Invasion Threshold in Heterogeneous Metapopulation Networks

Vittoria Colizza[1] and Alessandro Vespignani[2]

[1]*Complex Networks Lagrange Laboratory, Institute for Scientific Interchange, Torino 10133, Italy*
[2]*Complex Systems Group, School of Informatics, Indiana University, Bloomington, Indiana 47406, USA*

We study the dynamics of epidemic and reaction-diffusion processes in metapopulation models with heterogeneous connectivity patterns. In susceptible-infected-removed-like processes, along with the standard local epidemic threshold, the system exhibits a global invasion threshold. We provide an explicit expression of the threshold that sets a critical value of the diffusion/mobility rate below, which the epidemic is not able to spread to a macroscopic fraction of subpopulations. The invasion threshold is found to be affected by the topological fluctuations of the metapopulation network. The results presented provide a general framework for the understanding of the effect of travel restrictions in epidemic containment.

The role of heterogeneity has been acknowledged as a central question in the study of population biology of infectious diseases [1–3] and revamped recently with the evidence that a large number of real world networks exhibit complex topological properties [4–6]. These features, often mathematically encoded in a heavy-tailed probability distribution $P(k)$ that any given node has degree k, were shown to affect the system evolution, altering the threshold behavior and the associated dynamical phase transition [7–9]. These studies have mainly focused on networked systems where each node corresponds to a single individual, and only recently the study of the impact of heterogeneous topologies on bosonic systems, where nodes can be occupied by any number of particles, has been initiated [10]. Examples are provided by reaction-diffusion systems used to model a wide range of phenomena in chemistry and physics [11], and metapopulation epidemic models [2,3,12–16] where particles represent people moving across different subpopulations (nodes) such as city or urban areas, and the reaction processes account for the local infection dynamics.

Here we analyze epidemic metapopulation models characterized by an infection dynamics within each node (or subpopulation) that follows a Susceptible-Infected-Removed (SIR) model. The mobility rate p of individuals defines the coupling process among the subpopulations. In the real world, the networks representing the mobility pattern of individuals among different subpopulations are in many cases highly heterogeneous [17–21]. For this reason, the connectivity pattern of the metapopulation network is described as a random graph with arbitrary degree distribution $P(k)$. By using a mechanistic approach it is possible to show that along with the usual epidemic threshold condition $R_0 > 1$ on the basic reproductive number, the system exhibits a global invasion threshold setting the condition for the infection of a macroscopic fraction of the metapopulation system [22,23]. The threshold condition on R_0 ensures the local outbreak at the subpopulation

level [1,10], whereas the explicit expression obtained for global invasion threshold $R_* > 1$ provides a critical value for the diffusion rate p, below which the epidemic cannot propagate to a relevant fraction of subpopulations. We find that the global invasion threshold is affected by the topological fluctuations of the underlying network. The larger the network heterogeneity, the smaller the value of the critical diffusion rate above which the epidemic may globally invade the metapopulation system. The present results can be generalized to more realistic diffusion and mobility schemes and provide a framework for the analysis of realistic metapopulation epidemic models [24–28].

A simplified mechanistic (i.e., microscopic in the epidemic terminology) approach to the metapopulation spread of infectious diseases uses a markovian assumption in which at each time step the movement of individuals is given in terms of a matrix d_{ij} that expresses the probability that an individual in the subpopulation i is traveling to the subpopulation j. Several modeling approaches to the large-scale spread of infectious diseases [24–27,29] use this mobility process based on transportation networks combined with the local evolution of the disease. The markovian character lies in the assumption that at each time step the same traveling probability applies to all individuals in the subpopulation without having memory of their origin. This mobility scheme coupled with an infection dynamics at the local level can be generally viewed as equivalent to classic reaction-diffusion processes with no constraint on the occupation numbers N_i of each subpopulation. The total population of the metapopulation system is $N = \sum_i N_i$, and each individual diffuses along the edges with a diffusion coefficient d_{ij} that depends on the node degree, subpopulation size, and/or the mobility matrix. The metapopulation system is therefore composed of a network substrate connecting nodes—each corresponding to a subpopulation—over which individuals diffuse. We consider that each node i is connected to other k_i nodes according to

its degree, resulting in a network with degree distribution $P(k)$ and distribution moments $\langle k^\alpha \rangle = \sum_k k^\alpha P(k)$.

In the following, as a simplified diffusion process we assume that the mobility is equivalent to a diffusion rate along any given link of a node with degree k simply equal to $d_{kk'} = p/k$. This is obviously not the case in a wide range of real systems where the extreme heterogeneity of traffic is well documented, and more realistic processes will be considered elsewhere. This simple process, however, automatically generates a stationary distribution of occupation numbers that is better described by grouping subpopulations according to their degree k:

$$N_k = \frac{k}{\langle k \rangle} \bar{N}, \qquad (1)$$

where $\bar{N}$ is the average subpopulation size.

In each subpopulation j the disease follows an SIR model, and the total number of individuals is partitioned in the compartments $S_j(t)$, $I_j(t)$, and $R_j(t)$, denoting the number of susceptible, infected, and removed individuals at time t, respectively. The infection dynamics proceeds as follows. Each susceptible individual has a transition rate to the infected state expressed as $\beta I_j / N_j$, where β is the disease transmissibility rate and I_j / N_j is the force of infection in the homogeneous mixing assumption. Analogously, each infected individual enters the removed compartment according to the recovery rate μ. The basic SIR rules thus define a reaction scheme of the type $S + I \rightarrow 2I$ and $I \rightarrow R$, which conserves the number of individuals. The SIR epidemic model is characterized by the reproductive number $R_0 = \beta / \mu$, which defines the average number of infectious individuals generated by one infected individual in a fully susceptible population. The epidemic is able to generate a number of infected individuals larger than those who recover only if $R_0 > 1$, yielding the classic result for the epidemic threshold [1]. If the spreading rate is not large enough to allow a reproductive number larger than 1 (i.e., $\beta > \mu$), the epidemic outbreak will quickly die out. This result is valid at the level of each subpopulation and holds also at the metapopulation level where $R_0 > 1$ is a necessary condition to have the growth of the epidemic [10].

The intuitive result on the subpopulation epidemic threshold, however, does not take into account the effects due to the finite size of subpopulations, the discrete nature of individuals, and the stochastic nature of the reaction and diffusion processes. These effects have been shown to have a crucial role in the problem of resurgent epidemics, extinction, and eradication [22,23,30,31]. Also, in the present framework indeed each subpopulation may or may not transmit the infection to a neighboring subpopulation upon the condition that at least one infected individual is moving onto the noninfected subpopulations during the epidemic outbreak. Given an SIR model with $R_0 > 1$, the total number of infected individuals generated within a subpopulation and the mobility rate must be large enough

to ensure the seeding of other subpopulations before the end of the local outbreak [22,23].

As a simple example of this effect let us consider a metapopulation system in which the initial condition is provided by a single infection in a subpopulation with degree k and N_k individuals, given $R_0 > 1$. In the case of a macroscopic outbreak in a closed population, the total number of infected individuals during the outbreak evolution will be equal to αN_k, where α depends on the specific disease model and parameter values used. Each infected individual stays in the infectious state for an average time μ^{-1} equal to the inverse of the recovery rate, during which it can travel to the neighboring subpopulation of degree k' with rate $d_{kk'}$. We can therefore consider that on average the number of new seeds that may appear in a connected subpopulation of degree k' during the duration of the local outbreak is given by

$$\lambda_{kk'} = d_{kk'} \frac{\alpha N_k}{\mu}. \qquad (2)$$

In this perspective we can provide a characterization of the invasion dynamics at the level of the subpopulations, translating epidemiological and demographic parameters into Levins-type metapopulation parameters of extinction and invasion rate. Let us define D_k^0 as the number of diseased subpopulation of degree k at generation 0, i.e., those which are experiencing an outbreak at the beginning of the process. Each infected subpopulation will seed—during the course of the outbreak—the infection in neighboring subpopulations defining the set D_k^1 of infected subpopulations at generation 1, and so on. This corresponds to a basic branching process [22,32,33] where the nth generation of infected subpopulations of degree k is denoted D_k^n.

In order to describe the early stage of the subpopulations' invasion dynamics we assume that the number of subpopulations affected by an outbreak (with $R_0 > 1$) is small, and we can therefore study the evolution of the number of diseased subpopulations by using a branching process approximation relating D_k^n with D_k^{n-1}. Let us consider a metapopulation network with degree distribution $P(k)$ and V subpopulations and write the number of subpopulations of degree k invaded at the generation n as

$$D_k^n = \sum_{k'} D_{k'}^{n-1}(k'-1)\left[1 - \left(\frac{1}{R_0}\right)^{\lambda_{k'k}}\right] P(k|k')\left(1 - \frac{D_k^{n-1}}{V_k}\right). \qquad (3)$$

This equation assumes that each infected subpopulation of degree k' of the $(n-1)$th generation, $D_{k'}^{n-1}$, will seed the infection in a number $(k'-1)$ of subpopulations corresponding to the number of neighboring subpopulations k' minus the one that originally transmitted the infection, the probability $P(k|k')$ that each of the $k'-1$ not yet infected neighboring subpopulations has degree k, and the probability to observe an outbreak in the seeded subpopulation, i.e., $(1 - R_0^{-\lambda_{kk}})$ [34]. The last factor stems from the proba-

bility of extinction $P_{\text{ext}} = 1/R_0$ of an epidemic seeded
with a single infectious individual [34]. In order to obtain
an explicit result we will consider in the following that
$R_0 - 1 \ll 1$, thus assuming that the system is found to be
very close to the epidemic threshold. In this limit we can
approximate the outbreak probability as $1 - R_0^{-\lambda_{k'k}} \simeq$
$\lambda_{k'k}(R_0 - 1)$. The case of homogeneous diffusion of indi-
viduals $d_k = p/k$ with the stationary solution of Eq. (2) for
the subpopulation size yields $\lambda_{k'k} = p\bar{N}\alpha\mu^{-1}/\langle k \rangle$. In ad-
dition, we assume that at the early stage of the epidemic
$D_k^{n-1}/V_k \ll 1$, and we consider the case of uncorrelated
networks in which the conditional probability does not
depend on the originating node, i.e., $P(k|k') = kP(k)/\langle k \rangle$
[5], obtaining

$$D_k^n = (R_0 - 1)\frac{kP(k)}{\langle k \rangle^2}\frac{p\bar{N}\alpha}{\mu}\sum_{k'}D_{k'}^{n-1}(k' - 1). \quad (4)$$

By defining $\Theta^n = \sum_{k'}D_{k'}^n(k' - 1)$, the last expression can
be conveniently written in the iterative form

$$\Theta^n = (R_0 - 1)\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle^2}\frac{p\bar{N}\alpha}{\mu}\Theta^{n-1}, \quad (5)$$

which allows the increasing of infected subpopulations and
a global epidemic in the metapopulation process only if

$$R_* = (R_0 - 1)\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle^2}\frac{p\bar{N}\alpha}{\mu} > 1, \quad (6)$$

defining the global invasion threshold of the metapopula-
tion system. In other words, R_* is the analogue of the basic
reproductive number at the subpopulations level and is a
crucial indicator in assessing the behavior of epidemics in
metapopulation models. Its expression indeed contains the
probability of generating an outbreak in a neighbor sub-
population by means of mobility processes, $(R_0 - 1)p\bar{N}\alpha/(\mu\langle k \rangle)$ for $R_0 - 1 \ll 1$, times the factor
$\langle k^2 \rangle/\langle k \rangle - 1$, which also appears in the threshold condi-
tions characterizing phase transitions on complex networks
[7,33,35,36]. The explicit form of Eq. (6) can be used to
find the minimum mobility rate, ensuring that on average
each subpopulation can seed more than one neighboring
subpopulation. The constant α is larger than zero for any
$R_0 > 1$, and in the SIR case for R_0 close to 1 it can be
approximated by $\alpha \simeq 2(R_0 - 1)/R_0^2$ [34], yielding for the
SIR model a critical mobility value p_c, below which the
epidemics cannot invade the metapopulation system given
by the equation

$$p_c\bar{N} = \frac{\langle k \rangle^2}{\langle k^2 \rangle - \langle k \rangle}\frac{\mu R_0^2}{2(R_0 - 1)^2}. \quad (7)$$

The above condition readily tells us that the closer to the
epidemic threshold is the single subpopulation outbreak,
the larger the mobility rate has to be in order to sustain the
global spread into the metapopulation model. It is impor-
tant to stress that when R_0 increases, the small $(R_0 - 1)$
expansions are no longer valid and the invasion threshold is

obtained only in the form of a complicated implicit
expression.

In addition, Eq. (7) shows the dependence of the critical
mobility on the topological fluctuations of the mobility
network. The ratio $\langle k \rangle^2/\langle k(k - 1) \rangle$ is extremely small in
heavy-tailed networks, and it is vanishing in the limit of
infinite network size. This implies that the heterogeneity of
the metapopulation network is favoring the global spread
of epidemics by lowering the global spreading threshold. In
other words, the topological fluctuations favor the subpo-
pulation invasion and suppress the phase transition in the
infinite size limit. This finding provides a theoretical
framework and rationale for the evidence concerning the
inefficacy of travel restrictions in the containment of global
epidemics [27,37]. The simple plug-in of the actual num-
bers for modern transportation networks, the population
sizes, and realistic disease parameters in Eq. (7) indicates
that a reduction of 1 order of magnitude of the mobility is
not enough to bring the system below the invasion thresh-
old. While more complicated mobility schemes should be
considered for a precise calculation, this result sets the
framework for the understanding of mobility effects in
the spread and containment of infectious diseases.

In order to support the previous analytical finding we
have performed an extensive set of Monte Carlo numerical
simulations of the metapopulation system. The substrate
network is given by an uncorrelated complex network with
$P(k) \sim k^{-2.1}$ generated with the uncorrelated configuration
model [38] to avoid inherent structural correlations.
Network sizes of $V = 10^4$ and 10^5 nodes have been con-
sidered. The dynamics proceeds in parallel and considers
discrete time steps representing the unitary time scale τ of
the process. The reaction and diffusion rates are therefore
converted into probabilities, and at each time step in each
subpopulation j a susceptible individual is turned into an
infectious individual with probability $1 - (1 - \frac{\beta}{N_j}\tau)^{I_j}$ and
each infectious individual is subject to the recovery process
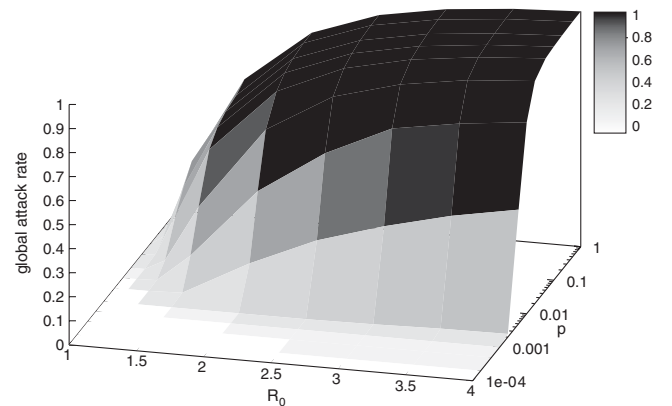and recovers with probability $\mu\tau$. The mobility is modeled

FIG. 1. Phase diagram of the metapopulation system. The final
epidemic size is shown as a function of the local threshold R_0
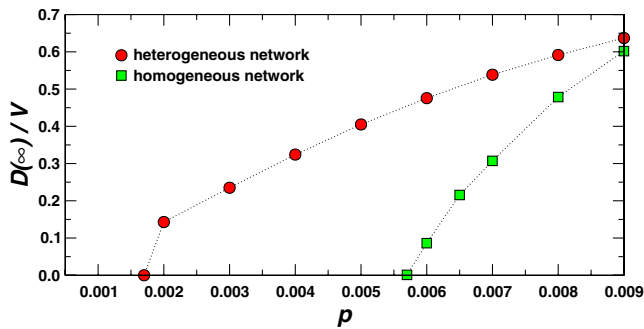and of the diffusion probability p.

FIG. 2 (color online). Effect of the network heterogeneity on the global epidemic threshold. The final fraction of diseased subpopulations $D(\infty)/V$ at the end of the global epidemic is shown as a function of the mobility rate p in a homogeneous network and a heterogeneous network.

assuming a diffusion probability for each individual along each link of the subpopulation of the form $d_{kk'} = p/k$.

A complete analysis of the system phase diagram is obtained by analyzing the behavior of the global attack rate $R(\infty)/N$, defined as the total fraction of cases in the metapopulation system at the end of the epidemic, as a function of both R_0 and p. Figure 1 reports the global attack rate surface in the p-R_0 space and clearly shows the effect of different couplings as expressed by the value of p in reducing the final size of the epidemic at a given fixed value of R_0. The smaller the value of R_0, the higher the coupling needs to be in order for the virus to successfully invade a finite fraction of the subpopulations, in agreement with the analytic result of Eq. (7). This provides a clear illustration of the varying global invasion threshold as a function of the reproductive rate R_0. Furthermore, it is possible to study the effect of the heterogeneity of the metapopulation structure on the global epidemic threshold. Figure 2 shows the results obtained by comparing two random metapopulation networks, one with Poissonian degree distribution (homogeneous network) and one with heavy-tailed $[P(k) \sim k^{-2.1}]$ degree distribution (heterogeneous network). Despite the two models having the same average degree and disease parameters, the fluctuations of the power-law network increase the value of R_*, thus lowering the critical value of the mobility.

The present analysis provides insights in setting a framework for the analysis of large-scale spread of epidemics in realistic mobility networks. Furthermore, these results open the path to future work aimed at analyzing refined metapopulation infection models.

[1] R. M. Anderson and R. M. May, *Infectious Diseases in Humans* (Oxford University Press, Oxford, 1992).
[2] H. W. Hethcote, Theor. Popul. Biol. **14**, 338 (1978).
[3] R. M. May and R. M. Anderson, Math. Biosci. **72**, 83 (1984).
[4] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[5] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks* (Oxford University Press, Oxford, 2003).
[6] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[7] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[8] A. L. Lloyd and R. M. May, Science **292**, 1316 (2001).
[9] M. Barthélemy *et al.*, J. Theor. Biol. **235**, 275 (2005).
[10] V. Colizza, R. Pastor-Satorras, and A. Vespignani, Nature Phys. **3**, 276 (2007).
[11] N. G. van Kampen, *Stochastic Processes in Chemistry and Physics* (Springer, Berlin, 2005).
[12] M. J. Keeling and P. Rohani, Ecology Letters **5**, 20 (2002).
[13] A. L. Lloyd and R. M. May, J. Theor. Biol. **179**, 1 (1996).
[14] B. T. Grenfell and J. Harwood, Trends Ecol. Evol. **12**, 395 (1997).
[15] B. T. Grenfell and B. M. Bolker, Ecology Letters **1**, 63 (1998).
[16] N. M. Ferguson *et al.*, Nature (London) **425**, 681 (2003).
[17] G. Chowell *et al.*, Phys. Rev. E **68**, 066102 (2003).
[18] C. L. Barrett *et al.*, Los Alamos National Laboratory Technical Report No. LA-UR-00-1725, 2000.
[19] A. Barrat *et al.*, Proc. Natl. Acad. Sci. U.S.A. **101**, 3747 (2004).
[20] R. Guimerá *et al.*, Proc. Natl. Acad. Sci. U.S.A. **102**, 7794 (2005).
[21] A. De Montis *et al.*, Environment and Planning B **35**, 905 (2007).
[22] F. Ball, D. Mollison, and G. Scalia-Tomba, Ann. Appl. Probab. **7**, 46 (1997).
[23] P. Cross *et al.*, Ecology Letters **8**, 587 (2005).
[24] R. F. Grais, J. H. Ellis, and G. E. Glass, Eur. J. Epidemiol. **18**, 1065 (2003).
[25] L. Hufnagel, D. Brockmann, and T. Geisel, Proc. Natl. Acad. Sci. U.S.A. **101**, 15 124 (2004).
[26] V. Colizza *et al.*, Proc. Natl. Acad. Sci. U.S.A. **103**, 2015 (2006).
[27] V. Colizza *et al.*, PLoS Medicine **4**, e13 (2007).
[28] S. Riley, Science **316**, 1298 (2007).
[29] L. A. Rvachev and I. M. Longini, Math. Biosci. **75**, 3 (1985).
[30] D. Watts *et al.*, Proc. Natl. Acad. Sci. U.S.A. **102**, 11 157 (2005).
[31] A. Vázquez, J. Theor. Biol. **245**, 125 (2007).
[32] T. E. Harris, *The Theory of Branching Processes* (Dover Publications, New York, 1990).
[33] A. Vázquez, Phys. Rev. Lett. **96**, 038702 (2006).
[34] N. T. Bailey, *The Mathematical Theory of Infectious Diseases* (Macmillan, New York, 1975), 2nd ed.
[35] D. S. Callaway *et al.*, Phys. Rev. Lett. **85**, 5468 (2000).
[36] R. Cohen *et al.*, Phys. Rev. Lett. **85**, 4626 (2000).
[37] T. D. Hollingsworth, N. M. Ferguson, and R. M. Anderson, Nature Medicine **12**, 497 (2006).
[38] M. Catanzaro, M. Boguña, and R. Pastor-Satorras, Phys. Rev. E **71**, 027103 (2005).