Scaling in the Global Spreading Patterns of Pandemic Influenza A and the Role of Control: Empirical Statistics and Modeling

Xiao-Pu Han[1,2], Bing-Hong Wang[1,3], Chang-Song Zhou[2,4], Tao Zhou[1,5], and Jun-Fang Zhu[1]
[1] *Department of Modern Physics, University of Science and Technology of China, Hefei 230026 China*
[2] *Department of Physics, Hong Kong Baptist University, Hong Kong, China*
[3] *the Research Center for Complex System Science, University of Shanghai for Science and Technology and Shanghai Academy of System Science, Shanghai, 200093 China*
[4] *Centre for Nonlinear Studies, and The Beijing-Hong Kong-Singapore Joint Centre for Nonlinear and Complex Systems (Hong Kong), Hong Kong, China*
[5] *Department of Physics, University of Fribourg, CH-1700, Fribourg,Switzerland*
(Dated: December 8, 2009)

The pandemic of influenza A (H1N1) is a serious on-going global public crisis. Understanding its spreading dynamics is of fundamental importance for both public health and scientific researches. In this paper, we investigate the spreading patterns of influenza A and find the Zipf's law of the distributions of confirmed cases in different levels. Similar scaling properties are also observed for severe acute respiratory syndrome (SARS) and bird cases of avian influenza (H5N1). To explore the underlying mechanism, a model considering the control effects on both the local growth and transregional transmission is proposed, which shows that the strong control effects are responsible for the scaling properties. Although strict control measures for interregional travelers are helpful to delay the outbreak in the regions without local cases, our analysis suggests that the focus should be turned to local prevention after the outbreak of local cases. This work provides not only a deeper understanding of the generic mechanisms underlying the spread of infectious diseases, but also an indispensable tool to decision makers to adopt suitable control strategies.

PACS numbers: 89.75.Da, 02.50.-r, 89.75.Hc

I. INTRODUCTION

A new global influenza pandemic has broken out. In the first three months, the epidemic spread to over 130 countries, and more than 10^5 people were infected by the novel virus influenza A (H1N1). H1N1 represents a very serious threat due to cross-species transmissibility and the risk of mutation to new virus with increased transmissibility. How to prevent the spreading becomes extremely urgent problems. Several early studies paid attention to this public issue from different perspectives [1, 2, 3, 4], and made known important information such as the biological activity of H1N1 virus and the patterns of early spreading. While every effort is being taken to develop antiviral and vaccination drugs, efficient reduction and prevention of the spreading could already be achieved by interventions of population contact. However, such interventions, like strict physical checking at the borders and enforced quarantine, are costly and highly controversial. It is therefore difficult to decide the control strategies: when should the schools be suspended and whether the border control should be reinforced or given up?

The detailed mechanism of transmission can differ significantly for different virus, the spreading patterns, however, may display common regularities due to generic contacting processes and control schemes. Many health organizations have collected large amount of information about the spreading of H1N1. In-depth analysis of these data, together with what we have known for SARS [5, 6], H5N1 [7, 8], foot-and-mouth epidemic [9, 10] and some other pandemic influenza [11, 12], may lead us to a more comprehensive understanding of the common spreading patterns that do not rely on the detailed biological features of virus. In this paper, we study the spreading patterns of influenza pandemic by both empirical analysis and modeling. Our main contributions are threefold: (i) The Zipf's law of the distribution of confirmed cases in different regions are observed in the spreading of H1N1, SARS and H5N1; (ii) A simple model is proposed, which does not rely on the biological details but can reproduce the observed scaling properties; (iii) The significant effects of control strategies are highlighted: the strong control for interregional travel is responsible for the Zipf's law and can sharply delay the outbreak in the regions without local cases, while the focus should be turned to local prevention after the outbreak of local cases. Our analysis provides a deeper understanding of the relationship between control and spreading, which is very meaningful for decision makers.

II. EMPIRICAL RESULTS

We first analyze the *cumulative* number n_i of laboratory confirmed cases of H1N1 of each country i till a given date (see the data description in *Methods and Materials*). Because n_i is growing, the distributions for different dates are normalized by the global total cases $N_T = \sum n_i$ till the corresponding dates for comparison. Fig. 1(a) and 1(b) report the Zipf's plots (see *Methods and Materials*) for the distributions of normalized n_i in different dates. The maximal rank corresponds to the number of regions with confirmed cases, which grows during the spreading.

The normalized distributions P surprisingly display scaling properties. Before the middle of May, P shows clearly a power-law type $P \sim r^{-\alpha}$ with an exponent $\alpha \approx 3.0$ (except the first data point, see Fig. 1(a)). Although the total cases N_T grows rapidly in this early stage, P for different dates seems to follow the same line in the log-log plot. After the middle of May, the middle part of the distribution grows more quickly, and meanwhile the virus spreads quickly to many more countries. The exponent α of the left part of P changes from about 3.0 to about 1.5, and an exponential tail emerges (Fig. 1(b)). After June 10, P can be well fitted by a power-law function with an exponential tail, for example, $P = 0.75r^{-1.53}e^{-r/40}$ for the data of July 6 (solid line, Fig. 1 (b)). The scaling properties are not special for the influenza A, but quite common in various diseases. To demonstrate this, we further analyze SARS in 2003 and the bird cases of avian influenza (H5N1) in 2008. As shown in Figs. 1(c) and 1(b), both of the two distributions show a power-law-like form, but the spreading range is much more limited (to only about 30 countries). For SARS $\alpha \approx 2.7$, and for H5N1, $\alpha \approx 2.0$.

Besides the Zipf's law, the *Heaps' law* [13] is another well-known scaling law observed in many complex systems, which describes a sublinear growth of the number of distinct sets as the increasing of the total number of elements belonging to those sets. Recent empirical analysis [14, 15] suggested that the Heaps' law and the Zipf's law usually coexist. As shown in Fig. 7, the number of infected countries M (before May 18, 2009) grows with the global total confirmed cases N_T in a power-law form as $M \sim N_T^{\lambda}$ with the Heaps' exponent $\lambda \approx 0.35$. Actually, Lü *et al.* [16] proved that if an evolving system has a stable Zipf's exponent, its growth must obey the Heaps' law with exponent $\lambda \approx 1/\alpha$ (an approximate estimation when $\alpha > 1$: the larger the α, the more accurate the estimation). Since before May 18, 2009, the system obeys the Zipf's law with exponent $\alpha \approx 3.0$, the empirical findings (Fig. 1(a) and Fig. 7 in *Supporting Information*) agree well with the theoretical analysis [16].

We also find that broad distribution of n_i is related to heterogeneity in different countries. Figs. 2(a) and 2(b) report the dependence between the number of confirmed cases n_i and the population and gross domestic product (GDP). A clearly hierarchical spreading pattern, similar to what were predicted by some theoretical models [17, 18], can be observed: the big and rich countries will be infected first, and then the disease spreads out to the global world. This can be understood that bigger and richer countries usually have more active population in international travel, and thus are of higher risk to be new spreading origins in the early stage of epidemic. The evolution of correlations between the confirmed cases n_i and population and GDP is reported in Fig. 2(c) by the *Kendall's Tau* (see *Methods and Materials*). Significantly positive correlations with a tendency of increase with time can be observed, confirming that the population and economic level are important factors for disease

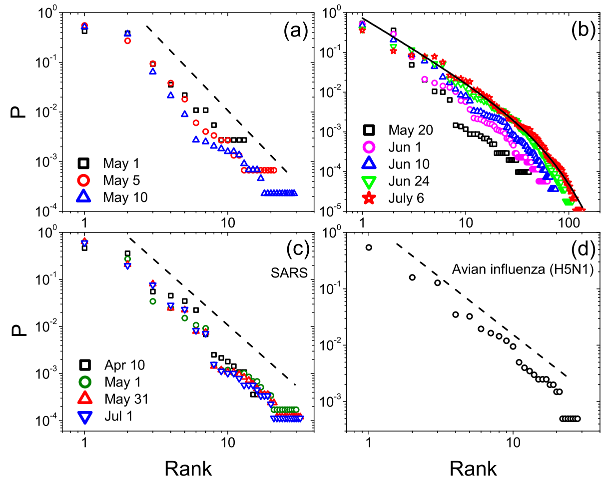

FIG. 1: (a) Zipf's distribution of the normalized number of H1N1 cases in different countries in a log-log plot with date before May 15, 2009. (b) Same as (a) but with date after May 15, 2009. (c) Zipf's distribution of the normalized number of probable SARS cases for different countries in 2003. (d) Zipf's distribution of the normalized number of H5N1 cases for different countries in the whole year of 2008. The dash and solid curves are fitting functions described in the text.

spreading. The global total confirmed cases N_T displays two phases of growth (Fig. 2(d)): in the early stage N_T increases exponentially with a high rate and then turns into a stable exponential growth with a much smaller rate, with the transition occurring around the middle of May. Such a transition may reflect the changes in the contacting rate among people due to imposed or self-adaptive control.

We next investigate the statistical regularities within a country. In Fig. 3, we compare the normalized distribution of confirmed cases in different states of USA and in different provinces of China: P of USA shows a much more homogeneous form with a large deviation from strict power-law distribution while P of China is close to a power-law with exponent $\alpha = 1.79 \pm 0.04$ (the Zipf's distribution of SARS cases of different provinces of China is also a power-law type with exponent $\alpha \approx 3$ [19]). These results show clearly that the spreading patterns in different countries can differ significantly. We believe that different strength of control and intervention measures adopted by different countries play an important role, as will be discussed in more detail in our model. To summarize, the empirical results show that the scaling properties in epidemic spreading process may widely exist at different regional levels and crossing various infectious diseases. In the following we try to obtain some insight into the generic mechanisms underlying these common properties.

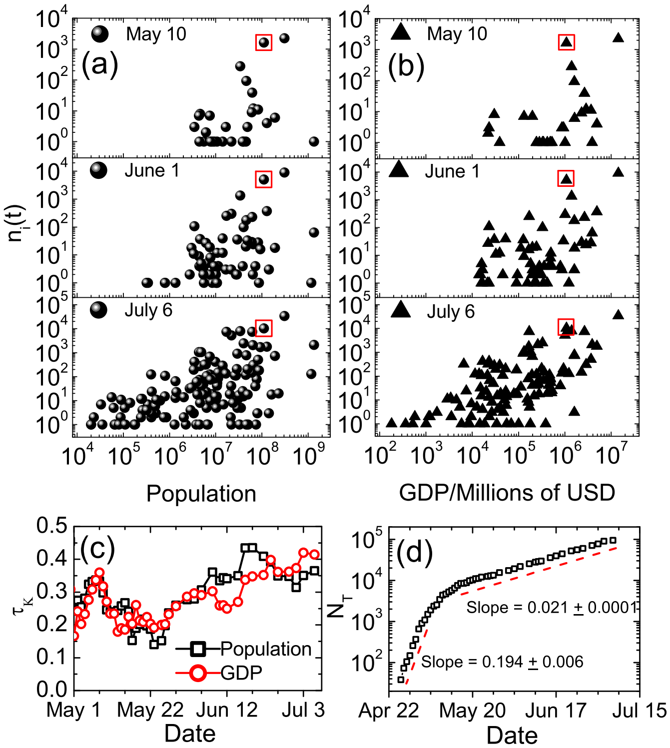

FIG. 2: (a) and (b) show the evolution of the dependence between the number of laboratory-confirmed cases n_i for different countries and the population and GDP of these countries, respectively. The data for Mexico where the disease initiated are highlighted with an open square. (c) The Kendall's Tau of the correlations between the number of confirmed cases and the population/GDP. (d) Growth of global total number N_T of laboratory-confirmed cases of Influenza A in the semi-log plot.

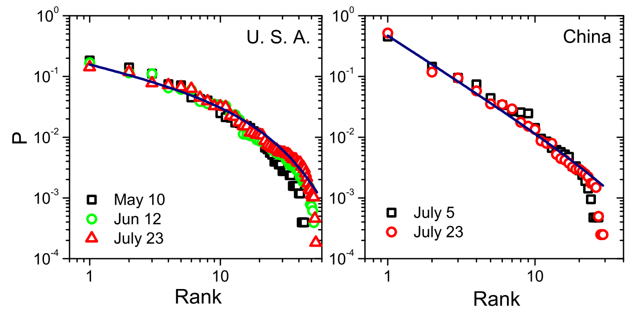

FIG. 3: Normalized distributions P of confirmed cases in different states of USA and different provinces of China. For USA, the solid curve is $P = 0.17r^{-0.51}e^{-r/19}$, fitting the data of July 23. For China, the solid line is of slope -1.79, fitting the data of July 23.

III. THE MODEL

The empirical results provoke some outstanding questions: how to understand the scaling properties in region distributions, which factors lead to the different spreading patterns for different regions, and what are the effects of control measures on the regional level spreading? We believe that the scaling properties have the origin at the generic contact process underlying the transmission of diseases, and the variation could result from the heterogeneity of the contact process in different diseases and regions. One most important heterogeneity may be the control strength. To build a generic model incorporating the effects of control, let us consider the actions taken by people when facing a serious epidemic spreading. In general, individual people try to take many approaches to reduce the probability of infection, such as using respirator, reducing the face-to-face social interactions, and disinfecting frequently. And also, many organizations usually take measures to prevent the spreading of epidemic, such as physical examinations in public transportation and schools, vaccinations and isolations for highly risky groups, and so on. If epidemic breaks out in a country, other countries may reinforce the health examinations at the borders for the travelers from that country. For example, measurement of body temperature is used in many airports since the outbreak of influenza A in Mexico, and some countries implement strict isolation for the infected persons and their close contacts. Similar but more strict measures have also been adopted in the control of SARS. These actions of individuals and social organizations can effectively change the structure of social contacts, reduce infection probability and affect the spreading patterns of epidemic [20, 21]. Such effect of imposed or self-adaptive controlling actions on reduced infection rate is the starting point of our modeling for the spreading process.

Different from many individual-based models, our model is in the regional level, so the detailed social contact structure [22, 23, 24] as well as the control methods and strategies in individual level [25, 26, 27] are not considered directly. In our model, a region (such as a country) is denoted by a node in a network with K nodes in total. The network is supposed to be fully-connected since in general there are direct contacts between almost all countries in the world. However, the strength of connections between the countries could be different due to the heterogeneity in various factors, such as population and economics. As we will show later, while such heterogeneity has some impact on the epidemic spreading, the most important ingredients are the strengths of control within and between the regions. Therefore, instead of employing the detailed information of real traffics, we generically denote the international traffic of a node as its strength s_i, and the weight of link between two node i and j is assumed to be symmetric and proportional to the products of the strengths s_i and s_j:

$$q_{ij} = s_i s_j / \sum_{k=1}^{K} s_k. \quad (1)$$

The spreading at time t from node j to i is proportional to the number of infected cases n_j of node j, together with a time-varying effective weight $w_{ij}(t)$ of the link, namely $w_{ij}(t)n_j(t)$. Here $w_{ij}(t)$ is related not only to the link strength q_{ij}, but also to the control strategy. Control measures are in general reinforced on the travelers from countries with large number of infected cases, and thus in our model the link weight is

$$w_{ij}(t) = q_{ij} n_j(t)^{-\beta_1}, \quad (2)$$

where β_1 is a free parameter. Effectively, we can take $w_{ij}(t) = 0$ if $n_j(t) = 0$. Note that while q_{ij} is symmetric, w_{ij} is in general asymmetric. This expression describes generically the effects of various control measures at the borders, without relying on the details at the individual level.

In this model, the update of the number of cases n_i of an arbitrary node i consists of two parts: a local infection growth and the global traveling infections:

$$\Delta n_i = \rho \left[a_i(t) n_i(t) + \frac{b}{\langle s \rangle} \sum_{j=1, j\neq i}^{K} w_{ij}(t) n_j(t) \right], \quad (3)$$

where ρ is a positive constant related to the basic transmissibility of the diseases, $\langle s \rangle$, the average value of s_i, is introduced for normalization, and the coefficient b denotes the relative contribution due to the transmission from other regions. Note that Δn is generally a real number while the real-world increment of infected cases must be integral. Therefore, we round Δn to the neighboring integer, namely to set $n_i(t+1) = n_i(t) + [\Delta n] + 1$ with probability p and $n_i(t+1) = n_i(t) + [\Delta n]$ with probability $1-p$, where $p = \Delta n - [\Delta n]$ ($[x]$ denotes the largest integer no larger than x).

The relative contribution by local infections, $a_i(t)$, is not constant, but reflects the strength of control within a region. In the same vein as the border control in Eq. 2, we describe the generic effects of local control by decaying $a_i(t)$ as a function of n_i with a free parameter β_2, namely

$$a_i(t) = \begin{cases} n_i(t)^{-\beta_2}, & \text{if } n_i(t)^{-\beta_2} > g \\ g, & \text{if } n_i(t)^{-\beta_2} \leq g. \end{cases} \quad (4)$$

Effectively, $a_i(t) = 0$ if $n_i = 0$. Here the decaying of a_i is limited by a constant g $(0 < g < 1)$, which accounts for the necessary social contacts in the daily life even under the outbreak of the epidemic. In reality, g is also related to the transmissibility and death rate of the disease.

IV. RESULTS OF THE MODEL

To focus on the effects of the control parameters β_1 and β_2, we first consider the simplest case of the model

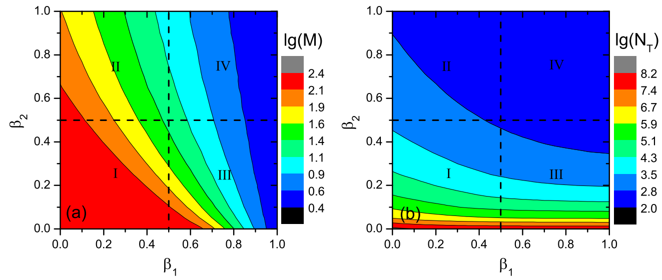

FIG. 4: Effects of the parameters β_1 and β_2 on spreading speed, measured by the logarithm of the infected range M (a) and the logarithm of the total cases N_T (b) at $t = 100$. Simulations run with $b = 0.06$, and $g = 0.2$, and all the data are averaged over 10^4 independent realizations.

in which s_i is uniform. In this case, $q_{ij} = 1/K$ and Eq. 3 is reduced to the minimal model

$$\Delta n_i = \rho \left[a_i(t) n_i(t) + \frac{b}{K} \sum_{j=1, j\neq i}^{K} n_j(t)^{1-\beta_1} \right]. \quad (5)$$

The impact of the heterogenity in s_i will be discussed later.

In the following, to represent a worldwide network of countries, the total number of nodes in our model is $K = 220$, the number of country. The model can also be used to represent the spreading within a county when regarding a node as a region within a country and ignoring the transmission from other countries. From Eq. 3, the parameter ρ does not affect the pattern of the normalized distribution P. ρ is thus fixed at 0.2 in all our simulations, which is close to the fast growing rate of the influenza A in the early stage of outbreak (see Fig. 2(d)). Initially, the epidemic starts at a random node k ($n_k(0) = 1$, and $n_l(0) = 0$ for $l \neq k$). To quantify the epidemic spreading with time, we compute the spreading range M (the number of nodes with $n_i > 0$) and the total cases N_T. The results of M and N_T in the minimal model at time $t = 100$ are shown in Fig. 4 in the parameter space (β_1, β_2). It is seen that both large β_1 and large β_2 can reduce the range of spreading M, but the control on the interregional borders by β_1 is more effective than β_2 (Fig. 4(a)). On the contrary, large β_2 is much more effective than β_1 to reduce the total number of cases N_T (Fig. 4(b)). We find that these patterns with respect to β_1 and β_2 in Fig. 4 are generic in the model for different parameters ρ, b and g and for different time during the spreading. These results imply that once a country has local epidemic outbreaks, its growth will be mainly driven by the local spreading but not the input of foreign cases.

The parameter space of β_1 and β_2 can be divided into four regimes, corresponding to various combinations of weak or strong local or interregional controls, as indicated in Fig. 4. Typical normalized distributions P obtained

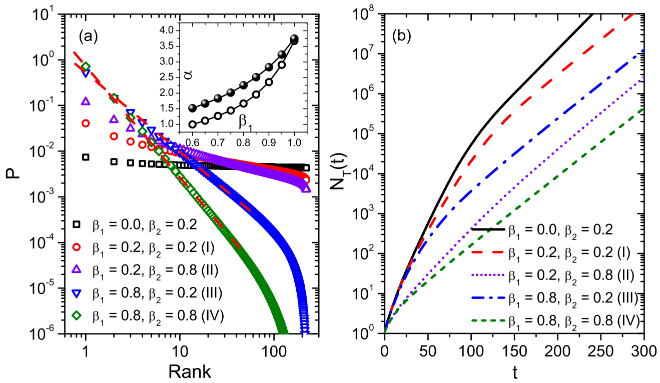

FIG. 5: (a) Typical normalized distribution P (at $t = 300$) in the regime of I, II, III, and IV in the parameter space (β_1, β_2) shown in Fig. 5. The two red dashed lines indicate the power-law functions with $\alpha = 1.67$ and $\alpha = 2.25$, respectively. The inserts show the dependence of α on β_1 for fixed $\beta_2 = 0.2$ (open circle) and $\beta_2 = 0.8$ (filled circle). (b) The corresponding growth of N_T with respect to time. The other parameters are the same as in Fig. 4.

in the four regimes are compared in Fig. 5. When β_1 is small (regimes (I) and (II)), the epidemics can spread to almost all nodes in short time, and P is rather homogeneous. When β_1 is large (regimes (III) and (IV)), the spreading across different region is suppressed, and P is rather inhomogeneous, manifested as a power-law-like form. Keeping β_2 fixed, the exponent α clearly increases with β_1 and a larger β_2 can slightly increase α further (inset, Fig. 5(a)). We have included a detailed discussion of the time evolution of the distributions P and their association to the Heaps' law in the *Supporting Information* (Figs. 8-9).

While β_1 has a sensitive impact on the interregional spreading and controls the heterogeneity of the distribution P, β_2 mainly affects the growth of total cases N_T, especially in the early stage (Fig. 5(b)). With stronger control at larger β_2, the fast growth of N_T in the early stage will be effectively suppressed and will be transformed to a slow exponential growth within shorter time. As seen in Eq. (4), β_2 only affects the growth in the very early stage after the epidemics appears in a region. The significant effect of β_2 on the growth of total cases N_T emphasizes the importance of early epidemic control, in agreement with the conclusion of previous studies on other diseases [9, 10].

Comparing the results from the four regimes, we can see that the spreading pattern in regime III (large β_1 and small β_2) is closer to the empirical observations of influenza A. In this regime, the range of α covers most of empirical results. For example, with $\beta_1 = 0.8$ and $\beta_2 = 0.2$, the distribution P can be well fitted by a power-law function with exponent 1.67 (Fig. 5(a)), and this value is close to the empirical exponent of influenza A on July 6 (Fig. 1(b)). Large β_1 and small β_2 is consistent with the real-world situation. While more efficient to implement control measures on the borders, e.g., to identify infected and suspected candidates and their close contacts for quarantine, it is much more difficult to get the same efficiency for the same control schemes in local communities. The relative lower death rate of influenza A is also likely to weaken the self-adaptive control and voluntary isolations of the individuals, leading to insignificant change of the contact patterns (e.g., much weaker than SARS). All these will render a lower efficiency in the local control, corresponding to a small β_2 and a larger g.

The parameter g expresses the background local growth speed under the effect of control measures which cannot be further reduced due to unavoidable social contacts. The other parameter b denotes the relative strength of interregional transmission. Both of them can affect the exponent of distribution P, and g also has a very sensitive impact on the growth of the total number N_T (see *Supporting Information*).

All the above discussions are based on the minimal model where the diversity of the nodes and the edges is ignored by assuming a uniform s_i. Now we study the impact of heterogeneous s_i and the effect of target control on the spreading of disease. While previous investigations have focused overwhelmingly on the impact of heterogeneity in the degree of complex networks [17, 24], here we study the effects of the heterogeneity in the intensity of nodes and links in globally coupled networks. We first take the real population of different countries as s_i in our model and investigate how does the initiation of the disease in countries with different ranks of populations influence the global spreading. When the disease starts in a country with a large s_i (Fig. 6(a), population rank $R_{ini} = 11$ as Mexico), the disease spreads out quickly and the spreading process displays a clear tendency from the node with large s_i to those with small s_i as seen by the evolution of the scatter plot of $n_i(t)$ vs.s_i and the Kendall's tau (Fig. 6(c)), which reproduces the main features in the empirical data in Fig. 2. On the contrary, when the initiation happens in a country with small population (Fig. 6(b), population rank $R_{ini} = 100$ as Libya), the disease is contained in the country where it is initiated for a period of time, and then the countries with the largest populations get infected soon and become new centers of spreading. τ_K is around zero in the very beginning when the diseases is contained and becomes negative when spreading to a few nodes with the largest s_i and quickly shift to positive values when the new centers take the leading role in the spreading Fig. 6(c)). The total cases N_T grows much faster in the first case (Fig. 6(d)). We have applied target control in our model (see *Supporting Information*), and we find that strong control just on one or two nodes with the largest s_i can sharply reduce the spreading by several orders of magnitude (Fig. 6(e)). This effect is similar to target immunization of the hubs in degree heterogeneous complex networks [27]. A more systemactic analysis of

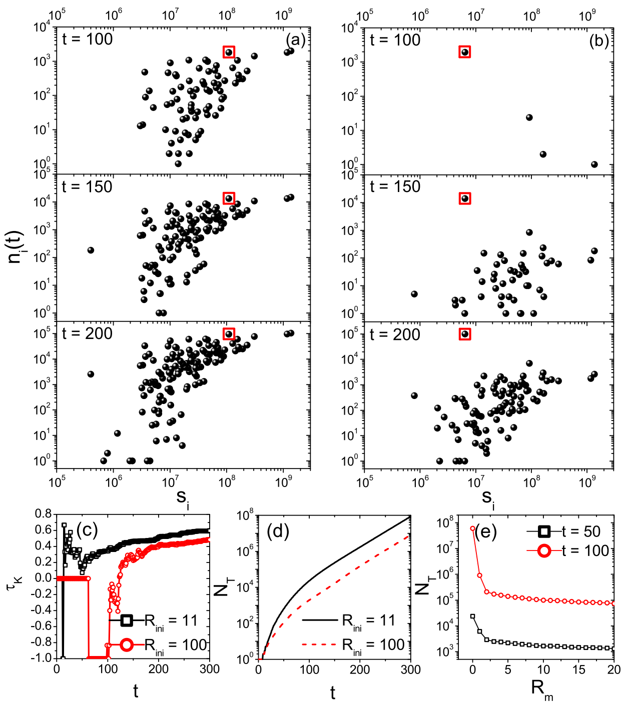

FIG. 6: Effects of heterogeneous s_i in the model where s_i is taken as the population of different countries. (a) and (b) show the evolution of of $n_i(t)$ vs s_i when the disease is initiated at different countries (highlighted by an open square). Simulations run with $\beta_1 = 0.8$, $\beta_2 = 0.2$, $b = 0.06$, and $g = 0.2$. (c) Evolution of τ_K between $n_i(t)$ and s_i. (d) The corresponding growth of $N_T(t)$. (a-d) are obtained from one realization of simulation. Statistical results from many realizations are shown in Fig. 13 in *Supporting Information*. (e) N_T at $t = 50$ and $t = 100$ when R_m nodes with the largest s_i are the targets of control with $\beta_1 = 0.8$ (see *Supporting Information*, the other parameters are the same as in (a-d)). Data of (e) are averaged from 10^4 independent runs with disease initiation at random nodes.

the effects of node heterogeneity and target control by considering a power-law distribution of s_i is included in *Supporting Information*. We find that even though the heterogeneity can accelerate the spreading, the strength of control plays the leading role to determine the patterns of spreading.

To summarize, power-law distribution of P with large exponent α appears in the situations with large β_1, small b and large g. This regime corresponds to the real situations that the epidemic control for the travelers is strong, the interregional contact is much weaker compared to that in local communities, and the change of local social contacts by the disease is not very significant. The epidemic control for the interregional travelers (large β_1) is most important condition for the emergence of the power-law type of P, since the power-law distribution cannot be generated when β_1 is close to zero no matter what other parameters are.

V. DISCUSSION

The statistics of region distributions of several pandemic diseases, including influenza A (H1N1), SARS and bird cases of avian influenza (H5N1) display obvious scaling properties in the spreading process at different levels. We study the origin of such scaling properties with a model of epidemic spreading at the regional level that incooperates the generic effects of intervention and control measures without the need of the structure details of social contacts and the particularity of the transmission of the diseases. Such a model is then able to capture the general principles underlying epidemic spreading and to reveal the generic impact of control measures. We elucidate that strict epidemic control on interregional travellers plays an important role in the emergence of the scaling properties.

The results of the model can cover the empirical statistics of H1N1 on both the region distribution and the growth of total cases, and are also consistent with the region distribution of SARS and H5N1. In particular, the exponent α of the empirical distribution P of H1N1 is about 3.0 in the early stage and changes to 1.53 on July 6, 2009, and α is about 2.7 for SARS and 2.0 for H5N1. In the stable spreading period, the α of H1N1 is smaller than SARS and H5N1. According to the understanding from our model, larger α indicates that the control measures are more strict and effective. This speculation is in agreement with the situation in SARS and H5N1 spreading. Because of high death rate and strong infection capability, SARS gave rise to strong social panic and attracted attentions from citizens to governments in the countries with outbreaks, such as China, and strict control measures were enforced in each public transportation systems and in daily life of people. As for H5N1, many efficient control measures were also taken to prevent the spreading, such as immunity for poultry and culling of livestock, etc. Large α in the early stage of the spreading of H1N1 could be related to stronger control effect due to overrating of the mortality of H1N1. Empirical results also showed that the distribution P of H1N1 in USA is more homogenous than in China. While there are probably several factors contributing to this difference, but the most obvious difference is in the control measures. China takes strict control policies, such as enforced quarantine and isolation for identified infectors and the close contacts, which are not so strict compared to those during the SARS spreading, but are stronger than USA.

Our main findings, i.e., interregional control mainly affects the spreading range and the form of the region distributions while local control sensitively impacts the growth of total cases, provide us a picture of epidemic control. For regions that have no or only a few local infected persons, strict control measures for interregional travellers can delay the local outbreaks significantly, but if large number of local cases have been broken out, these control methods for travellers are not so important. Instead, control methods and treatment for local communities will be much helpful. Recently, the focal point of the control policies for influenza A of many countries have turned to the treatment for infected persons. According to the conclusions of the present model, this strategy shift is reasonable. This model also indicates that the diversity of different regions will accelerate the spreading. Efficient prevention of the spreading could be achieved by enhanced control measures, especially for the giant regions.

In summary, a simple physical model basing on the abstraction of the generic contact processing and the effects of control can provide meaningful understanding of the scaling properties commonly observed in various pandemic diseases. It deepens our understanding of the relationship between the strength of control and the spreading process, and provide a meaningful guideline for the decision maker to adopt suitable control strategies.

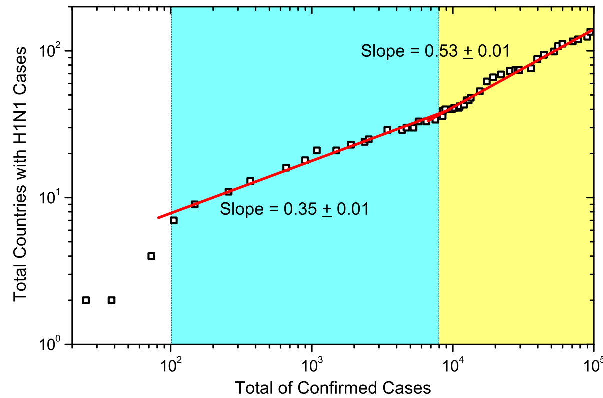

FIG. 7: Dependence between the number of infected countries, M, and the global total N_T of confirmed cases in log-log plot. The red lines respectively are the fitting line with slope 0.35 and 0.53.

VI. METHODS AND MATERIALS

A. Data Description

The cumulative number of laboratory confirmed cases of H1N1 of each country is available from the website of Epidemic and Pandemic Alert of World Health Organization (WHO) (http://www.who.int/), which started from April 26 to July 6, and updated each one or two days. After July 6, WHO stopped the update for each country since the global pandemic has broken out. The data for SARS and H5N1 are available from the websites of WHO and the World Organization for Animal Health (OIE) (http://www.oie.int/), respectively. The data for H1N1 cases of different states of USA is available on the website of Centers for Disease Control and Prevention (CDC) (http://www.cdc.gov/h1n1flu/), and the data of different provinces of China is available from Sina.com (http://news.sina.com.cn/z/zhuliugan/). The data for populations and GDPs of different countries are obtained from English Wikipedia http: // en.wikipedia.org/ wiki/ List_of_countries_by_population and http: // en.wikipedia.org/ wiki/ List_of_countries_by_GDP. There are three different lists of GDPs and what we used here is the one from the World Bank, which includes 182 countries. Among the 135 countries having reported the confirmed H1N1 cases until July 6, 22 of which do not have GDP data. They are all small countries and the number of confirmed cases in these countries is also quite few (the total of the 22 countries are 163 until July 6). We thus ignore them in evaluating the correlation in Fig. 2(c).

B. Zipf's Law and Power Law

Zipf's plot is widely used in the statistical analysis of the small-size sample [28], which can be obtained by first rearranging the data by decreasing order and then plotting the value of each data point versus its rank. The famous *Zipf's law* describes a scaling relation, $z(r) \sim r^{-\alpha}$, between the value of data point $z(r)$ and its rank r. As a signature of complex systems, the Zipf's law is widely observed [29, 30]. Indeed, it corresponds to a power-law probability density function $p(z) \sim z^{-\beta}$ with $\beta = 1 + \frac{1}{\alpha}$.

C. Kendall's Tau

In the empirical analysis, the numbers of confirmed cases, populations and GDPs for different countries are very heterogeneous, covering several orders of magnitude (e.g., the population of China is about 2×10^4 times larger than that of Dominica). Thus the classical measurement like the Pearson coefficient is not suitable in analyzing the correlations. We therefore use the rank-based correlation coefficient named *Kendall's Tau*. For two series $\overrightarrow{x} = \{x_1, x_2, \cdots, x_m\}$ and $\overrightarrow{y} = \{y_1, y_2, \cdots, y_m\}$, the Kendall's Tau is defined as [31]

$$\tau_K = \frac{2}{m(m-1)} \sum_{i<j} \mathtt{sgn}[(x_i - x_j)(y_i - y_j)], \qquad (6)$$

where $\mathtt{sgn}(x)$ is the signum function, which equals +1 if $x > 0$, -1 if $x < 0$, and 0 if $x = 0$. τ_K ranges from +1 (exactly the same ordering of $\overrightarrow{x}$ and $\overrightarrow{y}$) to -1 (reverse ordering of $\overrightarrow{x}$ and $\overrightarrow{y}$), and two uncorrelated series have $\tau_K \approx 0$.

D. On Power-Law Fitting

Most of the distributions P generated by simulations of our model with large β_1 (≥ 0.6) trend to a power-law-like type after several steps of evolution. In the fittings of simulation results, we firstly judge if the curve of P in this range is power-law-like. If yes, we fit the curve by linear function in log-log plots in using least square fit method to get the fitting parameters. The range of the power-law fittings of is from 2 to 50. If there is obvious deviation from power-law in this range, we do not use power-law to fit the curve. The only exception is the distribution P when $b = 0.02$ in Fig. 12(a), where the range is from 1 to 30, because the cut-off appears at rank = 30 due to slow spreading of the disease. All the power-law fitting results in the model does not show the error-bar (e.g., the dependence of α on various parameters of the model), because the fitting error on the power-law exponent is far less than the value of α for most cases after 10^4 averages (e.g., $\alpha = 1.666 \pm 0.003$ when $\beta_1 = 0.8$, $\beta_2 = 0.2$, $\rho = 0.2$, $b = 0.06$ and $g = 0.2$ in the minimal model).

VII. SUPPORTING INFORMATION

A. Heaps' Law for H1N1 Spreading

In the real-world spreading of influenza A, a scaling relation between the number of infected countries, M, and the global total N_T of confirmed cases was observed. As shown in Fig. 7, it displays a power-law dependence with exponent $\lambda \approx 0.35$ in the range from $N_T = 100$ to $N_T = 8000$ ($N_T = 8000$ corresponds to global total reported on May 18), and then turns to fast spreading with exponent $\lambda \approx 0.53$ in the range $N_T > 8000$.

This scaling relation, $M \sim N_T^{\lambda}$, is called the Heaps' law [13], which describes a sublinear growth of the number of distinct sets as the increasing of the total number of elements belonging to those sets. Note that, the Heaps' law only exists in a limited range because in the very early stage, the infected countries are too few to display any statistical regularities. As discussed in the paper, $\lambda \approx 1/\alpha$, thus larger λ after May 18 is consistent with smaller α (Fig. 1). While it has not yet been shown from the data available only till July 6, the growth of M will eventually saturate since the number of countries on the earth is limited.

B. The evolution of Zipf's distribution P vs the Heaps' plots in the model

The empirical results in Fig. 2(b) indicates that the Zipf's plot converges to a stable distribution before the spreading range M reaches saturation. The convergence to a stable distribution is inherent in our model. Let us consider a few nodes in the model with the largest n_i. The growth of such nodes is mainly determined by the

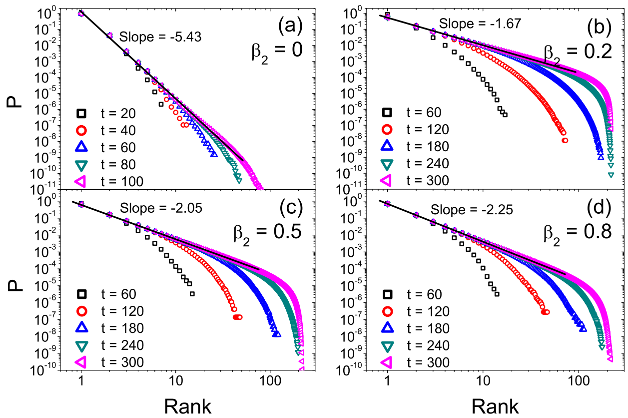

FIG. 8: Evolution of the normalized distributions P at different time steps of the model for various β_2. The other parameters are $\beta_1 = 0.8$, $b = 0.06$, $\rho = 0.2$ and $g = 0.2$. All the data are average over 10^4 independent runs.

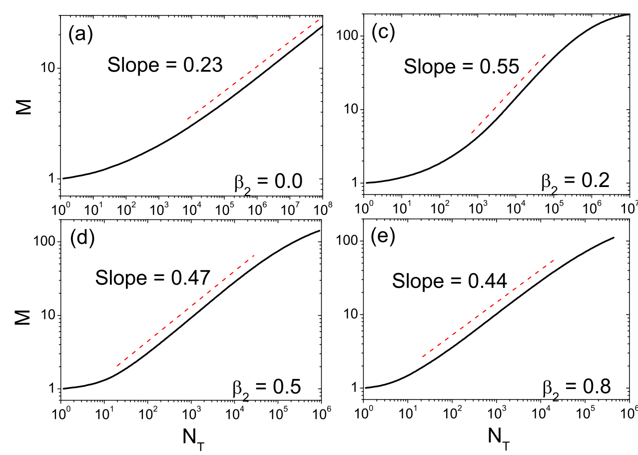

FIG. 9: Dependence between M and N_T generated by the model for different β_2, corresponding the distributions P shown in Figure S2.

local growth rate ρg, $n_i(t+1) \approx (1 + \rho g) n_i(t)$, since the number of infected case due to input from other nodes is much smaller and can be neglected and the local growth has shifted to a stable rate due to local control in Eq. 4. When considering a power -law distribution at time t, $P_t(r) \sim r^{-\alpha}$, the total global cases N_T are also mainly contributed by these a few nodes with the largest n_i, i.e., $N_T(t+1) \approx (1 + \rho g) N_T(t)$. Thus the normalized distribution at $t+1$ for these nodes is $P_{t+1}(r) = n_i(t+1)/N_T(t+1) \sim r^{-\alpha}$ which is invariant vs. time. This analysis is confirmed by the evolution of P at various parameters in Fig. 8. We can see that the distributions P at different time overlap for the nodes with the smallest rank r. The range of the forepart of the curve of P which

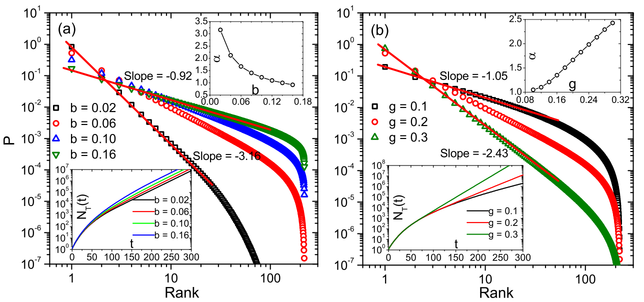

FIG. 10: (a) The normalized distributions P for different b. The two red lines are power-law functions with exponent $\alpha = 0.92$ and $\alpha = 3.16$, respectively. The top insert: dependence of α on b. Bottom insert: $N_T(t)$ vs. time for various b. Simulations run with $\beta_1 = 0.8$, $\beta_2 = 0.2$, and $g = 0.2$. (b) The same as (a), but for different g. The two red lines are power-law functions with exponent $\alpha = 1.05$ and $\alpha = 2.43$, respectively. Simulations run with $\beta_1 = 0.8$, $\beta_2 = 0.2$, and $b = 0.06$. All the data are average over 10^4 independent runs.

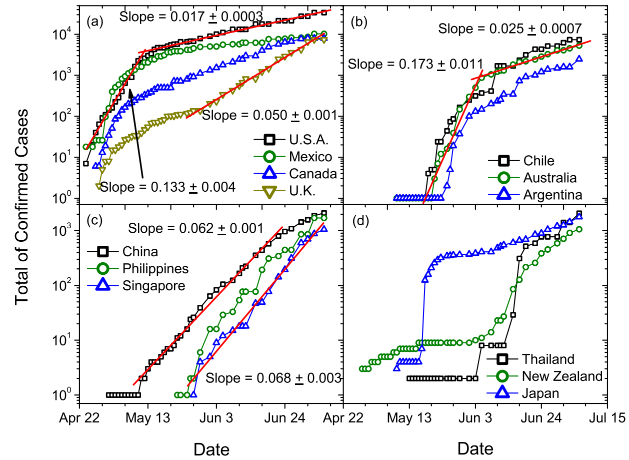

FIG. 11: The growth of total confirmed cases within a country. The four panels correspond to the four types.

can be well fitted by power-law extends along with the time evolution, and the cut-off tail will move to large ranks r till it reaches the system size K.

In our model, the scaling property in the distribution P is mainly contributed by large β_1. An extreme situation is that $\beta_1 > 0$ and $\beta_2 = 0$, namely, the effect of local control is ignored (the parameter g does not have any impacts). In this case, $n_i(t+1) \approx (1+\rho)n_i(t)$ for the nodes with the largest n_i and N_T increases with a rate close to ρ. P converges quickly to a power-law distribution when β_1 is large (Fig. 8(a)), and the exponent α is quite large because the early infected nodes grow very fast. On the other hand when β_2 is large enough, the growth of n_i and N_T will shift quickly to a stable rate ρg and P again converges to a power law distribution. α is significantly smaller than that at $\beta_2 = 0$ because the local control reduces significantly the growth rate of the early infected nodes, and n_i is not as heterogeneous. When β_2 is small, it takes a long period of time for infected nodes to achieve an stable exponential growth and consequently it takes many steps for P to converge. The convergent exponent α becomes larger when β_2 increases, because the newly infected nodes do not grow very fast under stronger local control so that the distribution become slightly more heterogeneous.

In the discussions of the results of our model, the evolution time of the model generally is set as 300 steps, because in the parameter settings in our discussion, most of the P distributions can show long range of power-law part and the exponent of the power-law part trends to stable after 300 steps of evolution.

From Fig. 8 we can also see that the distribution at a given time t has a cut-off at $r = M(t)$ where $P(r) = 1/N_T(t)$. When α is large (e.g., Fig. 8(a)), as an approximation we can assume that the power - law distribution extends to the cut-off point, i.e., $P(r) \approx 1/N_T(t) \sim M^{-\alpha}(t)$, and we get $M \sim N_T^{\lambda}$ where $\lambda = 1/\alpha$, implying that the Heaps' law can be observed in the process. The Heaps' plots corresponding to the Zipf's plots in Fig. 8 are shown in Fig. 9. We can see that the fitting exponent $\lambda \approx 1/\alpha$ as expected from the analysis. The plots also manifest the saturation of M when N_T becomes very large.

C. Effects of parameters b and g

Besides the two parameters β_1 and β_2 for the border and local control, the other two parameters b and g related to interregional and local contact rates can also significantly affect the spreading processes.

The parameter b in our model denotes the relative strength of interregional transmission. Large flow of interregional travels can also make the epidemic spread to most of the regions rapidly. As a result, the distribution P becomes more homogeneous with decreasing α when b is larger; however, b has only a slight impact on the growth pattern of N_T (Fig. 10(a)).

The parameter g expresses the background local growth speed which cannot be further reduced due to unavoidable social contacts even under the effect control measures. Under strong border control (large β_1), the number of infected cases n_i is mainly determined by g, growing exponentially with the rate ρg after an initial transient period, thus g has a very sensitive impact on the growth of the total number N_T (Fig. 10(b)). If g is large, earlier infected regions will have much more infected cases compared to later infected regions, leading to an inhomogeneous distribution P. At smaller g, the earlier and later infected regions do not differ very much in the number of infected cases, corresponding to more homogeneous distribution P with decreasing α (Fig. 10(b)). Different from the case of weak border control (small β_1), homogeneous P here dose not mean the rapid spreading; on the contrary, it denotes the situation that the spreading in each country is in a lower level.

D. Diversity of control in data and model

1. Growth of confirmed cases within a country: empirical data

We have investigated the growth of the number of confirmed cases n_i for all the countries with $n_i > 10^3$ until July 6, and found that the patterns of growth are quite diverse. As shown in Fig. 11, the growth patterns can be roughly classified into four different types. **Type I** (Fig. 11(a))—each curve in this type has a clear transition in the middle of May from a rapid breakout to a stably exponential growth. These are the countries got infected shortly after the breakout of H1N1 in Mexico. **Type II** (Fig. 11(b))—each curve in this type also shows a crossover from quick to slow growth as in Type I, but the initial infection in these countries were much late. **Type III** (Fig. 11(c))—each curve in this type exhibits a stably exponential growth without a pronounced crossover. This happens for countries with early or late initial infection. **Type IV** (Fig. 11(d))—curves in this type can not be classified into the former three types. These irregular curves usually include one or more very sharp growths, and tend to grow slowly in the later stage. Except the irregular Type IV, the other three types can be treated as two basic types: one has clear crossover from rapid to slow growth and the other shows stably exponential growth without such a pronounced crossover.

2. Effects of diversity of control in the model

The different growth patterns in the data can be qualitatively explained by the diverse effects of control in our model. To demonstrate this, we assume that parameters β_2 and g within countries are nonidentical and are randomly chosen from 0 to 1 for different nodes, while we fix the border control parameter β_1. In reality, all the important parameters ρ, β_1, β_2, g, and b can be different due to variation of contact structures (population, hygiene condition, culture, etc.) from country to country, but here we do not intend to fit the model precisely to the real data, but rather to demonstrate the concept and to prove the principle.

The results are summarized in Fig. 12 for two groups of nodes with early and late initial infections. In each group, we consider four combinations of the parameters β_1 and β_2. We can see that when β_1 is close to 1, the growth patterns are close to an exponential function regardless of β_2. While when β_2 is small, the growth will undergo

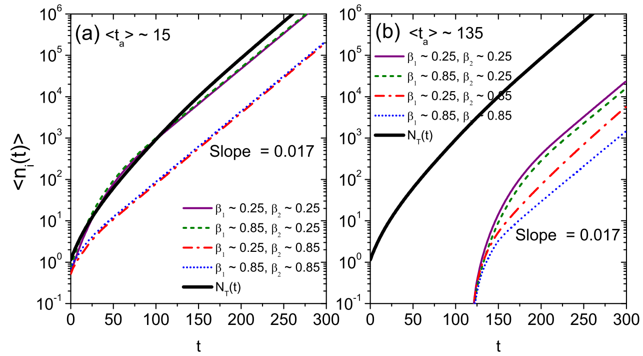

FIG. 12: The growth of averaged value of $n_i(t)$ for the nodes with different β_1, β_2 and the arrival time t_a. (a). For the nodes that $0 \leq t_a < 30$ ($t_a \sim 15$). (b). For the nodes that $120 \leq t_a < 150$ ($t_a \sim 135$). In the two panels, $\beta_1 \sim 0.25$ denotes the data averaged for the nodes that $0.20 \leq \beta_1 < 0.30$, and $\beta_1 \sim 0.85$ for $0.80 \leq \beta_1 < 0.90$, and also for the values of β_2. Simulations run with $\rho = 0.2$, $b = 0.06$, and $g = 0.2$. All the data are average over 10^4 independent runs.

a pronounced transition from fast to stably exponential function when β_2 is small, because it takes considerable time to change the local contact patterns (see Eq. 4). When β_2 is large, the growth will shift very quickly to the exponential function. The situation is the same for nodes with early or late initial infections except for the time-delay. In all the cases, the stable growth rates are close to ρg (the slope $\approx \rho g \log_{10} e$). Thus we can see that the two basic types of growth patterns in empirical data, i.e., with and without a pronounced transition, can be represented by different control parameters in the model. The model, however, does not include strong non-stationary ingredients that could lead to sudden increase of n_i observed in a few countries in Fig. 11(d).

We would like to point out that the growth patterns of n_i in the individual nodes at different parameters are similar to various growth patterns of the global total N_T in the model (Fig. 5(b) in the paper) and in empirical data (Fig. 2(d) in the paper). This provides justification that we an apply our model to the global level where each node represents a country, or to the level within a country where each node denotes a state/province. In the later case, N_T of the model represents the growth of the total cases of a country and is consistent with the growth of n_i in the former case.

E. Effects of heterogeneity in s_i

In the paper we discuss the effect of heterogeneity in s_i by taking s_i as the population of a country. Fig. 6 in the paper shows the evolution of the dependence between $n_i(t)$ and s_i in one realization of the model simulations. Fig. 13 display the statistics over many realizations: (a)

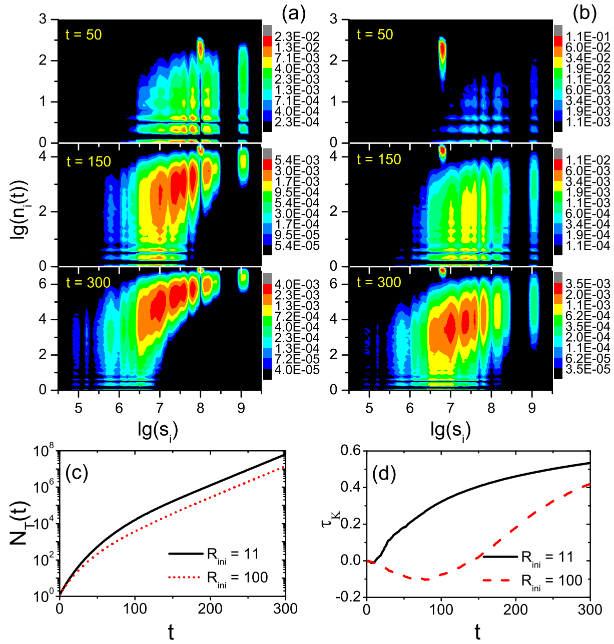

FIG. 13: Evolutions of the probability density in the space $(\log_{10} s_i, \log_{10} n_i)$ obtained from 10^4 realizations of simulations with the epidemic initiation at node with population rank $R_{ini} = 11$ (a) and $R_{ini} = 100$ (b). The corresponding growth of $N_T(t)$ (c) and evolution of Kendall's tau τ_K between $n_i(t)$ and s_i (d) averaged over all the realizations. Simulations run on $\beta_1 = 0.8$, $\rho = 0.2$, $b = 0.06$, and $g = 0.2$.

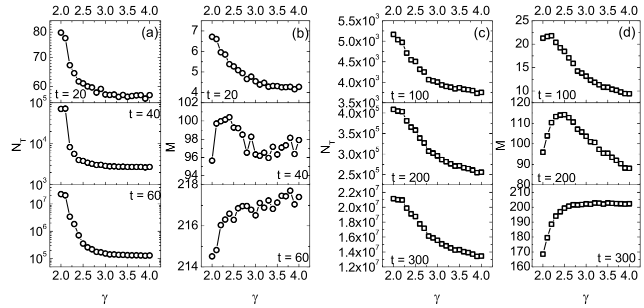

FIG. 14: Impacts of node heterogeneity (γ) on the total cases N_T and range M of the epidemic spreading without control (panels (a) and (b): $\beta_1 = 0$ and $\beta_2 = 0$) and with control (panels (c) and (d): $\beta_1 = 0.8$ and $\beta_2 = 0.2$). The simulations run on $\rho = 0.2$, $b = 0.06$, and $g = 0.2$.All of Data are averaged from 10^4 independent runs.

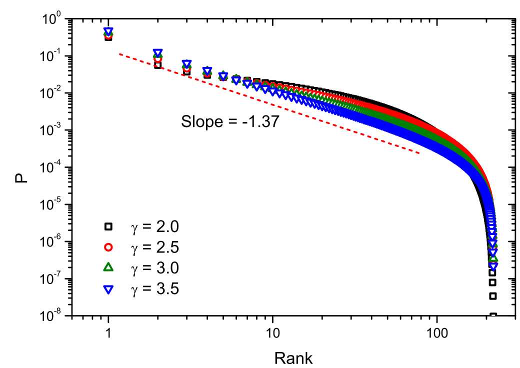

FIG. 15: The distributions P for different γ. Simulation runs on $\beta_1 = 0.8$, $\beta_2 = 0.2$, $\rho = 0.2$, $b = 0.06$, and $g = 0.2$. The red dashed line denotes the power-law with exponent -1.37. All of Data are averaged from 10^4 independent runs.

and (b) show the probability density function in the space $(\log_{10} s_i, \log_{10} n_i)$ with color scale. The spreading from the nodes with large s_i to those with small s_i becomes very evident in this presentation.

In the following we carry out a more systematic analysis of the impact of heterogeneous s_i and target control by considering power distributions of s_i, i.e., $P(s) \sim s^{-\gamma}$. Two spreading processes are compared. The first one is the situation without any control impacts, namely $\beta_1 = 0$ and $\beta_2 = 0$. In the second situation we consider strong border control ($\beta_1 = 0.8$ and $\beta_2 = 0.2$, typical parameter setting introduced in Fig. 5 in the paper).

Without control, the spreading is very fast even in the case of uniform s_i. Simulation results indicate that increased heterogeneity at smaller γ sharply accelerates the spreading by increasing the total cases N_T in all the spreading period (Fig. 14(a)). The impact of heterogeneity on the spreading range M is different in different period of the spreading: for stronger heterogeneity (smaller γ), M is larger in the early stage, but smaller in the later stage (Fig. 14(b)). With control, the spreading is significantly suppressed, and the acceleration by the heterogeneity is weaker: N_T does not increase so strongly when γ is smaller (Fig. 14(c)), while M displays similar non-monotonic, but relatively stronger dependence on γ (Fig. 14(d)). The non-monotonic impact of heterogeneity on M can be understood as follows. When s_i become rather heterogeneous, epidemic will rapidly arrive at the nodes with large s_i when initiated at a random node (see Fig. 13), so M is larger at smaller γ in the early stage. Then the epidemic mainly grows in a few nodes with the largest s_i and the majority of nodes with small s_i have rather weak connections between them, which make the spreading to new nodes more difficult, even though the total cases N_T is larger. In the situations with control, the spreading from these nodes have largest s_i and n_i to the nodes with small s_i is further reduced. As a result, the non-monotonic impact on M is more obvious in the situations with control. The enhanced spreading however, only makes the distribution P slightly more homogeneous (Fig. 15).

The comparison of these heterogeneous networks with

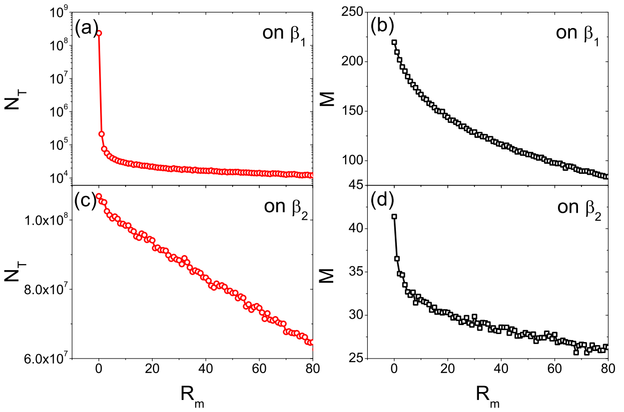

FIG. 16: (a) and (b), effects of the border control on the total cases N_T and range M vs. the number of large s_i nodes R_m (the maximum rank), simulations run on $\gamma = 2.0$, $\beta_2 = 0.2$, $\rho = 0.2$, $b = 0.06$, and $g = 0.2$, and $\beta_1 = 0.8$ for the nodes that rank of s_i equal or less than R_m, and $\beta_1 = 0$ for others. (c) and (d), effects of the local control on N_T and M vs. R_m, simulations run on $\gamma = 2.0$, $\beta_1 = 0.8$, $\rho = 0.2$, $b = 0.06$, and $\beta_2 = 0.2$ for the nodes that rank of s_i equal or less than R_m, and $\beta_2 = 0$ for others. All of Data are averaged from 10^4 independent runs.

the minimal models show that while strong heterogeneity in the nodes (countries) could be an accelerating factor, just like the effect of the heterogeneous degree distribution of complex networks [24], the strengths of control play a leading and dominant role in determining the epidemic spreading patterns. In fact, the impact of strong heterogeneity can be compensated with slightly increased border control parameter β_1.

Generally speaking, the nodes with large intensity on heterogeneous structures usually is a key towards the dynamics of the system. We have applied the target control to the first R_m nodes with the largest s_i. The spreading can be sharply decelerated by reducing both the total cases N_T and range M (Figs. 16(a) and (b)), when only a few nodes with the largest s_i are in strong border control (setting $\beta_1 > 0$ for the first R_m nodes and $\beta_2 = 0$ for others). Similar impact can be also observed in the local control on few large s_i nodes (setting $\beta_2 > 0$ for the front R_m nodes and $\beta_2 = 0$ for others, see Figs. 16 (c) and (d)). In this case, those nodes with $\beta_2 = 0$ grows very fast and the control on a few nodes does not reduce N_T very significantly. However, M is clearly reduced because the nodes with the largest s_i are usually the centers of spreading in the early stage and the target control within these nodes will reduce the spreading to other nodes. These impacts of control on the nodes with large s_i are quite similar to the targeted immunization strategy on hubs notes with the largest degrees in scale-free networks [27].

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant Nos. 10635040 and 10975126, the Specialized Research Fund for the Doctoral Program of Higher Education of China under Grant No. 20060358065. C.S.Z is supported by Hong Kong Baptist University. T.Z. acknowledges the valuable information provided by Joseph Wakeling and the National Natural Science Foundation of China under Grant No. 90924011.

[1] Neumann G, Noda T, Kawaoka Y (2009) Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* 459:931-939.

[2] Rohani P, Breban R, Stallknecht DE, Drake JM (2009) Environmental transmission of low pathogenicity avian influenza viruses and its implications for pathogen invasion. *Proc Natl Acad Sci USA* 106: 10365-10369.

[3] Fraser C, *et al.* (2009) Pandemic potential of a strain of Influenza A (H1N1): Early findings. *Science* 324: 1557-1561.

[4] Balcan D, *et al.* (2009) Seasonal transmission potential and activity peaks of the new influenca A (H1N1): a Monte Carlo likilihood analysis based on human mobility. *BMC Med* 7: 45.

[5] Hufnagel L, Brockmann D, Geisel T (2004) Forecast and control of epidemics in a globalized world. *Proc Natl Acad Sci USA* 101: 15124-15129.

[6] Masuda N, Konno N, Aihara K (2004) Transmission of severe acute respiratory syndrome in dynamical small-world networks. *Phys Rev E* 69: 031917.

[7] Small M, Walker DM, Tse CK (2007) Scale-free distribution of avian influenza outbreaks. *Phys Rev Lett* 99: 188702.

[8] Colizza V et al (2007) Modeling the Worldwide Spread of Pandemic Influenza: Baseline Case and Containment Interventions. *PLoS Med* 4(1): e13.

[9] Ferguson NM, Donnelly CA, Anderson RM (2001) Transmission Intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature* 413: 542-548.

[10] Ferguson NM, Donnelly CA, Anderson RM (2001) The Foot-and-Mouth Epidemic in Great Britain: Pattern of Spread and Impact of Interventions. *Science* 292: 1155-1160.

[11] Ferguson NM et al (2005) Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 437: 209-214.

[12] Longini Jr IM et al (2005) Containing Pandemic Influenza at the Source. *Science* 309: 1083-1087.

[13] Heaps HS (1978) *Information Retrieval: Computational and Theoretical Aspects* (Academic Press, Orlando).

[14] Zhang Z-K, Lü L, Liu J-G, Zhou T (2008) Empirical analysis on a keyword-based semantic system. *Eur Phys J B* 66: 557-561.

[15] Cattuto C, Barrat A, Baldassarri A, Schehr G, Loreto V (2009) Collective dynamics of social annotation. *Proc Natl Acad Sci USA* 106: 10511-10515.

[16] Lü L, Zhang Z-K, Zhou T (2009) Zipf's Law Results in Heaps' Law. unpublished.

[17] Barthélemy M et al (2004) Velocity and Hierarchical Spread of Epidemic Outbreaks in Scale-Free Networks. *Phys Rev Lett* 92: 178701.

[18] Yang R et al (2007) Epidemic spreading on heterogeneous networks with identical infectivity. *Phys Lett A* 364: 189-193.

[19] Wu ZL (2004) Scaling Law of SARS onset. *Int J Mod Phys B* 18: 2559-2563.

[20] Gross T, D'Lima CJD, Blasius B (2006) Epidemic dynamics on an adaptive network. *Phys Rev Lett* 96: 208701.

[21] Han XP (2007) Disease spreading with epidemic alert on small-world networks. *Phys Lett A* 365: 1-5.

[22] Moore C, Newman MEJ (2000) Epidemics and percolation in small-world networks. *Phys Rev E* 61: 5678-5682.

[23] Newman MEJ (2002) Spread of epidemic disease on networks. *Phys Rev E* 66: 016128.

[24] Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-Free networks. *Phys Rev Lett* 86: 3200-3203.

[25] Pastor-Satorras R, Vespignani A (2002) Immunization of complex networks. *Phys Rev E* 65:036104.

[26] Huerta R, Tsimring LS (2002) Contact tracing and epidemics control in social networks. *Phys Rev E* 66: 056115.

[27] Cohen R, Havlin S, ben-Avraham D (2003) Efficient immunization strategies for computer networks and populations. *Phys Rev Lett* 91: 247901.

[28] Zipf GK (1949) *Human Behaviour and the Principle of Least Effort: An introduction to human ecology* (Addison-Wesly, Cambridge).

[29] Newman MEJ (2005) Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* 46: 323-351.

[30] Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data. *SIAM Rev* (to be publsihed).

[31] Kendall M (1938) A New Measure of Rank Correlation. *Biometrika* 30: 81-89.