Community dynamics generates complex epidemiology through self-induced amplification and suppression

Zhenyuan Zhao[1,4], Juan Pablo Calderón[1,2,4], Chen Xu[3], Pak Ming Hui[3] & Neil F. Johnson[1]

[1]*Department of Physics, University of Miami, Florida FL 33126, U.S.A.*

[2]*Industrial Engineering, Universidad de Los Andes, Bogota, Colombia*

[3]*Department of Physics, The Chinese University of Hong Kong, Shatin, Hong Kong*

[4]*These authors contributed equally to the work*

The development of quantitative models of outbreaks is key to their eventual control[1,2], from human and computer viruses through to social (and antisocial) activities[3–8]. Standard epidemiological models can reproduce many general features of outbreaks[9]. Unfortunately, the large temporal fluctuations which often dominate real-world data are thought to require more complicated, system-specific models involving super-spreaders, specific social network topologies and rewirings, and birth-death processes[9–11]. However we show here that these large fluctuations have a generic explanation in terms of underlying community dynamics. Communities increasing (or decreasing) in size, act as instantaneous amplifiers (or suppressors) yielding a complex temporal evolution whose features vary dramatically according to the relative timescales of the community dynamics. We uncover, and provide an analytic theory for, a novel epidemiological phase transition driven by the population's response to an outbreak. An imminent epidemic will be suppressed if individual communities start to break up more frequently or join together less frequently, but will be amplified if the reverse is true.

There is great interest in developing models which can explain the time-dependence of infection levels in the outbreak of a biological virus and other social phenomena[1–9,12]. In the cases where there is some identifiable initial 'infection' (e.g. the first reported SARS case, the unveiling of a new webpage, the introduction of new explosive technology by Iraq insurgents, or the initiation of some particular activity or rumour[5,6,8]), it is reasonable to think in terms of susceptible individuals (S) subsequently becoming exposed to the virus, activity or information, undergoing a noticeable change in their symptoms or behaviour (i.e. infected, I) and then infecting others to whom they happen to be linked at that time. The infecteds then recover, or get removed, to a state where they can no longer infect others (R). As shown in Fig. 1 and the Supplementary Information, many real-world systems exhibit a surprisingly complex epidemiological response, accompanied by large fluctuations. The detailed origins of these large fluctuations are not in general known, yet their presence can have dramatic practical consequences – in particular if some unacceptably high threshold level suddenly gets exceeded or some disappearing phenomenon suddenly shows a resurgence. Our work suggests that such fluctuations emerge quite naturally as a result of the community dynamics within a population. The key feature of our analysis, in contrast to most existing works, is to allow for the fact that the links between individuals may change on a much faster, slower, or comparable timescale, to the underlying transmission process. Despite its simplicity, our dynamical community model produces infection profiles which span the entire spectrum of observed behaviours – from the large fluctuations in Fig. 1 through to the more standard unimodal SIR form[9] – simply by adjusting the timescales for the community dynamics. (See Supplementary Information for a full catalogue of behaviours).

Figure 2(a) illustrates our model, though we note that our main conclusions also hold for a wide variety of model variants[13,14]. In order to mimic the fact that modern-day human interactions feature long-distance travel and communication[1,2], in addition to rapidly evolving social networks in cyberspace, we choose the interactions in our model to be independent of any spatial lengthscale. In addition, our model allows the membership and size of a community, and the number of communities itself, to change over time on a timescale which is fast, slow or comparable to the intrinsic timescales τ_p and τ_q of viral transmission. At any one time, the links within an existing community are assumed to be strong, while those between communities are assumed to be so weak as to be negligible. The appropriate interpretation of a link, and hence community, will therefore depend on the application. For a biological virus, a meaningful community at time t would be the subset of people who happen to be together on a particular plane, or in a particular office, at that time. For a computer virus or rumour, it could be the subset who are logged on to a particular chatroom or website at that time. The population in our model, is the subset of the world's population who have a finite probability of coming in contact with each other in the near future. Given the application area in mind, the population size N may be quite large. At a given timestep t, a community is chosen randomly for possible membership change with a probability proportional to its size, in order to reflect the property that a priori anyone is equally likely to initiate such an event. All the possible real-world reasons for change in membership, are subsumed into two simple parameters: a community fragmentation probability $\nu_{\rm frag}$ and hence timescale $\tau_{\rm frag} \sim \nu_{\rm frag}^{-1}$; and a community coalescence probability $\nu_{\rm coal}$, and hence a timescale $\tau_{\rm frag} \sim \nu_{\rm frag}^{-1}$. During coalescence, another community is picked with probability again proportional to size, and the two join together. This

mimics the fact that two people in separate communities may suddenly become linked, and hence indirectly link together all members of both communities. Depending on $\tau_{\rm frag}/\tau_{\rm coal}$, the resulting community size-distributions may feature many small communities of size one or two, in which case the dominant coalescence and fragmentation processes correspond to a community losing or gaining just one or two individuals. At some arbitrary time $t = 0$ in the steady state, we infect one individual in the largest community (size N_c) but this can easily be varied. From then on, we apply the standard SIR process between individuals who happen to be linked at the timestep in question. This provides two further timescales: the standard timescale for viral transmission between a given infected and a susceptible in the same community ($\tau_p \sim p^{-1}$ where p is the standard SIR transmission probability[9]); and the standard timescale for recovery of a given infected ($\tau_q \sim q^{-1}$ where q is the standard SIR recovery probability[9]). More elaborate models are of course possible, but they would still feature an analogous set of timescales. Most importantly, the outputs in Figs. 1(c), (f), (i) and the Supplementary Information, confirm that we can obtain real-world behaviour without having to consider further complications such as hard-wired social network topologies, geographical heterogeneities, migration, fluctuating population size, super-spreaders, seasonalities, and birth-death processes. Indeed to our knowledge, alternative models with such features alone do not show such a wide range of realistic behaviour.

Figure 3 illustrates the sequence of events which drive the large fluctuations in a typical run from our model (blue curve in Fig. 3(a)). At a given timestep, the infection can only propagate along the transmission pathways which happen to be open at that timestep, and these will depend in turn on the instantaneous nature of the communities. It is tempting to imagine that the infection

profile (blue curve) can be reproduced by setting a stochastic SIR process on a suitably chosen complex network – for example, a weighted network obtained by aggregating the instantaneous community links (Fig. 2(b)) over some time-window T, as shown in Fig. 2(c). Such a network approach does not work, however, since it neglects the rapidly changing nature of the transmission pathways. The red curve corresponds to a stochastic SIR model on a $T = 1$ network (i.e. the $t = 0$ network in Fig. 2(b)) while the green curve is for $T = \infty$ as shown in Fig. 2(c). Results for other T values tend to lie in between these two curves. Figures 3(b) and 3(c) show that the infection profile is a subtle product of self-induced amplification and suppression events, leading to repeated resurgences and lingering infections. At any given time, an individual infected can become involved in a coalescence process with a largely susceptible community, thereby giving the community an initial infection seed. Before the infection dies out, this community then gets 'fed' with a fresh community of susceptibles, thereby amplifying the number of infecteds that it can produce. This larger community then fragments, releasing many infecteds who can then coalesce with other susceptible clusters. This whole process then repeats on multiple scales, yielding the observed large fluctuations. Elsewhere, we will report the results of a best-fit procedure for deducing the timescales $\tau_{\rm frag}$, $\tau_{\rm coal}$, τ_p and τ_q from the heights and separations of the fluctuation peaks across a wide range of empirical datasets and application areas.

Figure 4 explores the extent to which generalized continuous-time epidemiological theories[9] might account for these dynamical community processes at the level of the run-averaged behaviour. None of them can reproduce the initial jump caused by infection spread within the initially-infected community. A reasonable overall fit is obtained if the physiological transmission values (e.g. p

and q in the SIR model) are allowed to differ from their true values. However, the slower decay of the dynamical community model's run-averaged profile highlights the unique nature of the self-induced amplification effects. Interestingly, better agreement can be obtained if we introduce a fictitious double-infection state, $S \to I \to X \to I \to R$, where X is any non-infected intermediate state (see Supplementary Information). This suggests that in terms of run-averaged results, the effect of complex temporal dynamics in social space might be mimicked using continuous-time models at the expense of having to invoke a more complex (and possibly fictitious) physiology.

Ferguson recently emphasised the need to incorporate human reaction into epidemic models[1]. We now address this problem by allowing the population to spontaneously adjust its community dynamics (i.e. new $\nu_{\rm coal}$ and $\nu_{\rm frag}$) following a news announcement about the initial case. Prior to infection, we choose $\frac{\nu_{\rm frag}}{\nu_{\rm coal}}$ to be small since this guarantees a power-law community size-distribution as often observed in human systems[13–15]. Figure 5 reveals an unexpected phase transition, driven by the new values of $\nu_{\rm coal}$ and $\nu_{\rm frag}$. Defining the epidemic threshold as the point where the total number of infected individuals exceeds N_c, Fig. 5(c) in particular shows that epidemic spreading can now arise *below* the conventional threshold (i.e. $R_0 = p/q < 1$). We now outline a theory for this new transition, leaving fuller details to the Supplementary Information. In the fast community dynamics regime, the time-averaged and pair-averaged probability that any two objects are connected is given by $P \sim \frac{\nu_{\rm coal}}{N\nu_{\rm frag}}$. Since Fig. 5 only depends on long-time behaviours, we can adopt the continuous-time SIR theory with an effective infection probability given by $p \cdot P$ as opposed to p. The number of susceptibles in the long-time limit $S(\infty)$ with $N \gg 1$, is given by the solution $\overline{z}$

to the following generalized form of the standard self-consistent equation[9]:

$$z = \exp\left(-\frac{p\nu_{\rm coal}}{q\nu_{\rm frag}}(1-z)\right), \tag{1}$$

where $z \equiv S(\infty)/N$. The fraction $n_c \equiv N_c/N \ll 1$, hence the solution $\overline{z} \approx 1$ holds at the boundary. Expanding the exponential in Eq. (1) yields $\overline{z} = \frac{q\nu_{\rm frag}}{p\nu_{\rm coal}}$ and the boundary is described by $\frac{q\nu_{\rm frag}}{p\nu_{\rm coal}} = 1 - n_c$. Since $n_c \ll 1$, this boundary criterion becomes $R_0^* \equiv \frac{p\nu_{\rm coal}}{q\nu_{\rm frag}} = 1$, which yields an excellent approximation (dashed black lines in Fig. 5) to the exact phase boundary (white solid lines). This new epidemic threshold formula can be understood as follows. The standard SIR threshold $R_0 = p/q = 1$ corresponds to an infected's probability of transmission being equal to his probability of recovery. The effect of the fast community dynamics is to amplify the probability of transmission p by $\nu_{\rm coal}$, and the probability of recovery q by $\nu_{\rm frag}$ since fragmentation will postpone further infection of that community. Hence the renormalized threshold criterion becomes $(p\nu_{\rm coal})/(q\nu_{\rm frag}) = 1$. On average, a system which is below the conventional threshold (i.e. $p < q$) will yield an epidemic if $\nu_{\rm coal}/\nu_{\rm frag} \gg 1$, or equivalently $\tau_{\rm coal}/\tau_{\rm frag} \ll 1$. In terms of future control schemes, the imminent epidemic is *suppressed* by increasing the timescale for community coalescence with respect to the timescale for community fragmentation, but gets amplified if the reverse is true.

1. Ferguson, N. Capturing Human Behaviour. *Nature* 446, 733-733 (2007).

2. Hollingsworth, T.D., Ferguson, N.M., and Anderson, R.M. Will travel restrictions control the international spread of pandemic influenza? *Nature Medicine* 12, 497-499 (2006).

3. Levin, S. A., Grenfell, B., Hastings, A., and Perelson, A. S. Mathematical and computational challenges in population biology and ecosystems science. *Science* 275, 334-343 (1997).

4. Colizza, V., and Vespignani, A. Invasion Threshold in Heterogeneous Metapopulation Networks. *Phys. Rev. Lett.* 99, 148701 (2007).

5. Sornette, D., Deschâtres, F., Gilbert, T., and Ageon, Y. Endogenous Versus Exogenous Shocks in Complex Networks: An Empirical Test Using Book Sale Rankings. *Phys. Rev. Lett.* 93, 22 (2004).

6. Deschâtres, F., and Sornette, D. The Dynamics of Book Sales: Endogenous versus Exogenous Shocks in Complex Networks. *Phys. Rev. E* 72, 016112-016128 (2005).

7. Gross, T., Dommar, C., and Blasius, B. Epidemic Dynamics on an Adaptive Network. *Phys. Rev. Lett.* 96, 20-23 (2006).

8. McDonald, M., Suleman, O., Williams, S., Howison, S., and Johnson, N.F. Impact of unexpected events, shocking news, and rumors on foreign exchange market dynamics. *Phys. Rev. E* 77, 046110-046122 (2008).

9. Keeling, M., and Rohani, P. *Modeling Infectious Diseases in Humans and Animals*. (Princeton University Press, New York, 2007).

10. Small, M., Shi, P., and Tse, C.K. Plausible Models for Propagation of the SARS virus. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* E87-A, 2379-2386 (2004).

11. Watts, D.J., Muhamad, R., Medina, D.C., and Dodds, P.S. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proc. Natl. Acad. of Sci.* 102, 11157-11162 (2005).

12. Gonzalez, M.C., Hidalgo, C.A., and Barabási, A.L. Understanding individual human mobility patterns. *Nature* 453, 779-782 (2008).

13. Eguiluz, V.M., and Zimmermann, M.G. Transmission of Information and Herd Behavior: an Application to Financial Markets. *Phys. Rev. Lett.* 85, 5659-5662 (2000).

14. D'hulst, R. and Rodgers, G.J. Exact solution of a model for crowding and information transmission in financial markets. *Intern. Jour. Theor. and Appl. Fin.* 3, 609-612 (2000).

15. Johnson, N.F. Complexity in Human Conflict, in *Managing Complexity: Insights, Concepts, Applications* edited by Dirk Helbing (Springer, Berlin, 2008) p. 303-320.

Competing Interests The authors declare that they have no competing financial interests.

Correspondence Correspondence and requests for materials should be addressed to N.F.J. (email: njohnson@physics.miami.edu).

Figure 1 Outbreak dynamics in real-world systems. SARS in (a) and (d), the Code Red computer worm in (b), downloads of the Mozilla Firefox software in (g), news web portal hits in (h) and the sudden proliferation of IED explosion devices in Iraq in (e). Right-hand column ((c), (f), (i)) shows individual runs from our dynamical community model, in different parameter regimes. Details of the datasets, parameters, additional real-world examples, and a catalogue of additional runs, are given in the Supplementary Information.

Figure 2 Dynamical community model of outbreaks. (a) Schematic time-evolution of spreading in the presence of community dynamics (i.e. coalescence and fragmentation). Vertical axis shows the number of communities of a given size. (b) Instantaneous network representation at each timestep. (c) Weighted networks obtained by aggregating the network links over time-window T. As $T \to \infty$, the network becomes fully connected.

Figure 3 Infection profile from an individual run of our dynamical community model. (a) Blue curve is from Fig. 1(c) ($\nu_{\rm frag} = 0.05$, $\nu_{\rm coal} = 0.95$, $p = 0.001$ and $q = 0.001$). Using the same p and q values, the red curve corresponds to a stochastic SIR model on a static network obtained by time-aggregating the instantaneous network links over a time-window $T = 1$ (i.e. the $t = 0$ network in Fig. 2(b)). This is equivalent to stochastic SIR in a well-mixed population of size N_c. Likewise, the green curve is for $T = \infty$, which is equivalent to stochastic SIR in a well-mixed population of size N. (b) and (c) show the key dynamical process which yields amplification of the infection, and hence generates the realistic infection profiles of Fig. 1.

Figure 4 Run-averaged infection profile. Run-averaged results for our dynamical community model (1000 runs with $\nu_{\rm frag} = 0.001$, $\nu_{\rm coal} = 0.99$, $p = 0.01$ and $q = 0.001$) compared to best-fit versions for more standard SIR-like models. The parameter values (e.g. p and q) which are required to get such fits, are far from the actual values used for the dynamical community model.

Figure 5 Consequences of human reaction to an outbreak. Phase diagrams (a)-(c) show the numerically obtained phase transition (white line) separating spreading and no-spreading regimes, in a population which reacts to news of the initial infection by changing its community dynamics from $\nu_{\rm frag} = 0.001$ and $\nu_{\rm coal} = 0.99$, to the new values shown on the axes. Quantity shown is the fraction of the population who became infected over the lifetime of the disease, $R(\infty)/N$. (a) $p = 0.002$, $q = 0.001$ and hence $p > q$. (b) $p = 0.001$, $q = 0.001$ and hence $p = q$. (c) $p = 0.001$, $q = 0.002$ and hence $p < q$ which is below the conventional epidemic threshold. Black dashed line is the theoretical result $\frac{p\nu_{\rm coal}}{q\nu_{\rm frag}} = 1$. Shaded grey region is unphysical since $\nu_{\rm frag} + \nu_{\rm coal} > 1$.

Figure 1: Zhao et al.

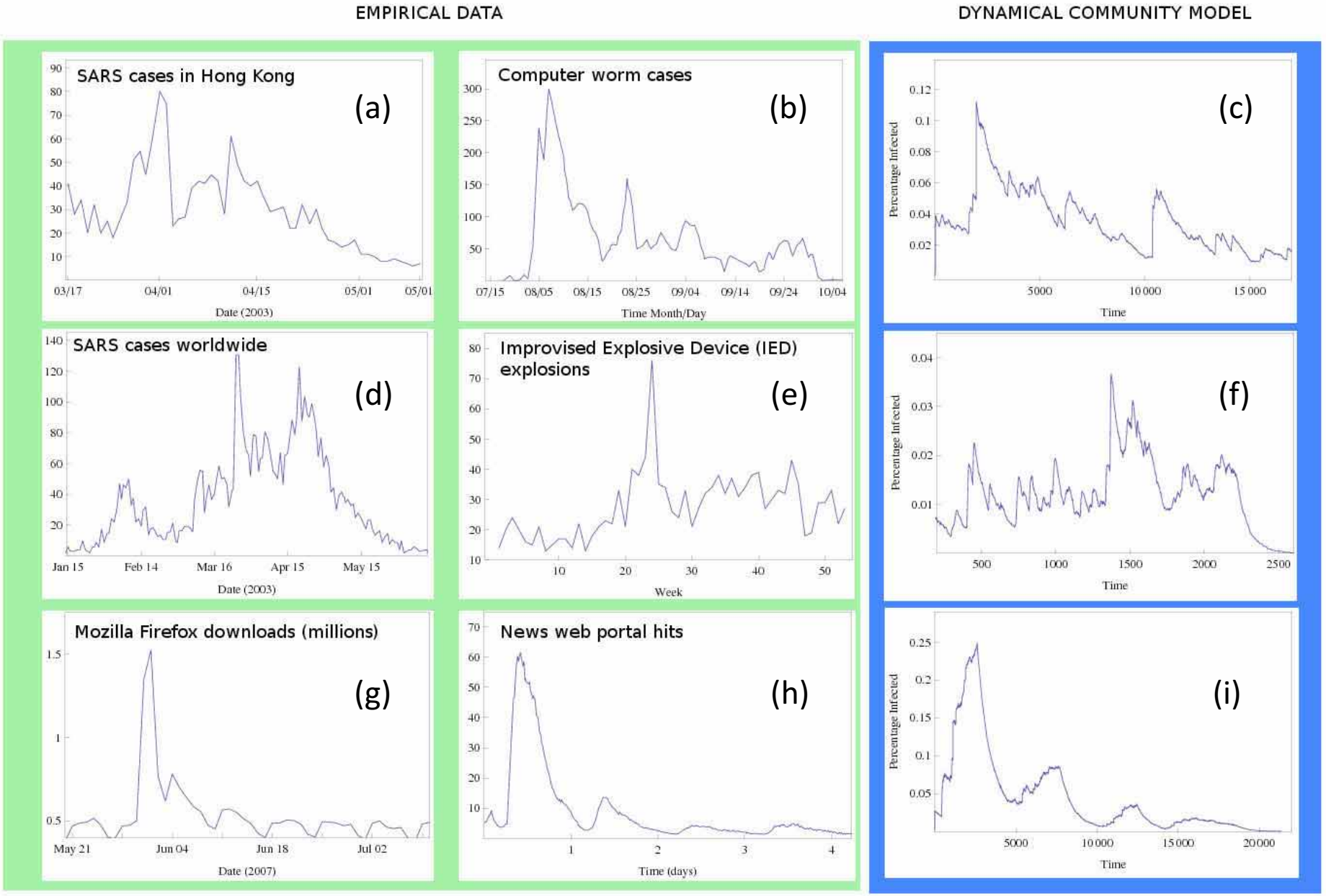


Figure 2: Zhao et al.

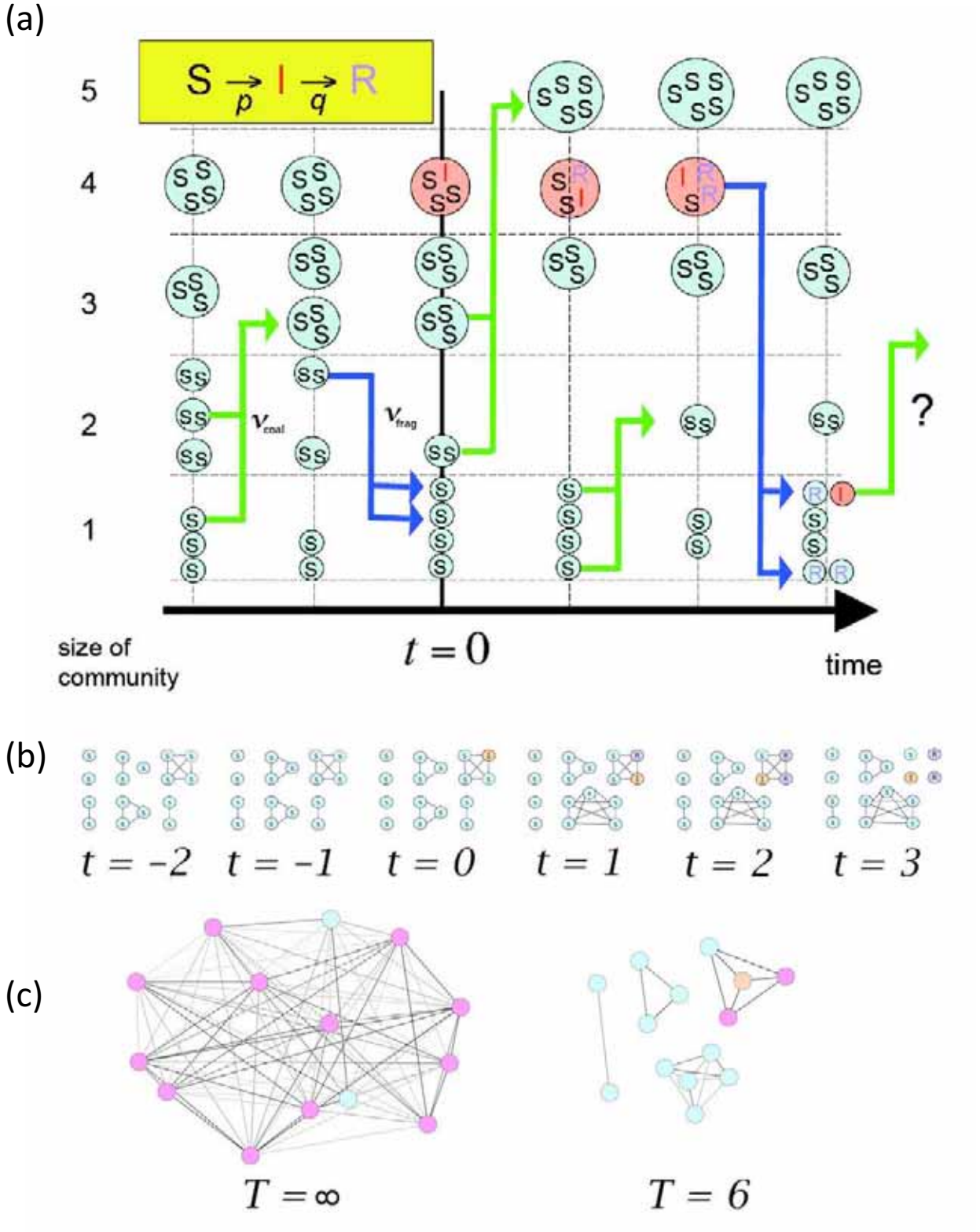


Figure 3: Zhao et al.

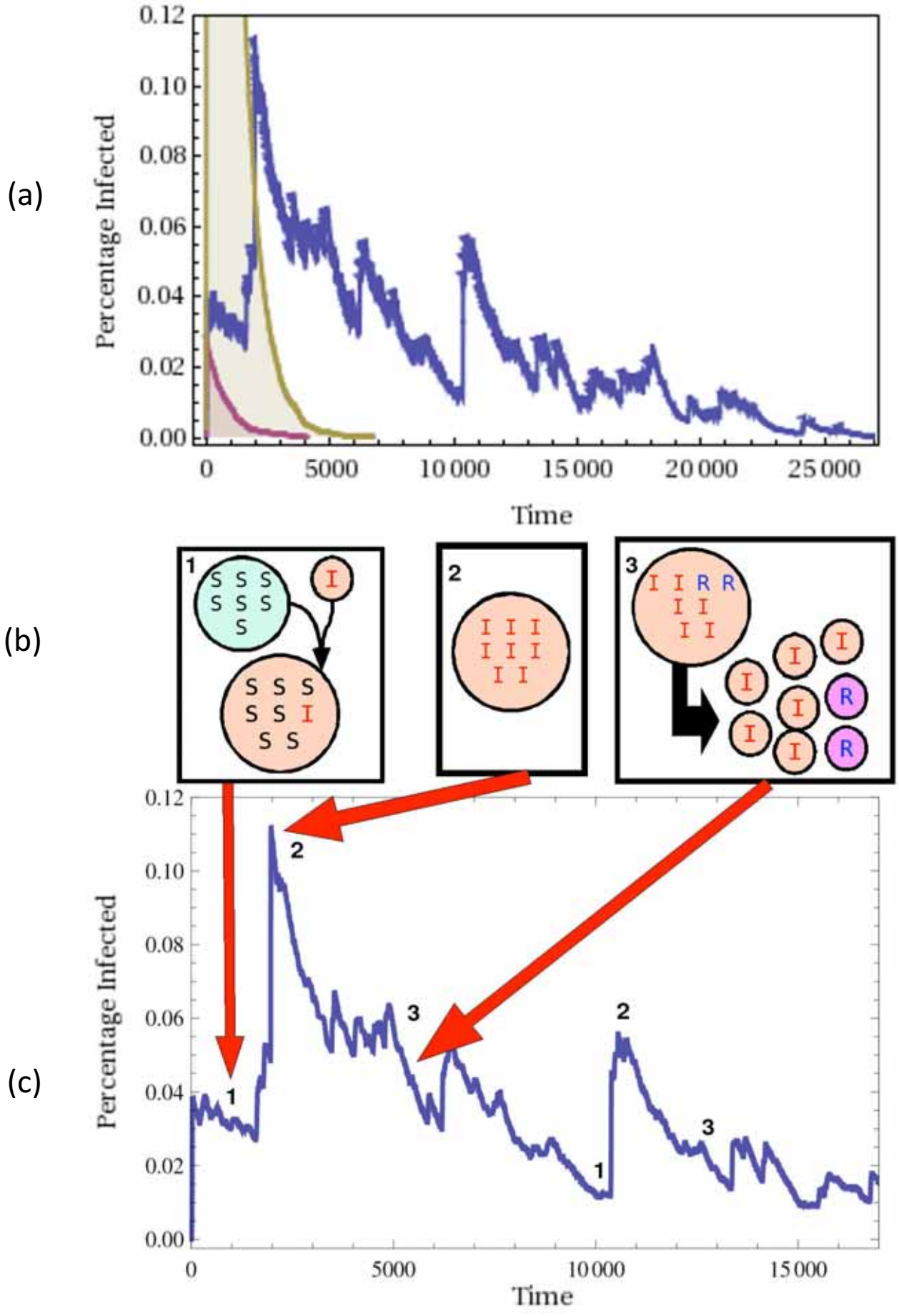


Figure 4: Zhao et al.

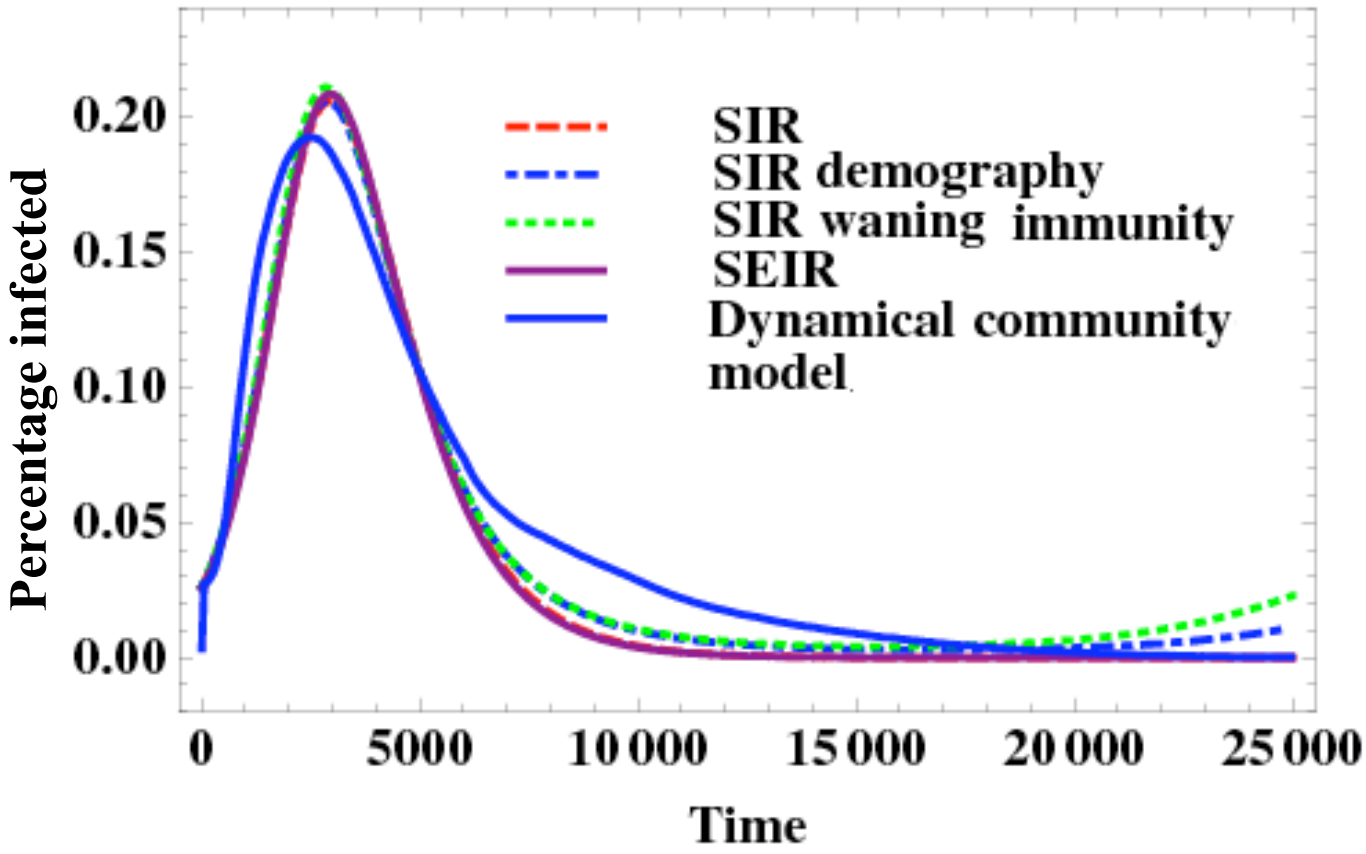
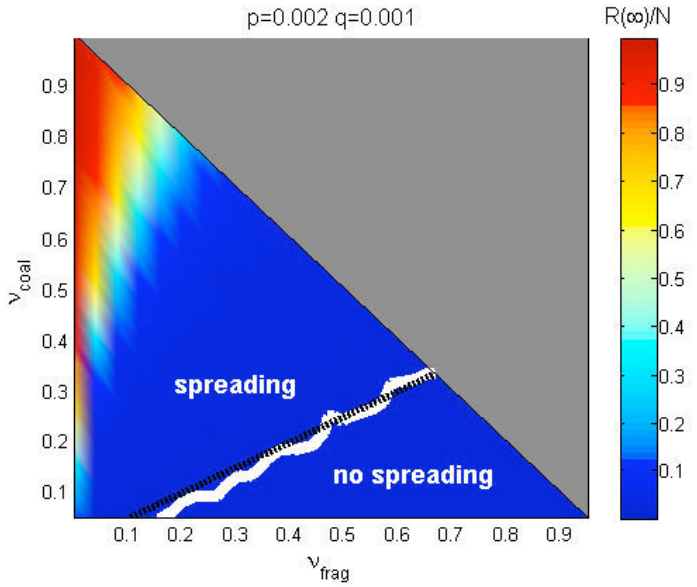
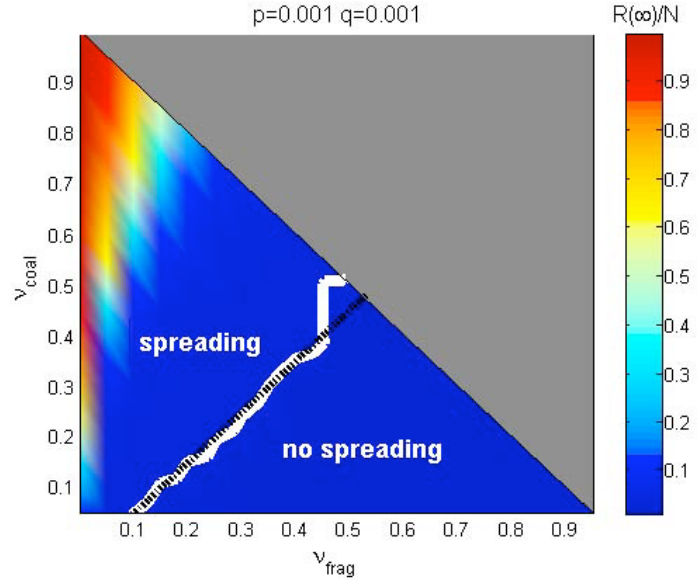


Figure 5: Zhao et al.

(a)



(b)



(c)