

Advantage of Being Multicomponent and Spatial: Multipartite Viruses Colonize Structured Populations with Lower Thresholds

Yi-Jiao Zhang,¹ Zhi-Xi Wu^{1,*}, Petter Holme,² and Kai-Cheng Yang³

¹*Institute of Computational Physics and Complex Systems, Lanzhou University, Lanzhou, Gansu 730000, China*

²*Tokyo Tech World Research Hub Initiative (WRHI), Institute of Innovative Research, Tokyo Institute of Technology, Yokohama 226-8503, Japan*

³*School of Informatics, Computing, and Engineering, Indiana University, Bloomington, Indiana 47408, USA*



(Received 6 February 2019; published 26 September 2019)

Multipartite viruses have a genome divided into different disconnected viral particles. A majority of multipartite viruses infect plants; very few target animals. To understand why, we use a simple, network-based susceptible-latent-infectious-recovered model. We show both analytically and numerically that, provided that the average degree of the contact network exceeds a critical value, even in the absence of an explicit microscopic advantage, multipartite viruses have a lower threshold to colonizing network-structured populations compared to a well-mixed population. We further corroborate this finding on two-dimensional lattice networks, which better represent the typical contact structures of plants.

DOI: [10.1103/PhysRevLett.123.138101](https://doi.org/10.1103/PhysRevLett.123.138101)

As the major cause of epidemics, viruses are posing a great threat to all living organisms. Among all types of viruses, multipartite (also named multiparticulate, multicomponent, or multicompartiment) viruses possess the strangest genetic organizations. Unlike other categories, such as nonsegmented and segmented viruses whose nucleic acid segments are encapsidated into a single virion (viral particle) that propagates as a whole, multipartite viruses have two or more segmented genomes packaged into separate virions, and each one of them can propagate independently [1]. Still, the complete replication cycle of a multipartite virus requires the full genome so that the concurrent presence of multiple segments is necessary for a successful infection [2]. This need for a multiplicity of infection events comes with an evolutionary disadvantageous cost, and factors counterbalancing it are an area of active research [3].

It has been shown that a large quantity of viruses, about 40% families and 17% species, do have multipartite genomes [4]. In the mean time, the number of known multipartite species is rapidly growing: 48 new species of multipartite viruses were identified from 2014 to 2015 [4]. All these facts signal that multipartitism is a successful evolutionary strategy. Apart from the peculiar genetic organization, perhaps the most striking characteristic about multipartite viruses is their strong preference for host populations—more than 90% of known multipartite virus species exclusively target plants and fungi [1,4], and only very few of them can infect animals, such as the recently discovered Jingmen tick virus [5] and the Guaico Culex virus [6]. The evolutionary origin of multipartite viruses and why they mainly infect plants remain puzzling even today.

To explain the evolutionary success of multipartitism, researchers have pointed out several possible advantages.

Still, we lack a complete theory of the evolutionary advantage of such a strange lifestyle [4]. Researchers have hypothesized that multipartitism leads to smaller genetic segments that may have lower mutation probability [7] and faster replication ability [8], the stability of the capsid would be higher for smaller virions [9], small segments facilitate the recombination of genomes in order to adapt new environment [10], and so on. However, solid experimental evidence for these possible advantages is scarce [4].

The theory of multipartite viruses has usually been carried out in a game theoretical framework. With this approach, one models the success of monopartite and multipartite forms in terms of their ability to spread and persist in a host population [3,11–14]. Such models typically conclude that multipartitism occurs as a way to compensate for the cost of high multiplicity of infection [3]. Specifically, it was shown that having multipartitism might be a successful adaptive strategy for viruses in new environments, especially in homogeneous spatial structure [14].

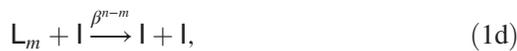
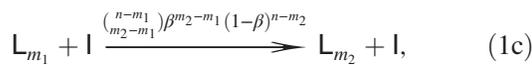
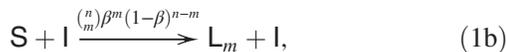
In this Letter, we approach this puzzle from the point of view of macroscopic ecological dynamics, thus seeking an alternative way to understand the rise and persistence of multipartitism. To this end, we propose a minimal model in the framework of network epidemiology [15–17]. We assume that all the virions are epidemically equivalent and that there is no competition dynamics among them that would complicate the situation. Instead, we focus on the effects of the contact structures of the hosts on the propagation of multipartite viruses. We construct interaction networks modeling different interaction patterns of real plant and animal populations.

To model the propagation of multipartite viruses into a networked epidemiological model, we first note that—unlike

monopartite viruses whose genetic material is wrapped up in a single virion that remains intact during the course of spreading—multipartite viruses have segmented genomes encapsidated in different virions that, in principle, propagate independently [1]. Successful propagation demands that all the different segments eventually reach the same host cell to complete the viral life cycle [2,13,14]. In such a scenario, the hosts can be divided into four compartments: susceptible (**S**, hosts uninfected by any virion); latent (**L**, those who are hosts of at least one, but not all, the different types of virions); infectious (**I**, those who have acquired the full viral genome and become infectious); removed (**R**, those who recovered or died from the viral infection).

We try two different types of underlying interaction structures—a static and an annealed network model [16–24]. In the latter case, the network connections are effectively reshuffled between any microscopic steps of the epidemic dynamics (while keeping the number of neighbors of each node fixed). The annealed structure mimics the interaction patterns in animals where the underlying network changes much faster than the propagation process of the virus. In contrast, a static topology is adopted for the interaction of the skeleton of plants. Though most multipartite viruses are transmitted by vector movements among plants [25], such a simplification is reasonable since plants are spatially distributed on the surface of earth, and, macroscopically, the interaction structure among plants changes at a much slower time scale compared to that of animals.

Suppose a multipartite virus is composed of n separately packaged nucleic acid segments. Let L_m denote the L -state individuals carrying m ($0 < m < n$) segments of the genome. I -state individuals can infect **S** and L_m ($m < n$) individuals, and recover at a constant rate μ . Upon contact with infected hosts, susceptible hosts may become either latent or infected (dependent on how many different types of virions have been transmitted). Similarly, latent hosts may acquire other viral segments by contacting with the infected ones, thus increasing m . If $n = m$, they become infectious. Finally, infected hosts will pass to the recovered state at some rate. We assume all types of virions have the same infection rate β when an infectious individual interacts with an individual in the **S** or **L** state. Then, the above reactions of the SLIR model can be summarized as



where m_1, m_2 are integers and satisfy $0 < m_1 < m_2 < n$. A graphical illustration of the SLIR model with $n = 3$ is shown in the Supplemental Material [26].

For simplicity, we do not account for higher-order plant-virus-vector interactions [27]. Moreover, in our model, all types of virions are epidemically equivalent, and each one of them can propagate independently without the involvement of any microscopic adaptive advantages, such as the competition, cooperation, and commensality [14].

It has been shown that a majority of multipartite viruses have two components, while those consisting of more than four segments are rare [4]. For this reason, we primarily use $n = 2$. Details of the modeling can be found in the Supplemental Material [26]. For the numerical simulation, we use the standard Gillespie algorithm [28]. In our study, unless otherwise stated, we let the networks have 10^5 to 10^6 nodes. Each data point presented below is an average over 10^3 to 10^4 runs of the simulation. In each realization, one node is randomly selected to be the seed of the infection, with all other nodes being susceptible.

We start our discussion with results from the annealed networks. The degree-based mean-field approach [15,16,29] is applied to analyze the spreading dynamics.

We observe the rich phenomena of the spreading on annealed networks. In particular, there exist two crucial β values, β_{c1} and β_{c2} , that divide the phase space into different regimes, see Fig. 1(a). When $\beta < \beta_{c1}$, the average fraction of recovered nodes at the end of the epidemic process, denoted by R_∞ , has only one stable fixed point 0, indicating a disease free state. When $\beta_{c1} < \beta < \beta_{c2}$, the stable fixed point 0 remains and two new fixed points R_- (unstable), R_+ (stable) appear. When $\beta > \beta_{c2}$, R_∞ has one unstable fixed point 0 and one stable fixed point R_+ . In this case, the system has a sudden transition to a state where it always evolves to R_+ , indicating that global outbreaks are

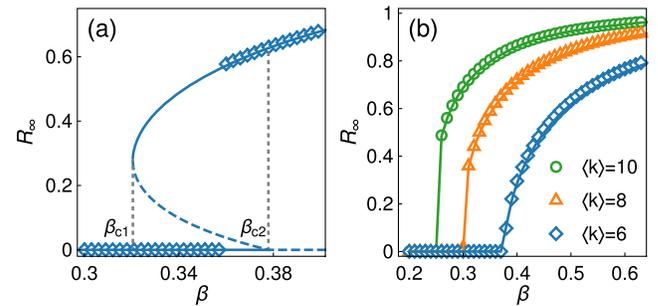


FIG. 1. Average density of recovered nodes R_∞ as a function of β for the SLIR process taking place on (a) annealed ER networks with a mean degree $\langle k \rangle = 6$, and (b) static ER networks with $\langle k \rangle = 10, 8, 6$. In panel (a), lines are theoretical results, where the stable solutions are represented by solid lines and unstable solutions by dashed lines. The diamonds are the results from stochastic simulations. In panel (b), the solid lines are numerical solutions from pair approximation equations and the symbols are simulation results.

inevitable. Thus, β_{c2} can be regarded as the epidemic threshold, and can be obtained by means of heterogeneous mean-field theory as

$$\beta_c^{\text{annealed}} = \beta_{c2} = \sqrt{\frac{\langle k \rangle}{\langle k^2 \rangle}}, \quad (2)$$

where $\langle k \rangle$ is the average degree of the interaction network. We perform extensive simulations on annealed Erdős-Rényi (ER) networks to validate our theoretical results. The numerical and theoretical thresholds agree well, see Fig. 2 and the Supplemental Material [26]. For annealed networks, we find theoretically that a discontinuous phase transition only appears when $\langle k \rangle > 1.91$. Most real interaction networks probably fulfill this criterion.

It is worth mentioning that for monopartite (including nonsegmented and segmented) viruses, the SLIR model reduces to the SIR model [16]. Consider an SIR model with infection rate β and recovery rate $\mu = 1$, the epidemic threshold on annealed networks is $\beta_{\text{SIR}}^{\text{annealed}} = \langle k \rangle / \langle k^2 \rangle$ [16].

Thus we have $\beta_c^{\text{annealed}} = \sqrt{\beta_{\text{SIR}}^{\text{annealed}}} > \beta_{\text{SIR}}^{\text{annealed}}$, since $\beta_{\text{SIR}}^{\text{annealed}} < 1$. In terms of epidemic threshold, monopartite viruses clearly have advantage over multipartite ones in propagation which coincides with our intuition. The major difference lies in the nature of the transition from epidemic-free state to global outbreak state. The SIR model always exhibits continuous transition whereas SLIR model shows discontinuous transition when the network is relatively dense. The discontinuous transition means that, just above the threshold, the multipartite viruses can spread to a large fraction of the population, making it difficult contain by interventions.

Now we turn to static networks modeling the situation when the interaction pattern of hosts changes slower than the propagation of the virus. A mean-field method is then no longer applicable since it neglects correlation between the dynamical states of the hosts in static networks; i.e.,

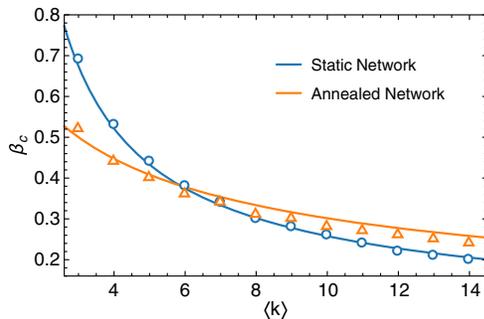


FIG. 2. Epidemic (or colonization) thresholds of the SLIR model on annealed and static ER networks with the increase of mean degree $\langle k \rangle$. The lines are theoretical results predicted, respectively, by Eqs. (2) and (3). The symbols represent the results from stochastic simulations.

a node is more likely to be infected if many of its neighbors are infected [30]. Thus, we resort to the pair approximation approach [31–33], which explicitly accounts for the dynamic correlations coming from infected pairs [30] (correlations that are by definition lacking in annealed networks [15]).

The outcomes of the propagation of multipartite viruses on static ER networks are shown in Fig. 1(b). The solid lines are numerical solutions of pair approximation equations, see details in the Supplemental Material [26]. The symbols represent simulation results which agree well with the analytical solutions. Figure 1(b) also displays that sudden outbreaks appear when the mean degree of the underlying interaction networks becomes sufficiently large, e.g., $\langle k \rangle = 8$ and 10. Yet, the transition in static networks is not discontinuous as in annealed networks since no saddle point bifurcation phenomenon occurs.

By solving the pair approximation equations, we are able to find the epidemic outbreak threshold β_c^{static} on static networks satisfies the following cubic equation

$$(1 - \langle k \rangle)\beta^3 + (3\langle k \rangle - 1)\beta^2 - 3\beta - 1 = 0. \quad (3)$$

A simple form of β_c^{static} is not available, but numerical solutions of Eq. (3) can be obtained easily. In Fig. 2, we compare the epidemic thresholds, obtained through either theoretical analysis or direct simulations, for the SLIR model on static networks and the annealed counterparts [34]. Interestingly and surprisingly, there is a crossover between β_c^{static} and $\beta_c^{\text{annealed}}$ as $\langle k \rangle$ grows. Specifically, we prove that the following relation always holds (see the Supplemental Material [26])

$$\beta_c^{\text{static}} < \beta_c^{\text{annealed}}, \quad (4)$$

when $\langle k \rangle$ satisfies

$$\langle k \rangle > \frac{5 + 3\sqrt{5}}{2} \approx 5.84. \quad (5)$$

Intuitively, epidemic spreading on annealed networks should happen easier than that on static networks because the frequent switching of neighbors enables the infectious nodes to contact more susceptible ones, which increases the basic reproduction number [30]. By contrast, our results here suggest that static topology favors the propagation of multipartite viruses than the annealed topology whenever local interactions are relatively dense (average interaction degree is greater than 5.84).

From the above, we know that the presence of L-state nodes induces both the discontinuous phase transition in annealed networks and the lower epidemic thresholds in dense static networks. On one hand, the frequently reshuffling of the annealed networks raises the chance for infectious hosts to come in contact with susceptible or latent hosts. On the other hand, it also pushes the L-state nodes far away from their infectious neighbors. Recall that L-state nodes are not

infectious due to their lack of a full genome. Thus, before the approach of a tipping point, L-state nodes are distributed evenly in the system, very likely resulting in pervasive outbreaks when the infection rate exceeds the critical value. The mechanism behind this dynamic phenomenon is akin to the explosive synchronization observed in multiplex networks [35], where microscopically suppressive rules give rise to macroscopic discontinuous transitions.

By contrast, on static networks, L-state nodes are restricted to stay at the surroundings of l-state nodes, which benefits prominently the coinfection of different virions, promoting in turn the successful propagation of the multipartite virus. This effect is especially pronounced on dense networks causing the lower thresholds on static networks in comparison to an annealed counterparts. More detailed analysis of the role of L-state nodes in epidemic spreading can be found in the Supplemental Material [26].

So far, we have mainly focused on bipartite viruses. It is straightforward to extend our theoretical analysis to larger numbers of viral segments. To illustrate this, Fig. 3 shows results for the tripartite case, with the same conclusions as the bipartite viruses.

In order to confirm that our findings are not bounded to random networks, we also consider the SLIR model on two-dimensional lattice networks (see graphical illustrations in the Supplemental Material [26]). Such geographically embedded networks are better models of the interaction structures between plants. To compare with annealed networks we investigate constantly rewiring regular random graphs (r-RRGs). Our simulations show that large outbreaks not only happen on rewiring regular random graphs but also on lattice networks provided the mean degree of interaction is high enough. Lower epidemic thresholds are once again verified in the simulations for dense static interaction networks, serving as a promising explanation for why multipartite viruses primarily infect plants, see Fig. 4.

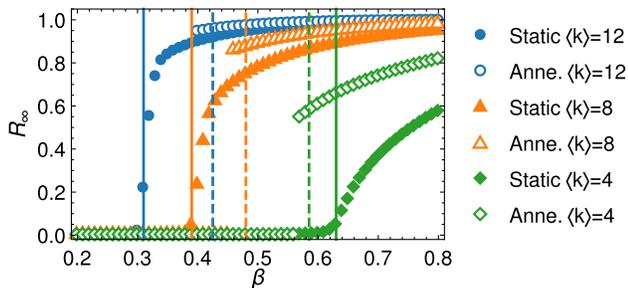


FIG. 3. Average density of recovered nodes R_∞ as a function of β for the SLIR process with $n = 3$ on static and annealed ER networks with a different mean degree $\langle k \rangle$. Filled and open symbols correspond, respectively, to the results obtained through stochastic simulations on static and annealed networks. Vertical solid and dashed lines show the theoretical colonization thresholds for static and annealed networks, respectively.

To clarify the critical behavior of the SLIR model on static networks, we carried out extensive simulations to measure the critical exponent α , characterizing the relationship $R_\infty \sim (\beta - \beta_c)^\alpha$ near the critical point. Basically, we find that a large exponent α of the SLIR model on static ER networks in comparison to the SIR model, suggesting that multipartite viruses above the threshold spread more pervasively compared to monopartite viruses [26].

In summary, we have constructed and analyzed a generic minimal SLIR model for the propagation of multipartite viruses on networked populations. We found discontinuous phase transitions for the SLIR model on annealed networks, whereas the transitions are continuous, but with a possibility of global outbreaks, on static networks. This implies that outbreaks of multipartite viruses can be very pervasive once the epidemic threshold is reached. The type of transition of the SLIR model is probably related to the turnover rate of L-state nodes, but a precise characterization is out of the scope of this Letter. The epidemic thresholds of the SLIR model on annealed and static networks are obtained analytically by solving the heterogeneous mean field equations and pair approximation equations, respectively. We showed that the epidemic thresholds on static networks are smaller than those on the annealed counterparts whenever the average degree of interaction becomes *sufficiently* large. The intuition is that locally dense and stable contacts makes

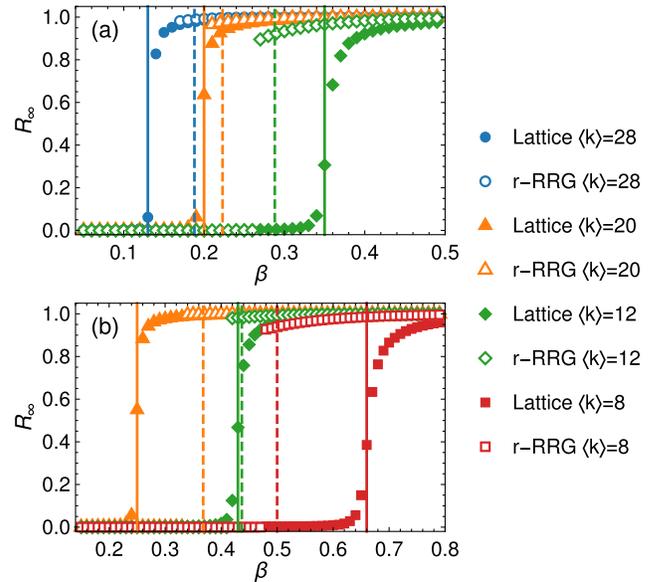


FIG. 4. Average density of recovered nodes R_∞ as a function of β for the SLIR process on two-dimensional square lattices and their annealed counterparts (i.e., frequently rewiring regular random graphs) with different mean degree $\langle k \rangle$. Panels (a) and (b) are for multipartite viruses with two and three viral segments, respectively. The filled and open symbols correspond, respectively, to the results obtained through stochastic simulations on static and annealed structures. The solid and dashed lines show the colonization thresholds yielded by theoretical predictions for static and annealed structures, respectively.

the delivery of the entire genome likelier. When the network changes faster, more hosts will be infected by some types of virions, but fewer will obtain the whole viral genome.

In this Letter, we used a network epidemiological framework to explain the viability of multipartite viruses, despite their apparently costly lifestyle, and why they primarily infect plants. We found that, although multipartite viruses need to be more infectious than monopartite viruses in order to spread out in general, the outbreak is usually more pervasive once the threshold is reached. Moreover, we show that multipartite viruses are mostly favored to spread in plants whose contact structures remain relatively static during the viral cycle. The counter-intuitive behavior is believed to be particularly relevant to the multipartite genetic organization of the viruses themselves.

Our current model has included as few virological features as possible. For instance, we ignored the possibility of a difference in transmissibility of different types of virions. Nonetheless, even in the absence of the explicit microscopic advantage, multipartitism is verified to be advantageous for the viruses to colonize (globally) structured populations. Our work thus provides a promising avenue to elucidate the rise and persistence of multipartitism from the macroscopic ecological dynamics viewpoint. It is convenient to incorporate more realistic mechanisms, in terms of empirical knowledge of the molecular, ecological, and evolutionary features, into our minimal model to better characterize the propagation of multipartite viruses in the real world [4,36].

Finally, even though multipartite viruses mainly target plants, the recent discovery of multipartite viruses on mosquitoes and other insects hints at the possibility of transmitting to animals as well as human beings, which calls for elaborated extensions of our minimal model [5,6].

This work was supported by the National Natural Science Foundation of China (Grant No. 11975111). P.H. was supported by JSPS KAKENHI Grant No. JP 18H01655.

*Corresponding author.
eric0724@gmail.com

- [1] A. Sicard, Y. Michalakis, S. Gutiérrez, and S. Blanc, *PLoS Pathogens* **12**, e1005819 (2016).
- [2] J. García-Arriaza, S. C. Manrubia, M. Toja, E. Domingo, and C. Escarmís, *J. Virol.* **78**, 11678 (2004).
- [3] A. Lucía-Sanz, J. Aguirre, and S. Manrubia, *Curr. Opin. Virol.* **33**, 89 (2018).
- [4] A. Lucía-Sanz and S. Manrubia, *npj Syst. Biol. Appl.* **3**, 34 (2017).
- [5] X.-C. Qin, M. Shi, J.-H. Tian, X.-D. Lin, D.-Y. Gao, J.-R. He, J.-B. Wang, C.-X. Li, Y.-J. Kang, B. Yu, D.-J. Zhou, J. Xu, A. Plyusnin, E. C. Holmes, and Y.-Z. Zhang, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6744 (2014).
- [6] J. T. Ladner *et al.*, *Cell Host Microbe* **20**, 357 (2016).
- [7] J. Pressing and D. C. Reaney, *J. Mol. Evol.* **20**, 135 (1984).
- [8] S. Nee, *J. Mol. Evol.* **25**, 277 (1987).
- [9] S. Ojosnegros, J. Garca-Arriaza, C. Escarmis, S. C. Manrubia, C. Perales, A. Arias, M. G. Mateu, and E. Domingo, *PLoS Genet.* **7**, e1001344 (2011).
- [10] L. Chao, *J. Theor. Biol.* **153**, 229 (1991).
- [11] S. Martín and S. F. Elena, *J. Gen. Virol.* **90**, 2815 (2009).
- [12] S. F. Elena, G. P. Bernet, and J. L. Carrasco, *Curr. Opin. Virol.* **8**, 62 (2014).
- [13] J. Iranzo and S. C. Manrubia, *Proc. R. Soc. Ser. B: Biol. Sci.* **279**, 3812 (2012).
- [14] E. Valdano, S. Manrubia, S. Gómez, and A. Arenas, *PLoS Comput. Biol.* **15**, e1006876 (2019).
- [15] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, *Rev. Mod. Phys.* **87**, 925 (2015).
- [16] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. Lett.* **86**, 3200 (2001).
- [17] M. E. J. Newman, *Phys. Rev. E* **66**, 016128 (2002).
- [18] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, in *Proceedings of the 22nd International Symposium on Reliable Distributed Systems, 2003* (IEEE, Florence, 2003), pp. 25–34.
- [19] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno, *Europhys. Lett.* **89**, 38009 (2010).
- [20] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, *Phys. Rev. Lett.* **85**, 4626 (2000).
- [21] M. Boguñá, C. Castellano, and R. Pastor-Satorras, *Phys. Rev. Lett.* **111**, 068701 (2013).
- [22] C. Castellano and R. Pastor-Satorras, *Phys. Rev. Lett.* **105**, 218701 (2010).
- [23] A. V. Goltsev, S. N. Dorogovtsev, J. G. Oliveira, and J. F. F. Mendes, *Phys. Rev. Lett.* **109**, 128702 (2012).
- [24] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, England, 2008).
- [25] B. Moury, F. Fabre, and R. Senoussi, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17891 (2007).
- [26] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.123.138101> for the detailed definition and the analytical treatments of the SLIR model on both annealed and static networks. Additional quantitative discussions, analysis, and explanations of the results are also discussed.
- [27] R. G. Dietzgen, K. S. Mann, and K. N. Johnson, *Viruses* **8**, 303 (2016).
- [28] W. Cota and S. C. Ferreira, *Comput. Phys. Commun.* **219**, 303 (2017).
- [29] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Rev. Mod. Phys.* **80**, 1275 (2008).
- [30] C.-R. Cai, Z.-X. Wu, M. Z. Q. Chen, P. Holme, and J.-Y. Guan, *Phys. Rev. Lett.* **116**, 258301 (2016).
- [31] M. J. Keeling, *Proc. R. Soc. Ser. B: Biol. Sci.* **266**, 859 (1999).
- [32] K. T. D. Eames and M. J. Keeling, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13330 (2002).
- [33] I. Z. Kiss, G. Röst, and Z. Vizi, *Phys. Rev. Lett.* **115**, 078701 (2015).
- [34] It is also interesting to compare the epidemic threshold of the SLIR model to that of the SIR model on heterogeneous

static networks. According to Refs. [15,30], the upper limit of the epidemic threshold of the SIR model on finite-size static networks is $\beta_{\text{SIR}}^{\text{static}} = [\langle k \rangle / (\langle k^2 \rangle - 2\langle k \rangle)]$. Via some straightforward algebra, one can verify $\beta_{\text{SLIR}}^{\text{static}} > \beta_{\text{SIR}}^{\text{static}}$ for any $\langle k \rangle > 1$ [26].

- [35] X. Zhang, S. Boccaletti, S. Guan, and Z. Liu, *Phys. Rev. Lett.* **114**, 038701 (2015).
- [36] T. Hayakawa, K. Kojima, K. Nonaka, M. Nakagaki, K. Sahara, S. Asano, T. Iizuka, and H. Bando, *Virus. Res.* **66**, 101 (2000).