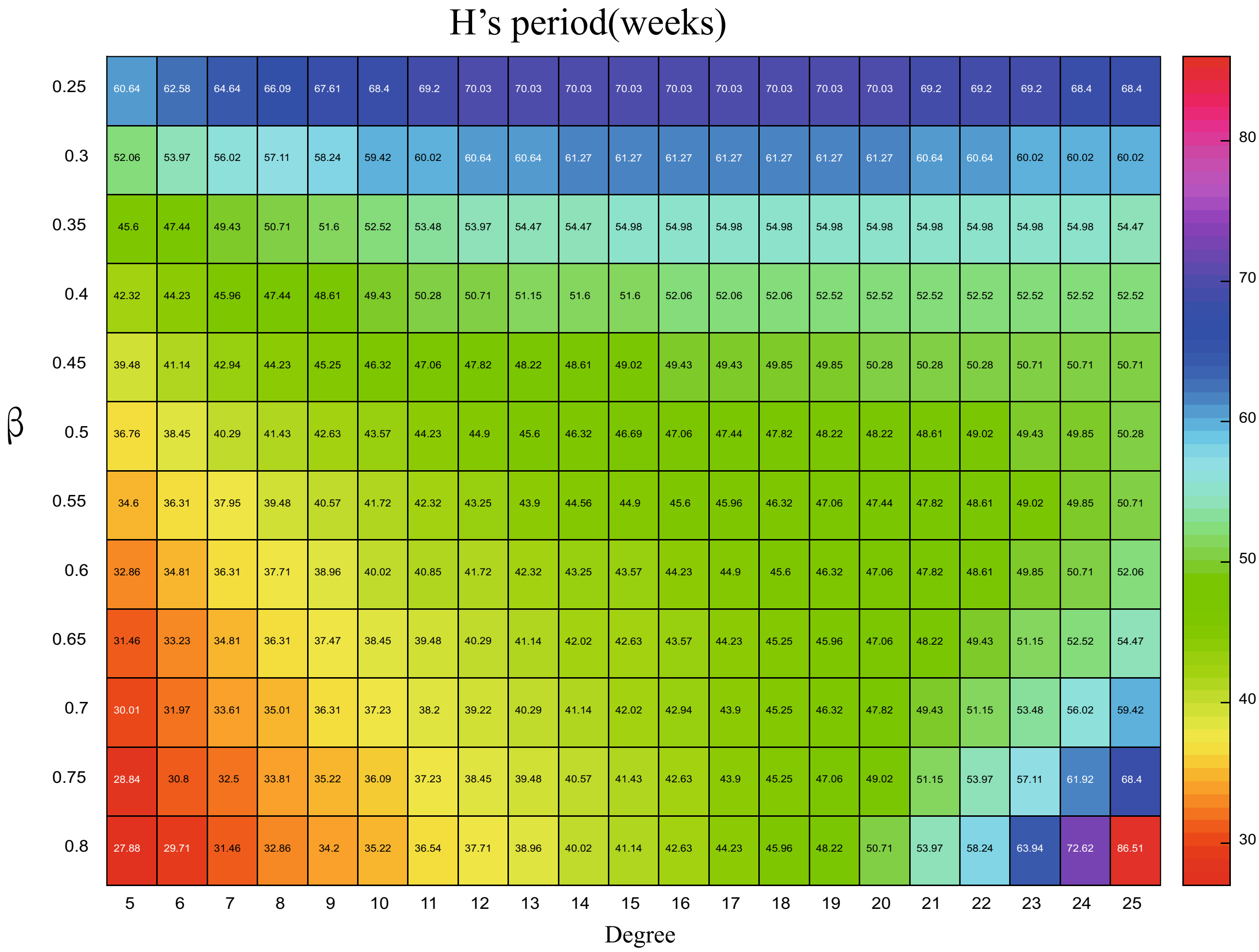

Figure 6. The period of H with different degree and β, when $\rho = 0.004$.

We can have a better understanding of the interaction between the deterministic and stochastic components of the system by using the method of van Kampen expansion. The analytical solution of PSD agrees well with the simulation data. The result reveals that for the infected people who go to the sentinel hospitals, i.e. H, the relationship between the ρ, β and period is nonlinear, even for the hospitals hospitalization rate ρ, just an appropriate one is benefit for the patients. On the other hand, the nonlinear relationship between the three is also suitable for the infected people I in most cases, except that the relationship between ρ (or β) and period is monotonous when other parameters are fixed.

Our work emphasizes the importance of heterogeneous contact network on the periodic outbreak of influenza, which further validates the truth that the combined action of the non-linear system and stochasticity may results many novel phenomena. On the other hand, our study also show that there exists a very close relation between the phenomenon of multi-cycle and the heterogeneity of contact.

Limitations. To obtain the specific analytical solutions are difficult for the high dimensional coupling system, so we can not give the explicit formula about the relationship between the period, degree, β and ρ, and just have a intuitive understanding from the numerical simulation.

The influenza date we obtained from the sentinel hospitals are weekly, so we set the time-boxed intervals measured in week. Comparing to the daily one, the weekly contact network will be not very precise and may cause the repeatedly contact problem, which one we omitted in this paper.

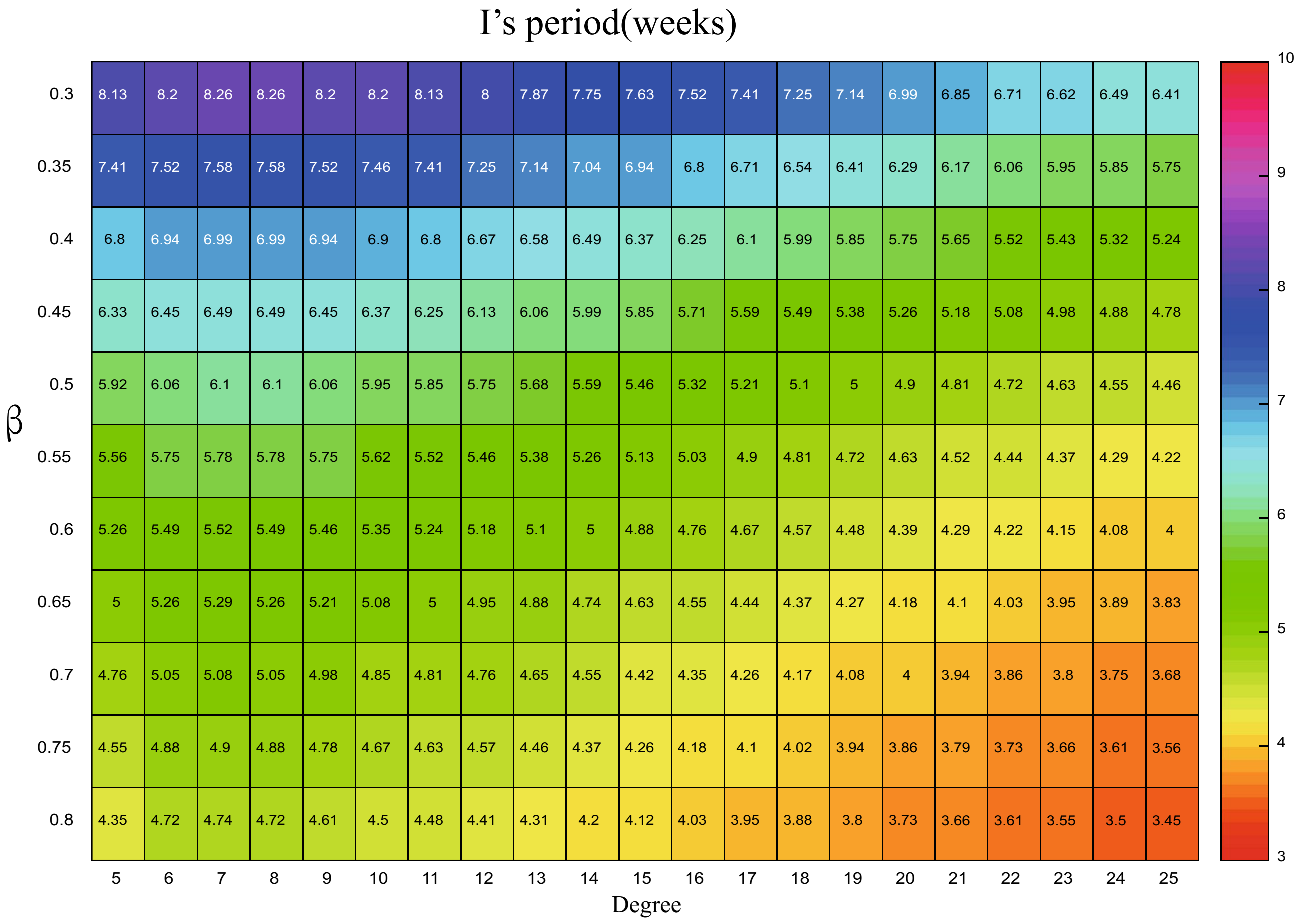

Figure 7. The period of I with different degree and β, when $\rho = 0.004$.

The approach of van Kampen is limited to the following two conditions: a formal treatment of demographic stochasticity; the existence of stable non-zero solution, for simpleness we do not take into account the changing population, seasonal forcing and the correlation between the degree. In essence, the migration of population and behavioral changing in flu season is unavoidable, which is missing in our work.

Possible extensions. It is easy to adapt this method to many diseases exhibited fluctuations, such as malaria, cholera, hand foot and mouth disease and so on. Especially for childhood diseases, similar study may gives a good idea on avoiding frequent illness. At the same time, if there are several strains that coexist for a disease, such as influenza, our method may also deduce the period for each strain, which may be good news for the vaccine's manufacturers.

We can find that from Fig. 10, even among the very nearest cities, like Beijing and Tianjin, Jiangsu and Shanghai, though the climate differences can be ignored, the wavelet power spectrum is greatly different. This may prove the importance of the contact network. On the other hand, even though the climate characteristics of city are different from north part of China (Beijing and Tianjin) to the south part (Jiangsu and Shanghai), it is a surprise phenomenon that Tianjin and Shanghai exhibit the similar periodic law. Furthermore, the same method can also be used to the influenza data of tropical cities, then comparisons with existing temperate data may reveal the dominant factors. Which one is the key factor lead to this? Temperature, humidity, wind speed, haze or human behavior? It deserves our further investigation.

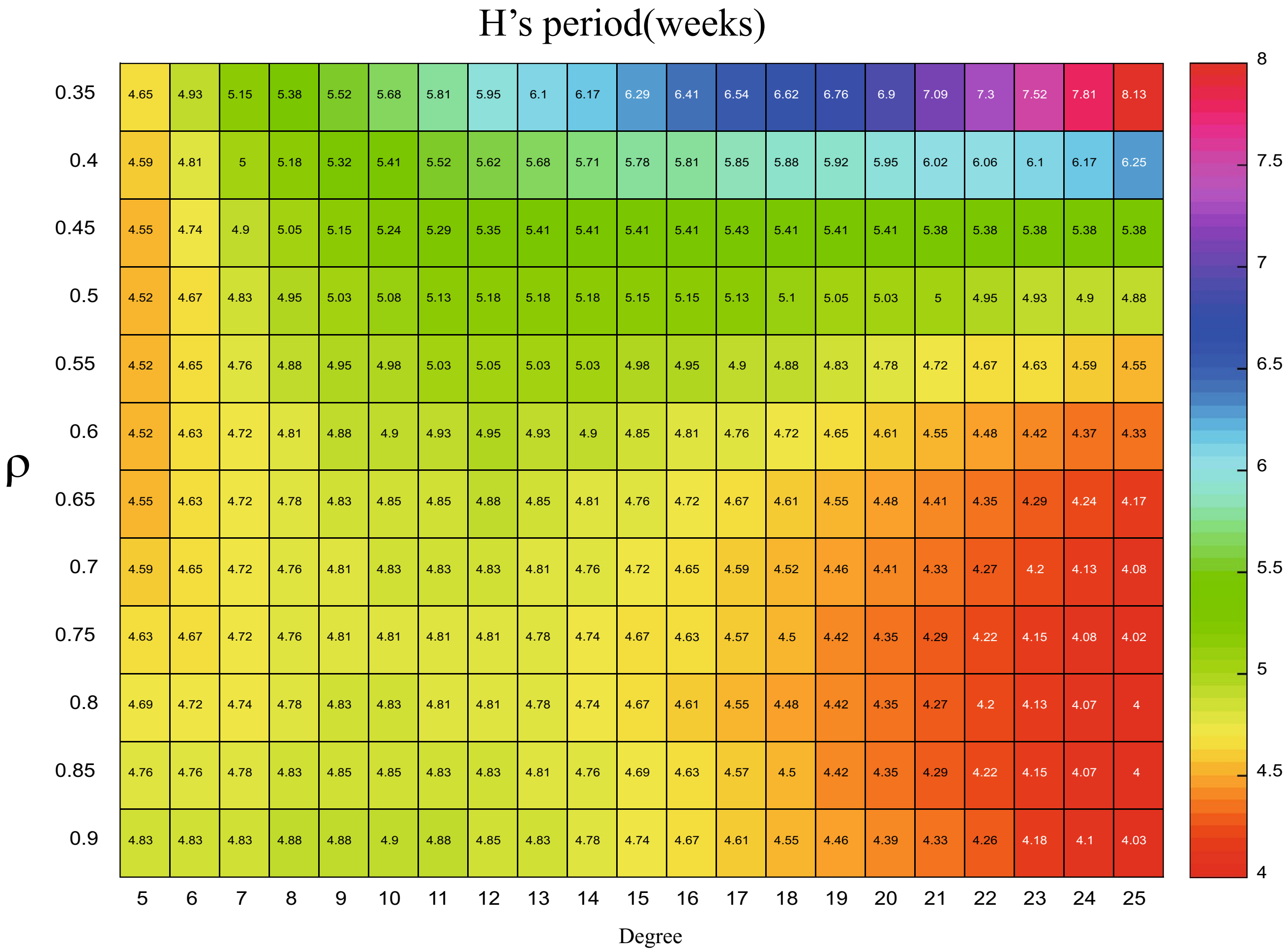

Figure 8. The period of H with different degree and ρ, when $\beta = 0.8$.

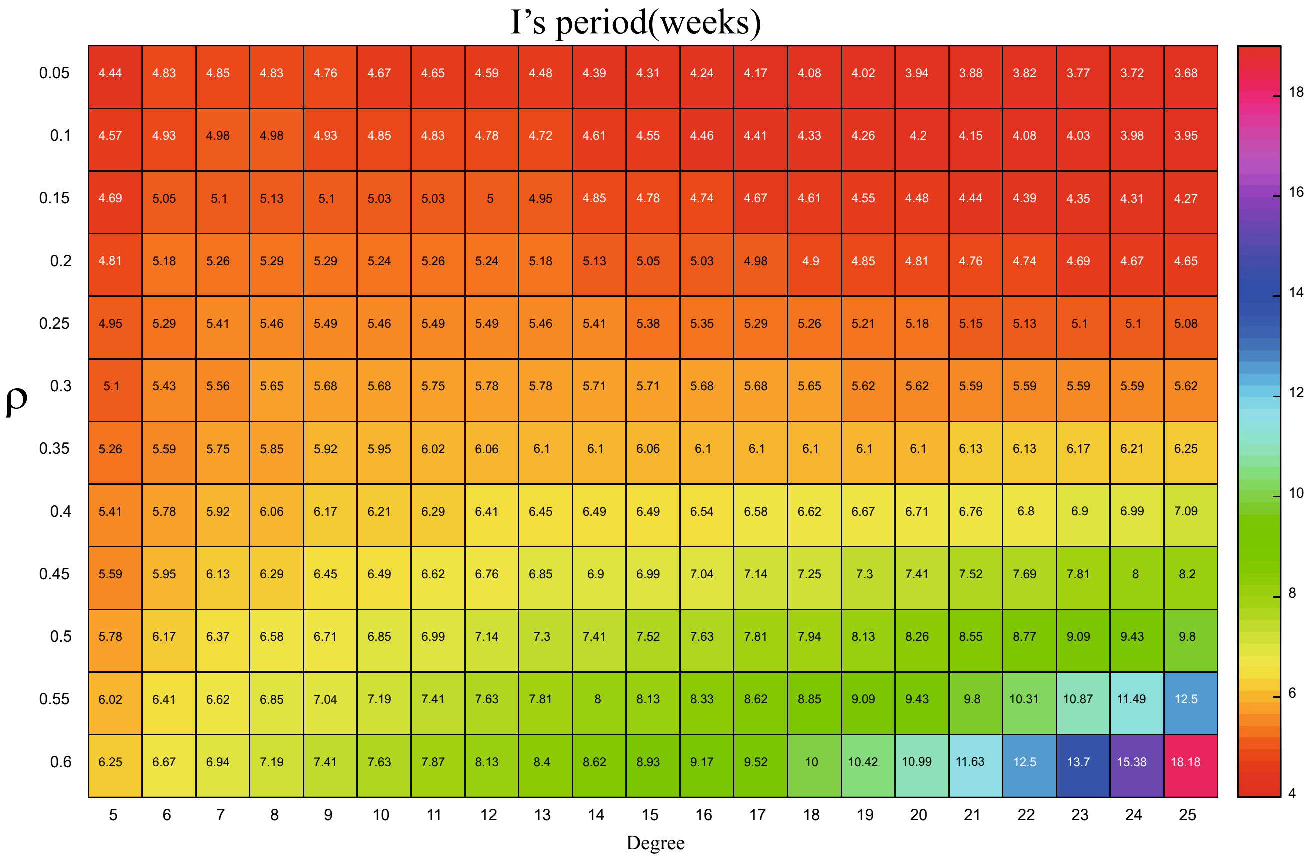

Figure 9. The period of I with different degree and ρ, when $\beta = 0.8$.

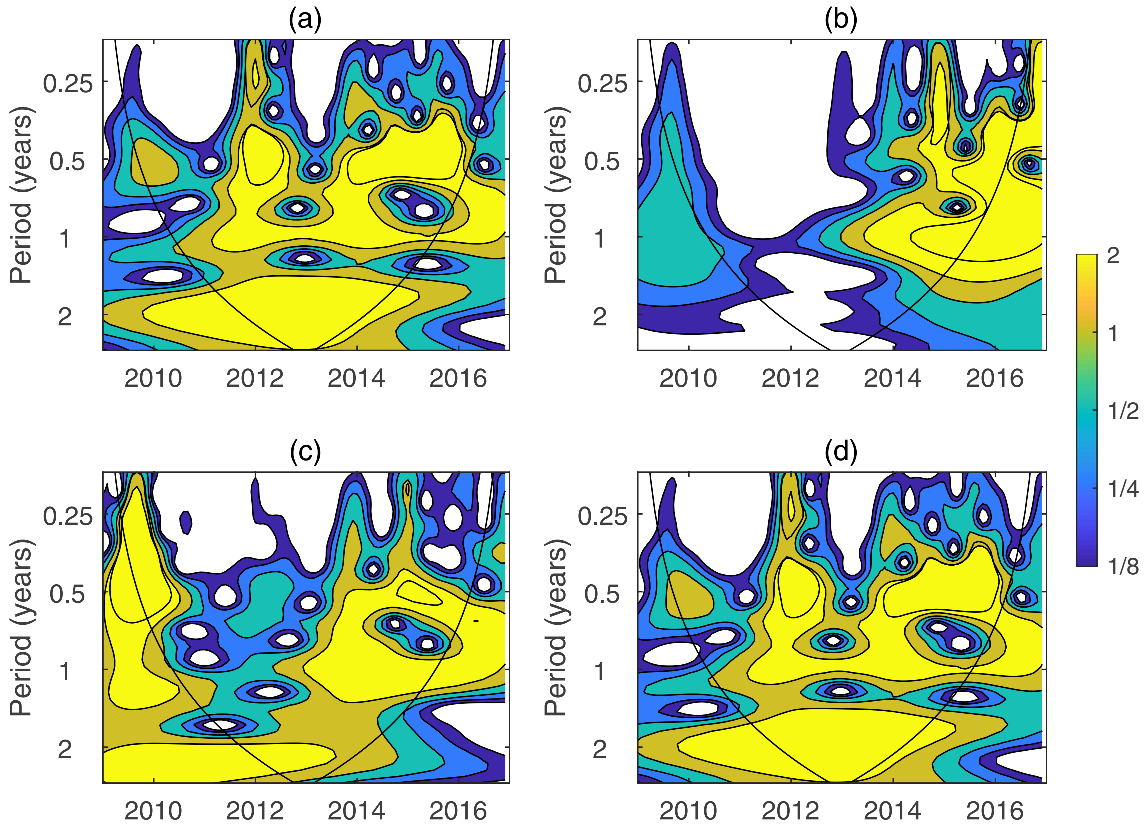

Figure 10. The results of wavelet analysis, where the monthly influenza data of (**a**) Shanghai, (**b**) Beijing, (**c**) Jiangsu province and (**d**) Tianjin obtained from CDC since 2009–2016.

Received: 2 February 2021; Accepted: 24 September 2021
Published online: 25 October 2021

References

1. Colizza, V., Barrat, A., Barthelemy, M. & Vespignani, A. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl. Acad. Sci.* **103**, 2015–2020 (2006).
2. Pei, S., Kandula, S., Yang, W. & Shaman, J. Forecasting the spatial transmission of influenza in the united states. *Proc. Natl. Acad. Ences USA* **115**, 2752–2757 (2017).
3. Merler, S. & Ajelli, M. The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proc. Royal Soc. B Biol. Sci.* **277**, 557–565 (2010).
4. Colijn, C. *et al.* What is the mechanism for persistent coexistence of drug-susceptible and drug-resistant strains of streptococcus pneumoniae. *J. Royal Soc. Interface* **7**, 905–919 (2010).
5. Dietz, K. & Heesterbeek, J. A. P. Daniel bernoullis epidemiological model revisited. *Math. Biosci.* **180**, 1–21 (2002).
6. EnKo, P. D. On the course of epidemics of some infectious diseases. *Int. J. Epidemiol.* **18**, 749–755 (1989).
7. Bacar, N. *A Short History of Mathematical Population Dynamics* (Springer, New York, 2011).
8. Kermack, A. G. & McKendrick, W. O. A contribution to the mathematical theory of epidemics. *Proc. Royal Soc. Math. Phys. Eng. Ences* **115**, 700–721 (1927).
9. Newman, M. E. J. *Networks: An Introduction* (Oxford University Press, New York, 2010).
10. Keeling, M. J. & Eames, K. T. D. Networks and epidemic models. *J. Royal Soc. Interface* **2**, 295–307 (2005).
11. Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200–3203 (2001).
12. Peng, X., Xu, X. J., Small, M., Fu, X. & Jin, Z. Prevention of infectious diseases by public vaccination and individual protection. *J. Math. Biol.* **73**, 1561–1594 (2016).
13. Miller, J. C. A note on a paper by erik volz: Sir dynamics in random networks. *J. Math. Biol.* **62**, 349–358 (2011).
14. Pastor-Satorras, R. & Vespignani, A. Epidemic dynamics and endemic states in complex networks. *Phys. Rev. E* **63**, 066117 (2001).
15. Lindquist, J., Ma, J., Driessche, P. V. D. & Willeboordse, F. H. Effective degree network disease models. *J. Math. Biol.* **62**, 143–164 (2011).
16. Stone, L., Olinky, R. & Huppert, A. Seasonal dynamics of recurrent epidemics. *Nature* **446**, 533–536 (2007).
17. Rohani, N. P. Noise, nonlinearity and seasonality: The epidemics of whooping cough revisited. *J. Royal Soc. Interface* **5**, 403 (2008).
18. Ponciano, J. M., Capistrán, M. A. & Pascual, M. First principles modeling of nonlinear incidence rates in seasonal epidemics. *PLoS Comput. Biol.* **7**, e1001079 (2011).
19. Keeling, M. J., Rohani, P. & Grenfell, B. T. Seasonally forced disease dynamics explored as switching between attractors. *Phys. D Nonlinear Phenomena* **148**, 317–335 (2001).
20. Alonso, D. & Mckane, A. J. Extinction dynamics in mainland-island metapopulations: An n-patch stochastic model. *Bull. Math. Biol.* **64**, 913–958 (2002).
21. Chaffee, J. & Kuske, R. The effect of loss of immunity on noise-induced sustained oscillations in epidemics. *Bull. Math. Biol.* **73**, 2552–2574 (2011).
22. Earn, D. *et al.* A simple model for complex dynamical transitions in epidemics. *Science* **287**, 667–670 (2000).

23. Van Kampen, N. G. *Stochastic Processes in Physics and Chemistry (Third Edition)* (The Netherlands:Elsevier, 1992).
24. Alonso, D., Mckane, A. J. & Pascual, M. Stochastic amplification in epidemics. *J. Royal Soc. Interface* **4**, 575–582 (2007).
25. Wang, R. H., Jin, Z., Liu, Q. X., Koppel, J. V. D. & Alonso, D. A simple stochastic model with environmental transmission explains multi-year periodicity in outbreaks of avian flu. *PLoS ONE* **7**, e28873 (2012).
26. Torrence, Christopher & Compo, Gilbert P. A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* **79**, 61–78 (1998).
27. Cazelles, B., Chavez, M., De Magny, G. C., Guegan, J. & Hales, S. Time-dependent spectral analysis of epidemiological time-series with wavelets. *J. Royal Soc. Interface* **4**, 625–636 (2007).
28. Gillespie, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**, 403–434 (1976).
29. Jin, Z. *et al.* Modelling and analysis of influenza a (h1n1) on networks. *BMC Public Health* **11**, 1–9 (2011).

Acknowledgements

This work is supported by the National Natural Science Foundation of China (61873154,11871316,11971279), Shanxi Key Laboratory (201705D111006), National Key Research and Development Program of China (2016YFD0501500) and Shanxi Scientific and Technology Innovation Team (201705D131028-5).

Author contributions

H.Y. constructed the model, performed the data analysis and wrote the manuscript; Z.J. helped perform the analysis with constructive discussions.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-00260-3.

Correspondence and requests for materials should be addressed to Z.J.