Supplementary Information
Phase transitions in contagion processes mediated by recurrent mobility patterns

Duygu Balcan[1,2] and Alessandro Vespignani[1,2,3⋆]

December 12, 2010

[1]Center for Complex Networks and Systems Research (CNetS), School of Informatics and Computing, Indiana University, Bloomington, IN 47408, USA

[2]Pervasive Technology Institute, Indiana University, Bloomington, IN 47406, USA

[3]Institute for Scientific Interchange (ISI), Torino, Italy

⋆email: alexv@indiana.edu

Invasion Threshold

The global behavior of the contagion process is determined by the largest eigenvalue R_* of the subpopulation next generation matrix G as detailed in the Methods section of main paper. If the eigenvalue $R_* > 1$ we have that the subpopulation invasion process is supercritical and the disease will be able to globally spread across subpopulations. This is equivalent to define a subpopulation reproductive number R_*[1–5] that in structured metapopulation systems is equivalent to basic reproductive number R_0 at the single population level:

$$R_* = \frac{2\overline{N}(R_0-1)^2\rho}{R_0^2(1+\langle k\rangle/\langle k^2\rangle+\rho)} F(\langle k\rangle, \langle k^2\rangle, \langle k^3\rangle, \langle k^4\rangle) \quad , \tag{1}$$

where

$$F(\langle k\rangle,\langle k^2\rangle,\langle k^3\rangle,\langle k^4\rangle) \equiv \frac{1}{\langle k\rangle\langle k^2\rangle}\left[\langle k^3\rangle - \langle k^2\rangle + \left(\langle k^4\rangle - \langle k^3\rangle\right)^{1/2}\left(\langle k^2\rangle - \langle k\rangle\right)^{1/2}\right] \quad . \quad (2)$$

The infectious diseases will spread globally in the metapopulation system only if $R_* > 1$. Thus, by setting $R_* = 1$, we can define an epidemic threshold relation for the mobility ratio ρ,

$$\rho_c = \frac{1 + \langle k\rangle/\langle k^2\rangle}{2\overline{N}(1 - R_0^{-1})^2 F(\langle k\rangle,\langle k^2\rangle,\langle k^3\rangle,\langle k^4\rangle) - 1} \quad , \quad (3)$$

below which the infection remains confined to a small number of subpopulations. In an infinite metapopulation system the threshold is defined rigorously and the fraction of infected subpopulations is zero below the threshold and finite only if the mobility parameters set the ratio ρ above the threshold value. The threshold value is defined for the ratio between the rates characterizing the mobility process. This condition is therefore twofold on the mobility dynamics if we fix one of the two parameters σ and τ, and let the other parameter free. On one hand the threshold relation is $\sigma_c = \rho_c\, \tau^{-1}$,

$$\sigma_c = \frac{(1 + \langle k\rangle/\langle k^2\rangle)\tau^{-1}}{2\overline{N}(1 - R_0^{-1})^2 F(\langle k\rangle,\langle k^2\rangle,\langle k^3\rangle,\langle k^4\rangle) - 1} \quad . \quad (4)$$

This intuitively states that the rates of diffusion has to be large enough ($\sigma > \sigma_c$) to guarantee the spreading of the disease. Interestingly, however, we can also define the threshold relation for τ by $\tau_c = \rho_c\sigma^{-1}$,

$$\tau_c = \frac{(1 + \langle k\rangle/\langle k^2\rangle)\sigma^{-1}}{2\overline{N}(1 - R_0^{-1})^2 F(\langle k\rangle,\langle k^2\rangle,\langle k^3\rangle,\langle k^4\rangle) - 1} \quad , \quad (5)$$

that is telling that the global spreading of the contagion porcess can be achieved by reducing the return rates of individuals; in other words by extending the visit times of individuals in nearby subpopulations ($\tau > \tau_c$). This last conditions however breaks down when τ becomes much larger than the contagion time scale thus breaking the time-scale separation[6] assumption used here.

Another very interesting feature of the above threshold condition is the explicit effect of the network topology encoded in the moments of degree distribution $\langle k\rangle$, $\langle k^2\rangle$, etc. As already been

observed in the case of Markovian diffusion case[4,5], the heterogeneity of the network favors the global spread of the epidemic by lowering the threshold value. Indeed, for heavy-tailed degree distribution $P(k) \sim k^{-\gamma}$ with $\gamma > 1$, the nth moment scales as $k_{\max}^{n+1-\gamma}$ if $n \geq \gamma - 1$ and $k_{\max} \gg k_{\min}$. This means that for $n \geq \gamma - 1$, the nth moment of the degree distribution tends to diverge in the infinite size limit of the network, as in this limit $k_{\max} \to \infty$, virtually reducing the threshold to zero. Even at finite size, however, the threshold value is generally smaller the higher the network heterogeneity is. In order to make this last statement transparent, we will turn our attention to the scaling of the moments of degree distribution for very large system sizes. In the case of $1 < \gamma < 2$, the term $\langle k \rangle / \langle k^2 \rangle$ in the nominator of ρ_c scales as

$$\frac{\langle k \rangle}{\langle k^2 \rangle} \sim k_{\max}^{-1} \quad . \tag{6}$$

In the range $2 < \gamma < 3$, the scaling is

$$\frac{\langle k \rangle}{\langle k^2 \rangle} \sim k_{\max}^{\gamma-3} \quad . \tag{7}$$

In all the other cases of $\gamma > 3$, the term $\langle k \rangle / \langle k^2 \rangle$ has a finite value. That means that the nominator of ρ_c is finite for any $\gamma > 1$. Now lets turn our attention to the denominator of ρ_c and analyze the scaling of $F(\langle k \rangle, \langle k^2 \rangle, \langle k^3 \rangle, \langle k^4 \rangle)$. In the case of $1 < \gamma < 2$, F scales as

$$F(\langle k \rangle, \langle k^2 \rangle, \langle k^3 \rangle, \langle k^4 \rangle) \sim \frac{\langle k^3 \rangle + \langle k^4 \rangle^{1/2} \langle k^2 \rangle^{1/2}}{\langle k \rangle \langle k^2 \rangle} \sim k_{\max}^{\gamma-1} \quad . \tag{8}$$

In the case of $2 < \gamma < 3$, second moment $\langle k^2 \rangle$ in the denominator and higher moments in the numerator are dominant, leading to the scaling relation:

$$F(\langle k \rangle, \langle k^2 \rangle, \langle k^3 \rangle, \langle k^4 \rangle) \sim \frac{\langle k^3 \rangle + \langle k^4 \rangle^{1/2} \langle k^2 \rangle^{1/2}}{\langle k^2 \rangle} \sim k_{\max} \quad . \tag{9}$$

In the range $3 < \gamma < 4$, only the third moment $\langle k^3 \rangle$ in the numerator dominates, thus

$$F(\langle k \rangle, \langle k^2 \rangle, \langle k^3 \rangle, \langle k^4 \rangle) \sim \langle k^3 \rangle \sim k_{\max}^{4-\gamma} \quad . \tag{10}$$

In the range $4 < \gamma < 5$, only the fourth moment $\langle k^4 \rangle$ in the numerator dominates, leading to

$$F(\langle k \rangle, \langle k^2 \rangle, \langle k^3 \rangle, \langle k^4 \rangle) \sim \langle k^4 \rangle^{1/2} \sim k_{\max}^{(5-\gamma)/2} \quad . \tag{11}$$

The above expressions state that for any heavy-tailed degree distribution with exponent $\gamma < 5$, $F(\langle k\rangle, \langle k^2\rangle, \langle k^3\rangle, \langle k^4\rangle)$ tents to diverge in the limit of infinite network size, which in turn pushes the threshold value ρ_c to zero. While, on the other hand, if $\gamma > 5$ then $F(\langle k\rangle, \langle k^2\rangle, \langle k^3\rangle, \langle k^4\rangle)$ has a finite value.

Computational model

Inside each subpopulation we consider an SIR epidemic model[7], in which each individual is classified by one of the discrete disease states at any point in time. The rate at which a susceptible person in subpopulation i acquires the infection, the so-called force of infection[7] λ_i, is determined by interactions with infectious individuals. The force of infection λ_i acting on each susceptible individual in subpopulation i has been assumed to follow the mass action principle

$$\lambda_i(t) = \beta \frac{I_i^*(t)}{N_i^*(t)} \quad , \tag{12}$$

where β is the transmission rate of infection and $I_i^*(t)/N_i^*(t)$ is the prevalence of infectious individuals in the subpopulation. Each person in the susceptible compartment (S) contracts the infection with probability $\lambda_i(t)\Delta t$ and enters the infectious compartment (I), where Δt is the time interval considered. Each infectious individual permanently recovers with probability $\mu\Delta t$, entering the recovered compartment (R).

Synthetic subpopulation networks

Generation of substrate networks. In order to compare with theoretical calculations, topologically uncorrelated random graphs have been considered. In this case, analytical calculations show that epidemic invasion threshold only depends on the degree distribution of the subpopulation networks. In order to verify this result, two different network topologies have been generated:

- Erdős-Rényi graphs[8] have been synthetized by assigning a link between each pair of

nodes with probability $\langle k\rangle/(V-1)$, where V is the number of nodes and $\langle k\rangle$ is a prescribed average node degree.

- Networks with power-law degree distribution, $P(k) \sim k^{-\gamma}$ with $k_{\min} \leq k \leq k_{\max}$, have been generated by uncorrelated configuration model[9,10]. All the scale-free networks have been generated by setting $\gamma = 2.1$ and $k_{\min} = 2$.

For the sake of comparison, the average degree of Erdős-Rényi graphs has been set to that of scale-free networks.

Subpopulation sizes. From a pool of $\overline{N}V$ people, a population size N_i is assigned to each subpopulation i, defining its permanent residents. The population size is chosen at random from a multinomial distribution with probability proportional to k_i, which ensures the metapopulation system to obey $N_k = \overline{N}k/\langle k\rangle$.

Mobility parameters. The rate σ_{ij} at which a resident of subpopulation i commutes to a neighboring subpopulation $j \in \upsilon(i)$ assumes

$$\sigma_{ij} = \sigma \frac{N_j}{N_i + \sum_{\ell\in\upsilon(i)} N_\ell} \quad . \tag{13}$$

Each resident in subpopulation i leaves its origin and visits subpopulation j with probability $\sigma_{ij}\Delta t$. A commuter in subpopulation j returns back to its resident subpopulation i with probability $\tau^{-1}\Delta t$.

Simulations have been initialized with $I(0) = 10$ infectious individuals, seeded randomly in a single subpopulation of degree $k_{\min}$, while the rest of the population is assumed to be susceptible to infection.

Real-world subpopulation networks

Realistic simulations have been performed using the county to country commuting network in the continental United States[11]. The network consists of about $3,100$ nodes, each of which corresponds to a US county. Weighted link from node i to node j represents the daily number of

commuters from county i to county j. Thus, the population size of each node and the commuting rates among them are fully determined by the data. The return rate τ^{-1}, however, has been set to $\tau^{-1} = 3\ day^{-1}$ corresponding to a regular working day (8 hours). Simulations have been initialized with $I(0) = 10$ infectious individuals seeded in Los Angeles County, California.

Statistical analysis

Since we aim at determining the epidemic invasion threshold, we have let the metapopulation system run until the infection dies out. In the numerical results presented in main paper, all the realizations resulting in at least one diseased subpopulation have contributed to the statistical analysis. For each set of parameters, we have generated at least $1,000$ system realizations. Since the subpopulation networks and dynamical processes on them are subject to fluctuations in the case of synthetic populations, we have sampled at least $10-20$ network realizations and $100-200$ dynamical realizations on each of them. While in the case of the real-world scenarios, we have generated at least $1,000$ dynamical realizations.

Contagion and mobility dynamics

We will follow the notations defined in main paper and represent each individual by its disease state X, its permanent subpopulation i and its present subpopulation $j \in \upsilon(i)$. Since all the individuals sharing the same three indices $(X,\ i,\ j)$ are identical in terms of the dynamical processes, we are going refer to the number of such individuals at time t by $X_{ij}(t)$. Then, by definition, the instantaneous compartment size $X_j^*(t)$ in subpopulation j can be expressed as

$$X_j^*(t) \quad = \quad X_{jj}(t) + \sum_{\ell \in \upsilon(j)} X_{\ell j}(t) \quad , \tag{14}$$

and the total number of individuals as $N_j^* = \sum_X X_j^*$. The number of individuals in each compartment X with a residence in i and present in j is subject to discrete and stochastic dynamical processes defined by disease and transport operators. The disease operator $\mathcal{D}_j$ represents the change due to the compartment transition induced by the infection dynamics, and the transport

operator Ω_X represents the variation due to mobility.

The term $\mathcal{D}_j$ can be written as a combination of a set of transitions $\{\mathcal{D}_j(X,Y)\}$, where $\mathcal{D}_j(X,Y)$ represents the number of transitions from compartment X to Y and is simulated as an integer random number extracted from a multinomial distribution. Then the change due to infection dynamics reads as

$$\mathcal{D}_j(X) = \sum_Y \left[\mathcal{D}_j(Y,X) - \mathcal{D}_j(X,Y)\right] \quad . \tag{15}$$

As a concrete example let us consider the temporal change in the infectious compartment. There is only one possible transition from the compartment, that is to the recovered compartment. The number of transitions is extracted from the binomial distribution

$$\Pr^{\mathrm{Binom}}(I_{ij}(t), p_{I_{ij}\to R_{ij}}) \quad , \tag{16}$$

determined by the transition probability

$$p_{I_{ij}\to R_{ij}} \quad = \quad \mu\Delta t \quad , \tag{17}$$

and the number of individuals in the compartment $I_{ij}(t)$ (its size). This transition causes a reduction in the size of the compartment. The increase in the compartment size is due to the transitions from susceptible into infectious compartment. This is also a random number extracted from the binomial distribution

$$\Pr^{\mathrm{Binom}}(S_{ij}(t), p_{S_{ij}\to I_{ij}}) \quad , \tag{18}$$

given by the chance of contagion

$$p_{S_{ij}\to I_{ij}} \quad = \quad \lambda_j(t)\Delta t \quad , \tag{19}$$

and the number of attempts equal to the number of susceptibles $S_{ij}(t)$. After extracting these numbers from the appropriate distributions, we can calculate the total change $\mathcal{D}_j(I)$ in infectious compartment as

$$\mathcal{D}_j(I) = \mathcal{D}_j(S,I) - \mathcal{D}_j(I,R) \quad . \tag{20}$$

Transport operator Ω_X expresses the total change in compartment sizes due to the commuting of permanent residents of subpopulation i back and forth. The variation in X_{ij} can be decomposed into $\Omega_X^{\rightarrow}(i,j)$ and $\Omega_X^{\leftarrow}(j,i)$ as

$$\Omega_X = \Omega_X^{\rightarrow}(i,j) - \Omega_X^{\leftarrow}(j,i) \quad . \tag{21}$$

The first term $\Omega_X^{\rightarrow}(i,j)$ represents an increase that is caused by the departing residents of subpopulation i to visit subpopulation j. The $\Omega_X^{\rightarrow}(i,j)$ is a random number extracted from the multinomial distribution

$$\Pr^{\text{Multinom}}(X_{ii}(t), \{p_{X_{ii}\rightarrow X_{i\ell}} | \ell \in \upsilon(i)\}) \quad , \tag{22}$$

determined by the probability of commuting to subpopulation j

$$p_{X_{ii}\rightarrow X_{ij}} = \sigma_{ij}\Delta t \quad , \tag{23}$$

and the number of such trails $X_{ii}(t)$. The second term $\Omega_X^{\leftarrow}(j,i)$ corresponds to a reduction in X_{ij} and is due to the return trips from subpopulation j to permanent subpopulation i. The $\Omega_X^{\leftarrow}(j,i)$ is also a random number extracted from the binomial distribution

$$\Pr^{\text{Binom}}(X_{ij}(t), p_{X_{ij}\rightarrow X_{ii}}) \quad , \tag{24}$$

given by the probability of returning home

$$p_{X_{ij}\rightarrow X_{ii}} = \tau^{-1}\Delta t \quad , \tag{25}$$

and the size of the compartment $X_{ij}(t)$.

We have assumed that the infection does not alter people’s behavior, i.e., all the compartments are identical in their mobility.

References

1. Ball, F., Mollison, D. & Scalia-Tomba, G. Epidemics with two levels of mixing. *Ann. Appl. Probab.* **7,** 46−89 (1997).

2. Cross, P., Lloyd-Smith, J. O., Johnson, P. L. F. & Wayne, M. G. Duelling timescales of host movement and disease recovery determine invasion of disease in structured populations. *Ecol. Lett.* **8,** 587−595 (2005).

3. Cross, P., Johnson, P. L. F., Lloyd-Smith, J. O. & Wayne, M. G. Utility of R_0 as a predictor of disease invasion in structured populations. *J. R. Soc. Interface* **4,** 315−324 (2007).

4. Colizza, V. & Vespignani, A. Invasion threshold in heterogeneous metapopulation networks. *Phys. Rev. Lett.* **99,** 148701 (2007).

5. Colizza, V. & Vespignani, A. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. *J. Theor. Biol.* **251,** 450−467 (2008).

6. Keeling, M. J. & Rohani, P. Estimating spatial coupling in epidemiological systems: a mechanistic approach. *Ecol. Lett.* **5,** 20−29 (2002).

7. Keeling, M. J. & Rohani, P. *Modeling Infectious Diseases in Humans and Animals* (Princeton Univ. Press, Princeton, 2008).

8. Erdős, P. & Rényi, A. On random graphs. *Publ. Math.* **6,** 290−297 (1959).

9. Molloy, M. & Reed, B. The size of the largest component of a random graph on a fixed degree sequence. *Combinatorics, Probab. Comput.* **7,** 295−306 (1998).

10. Catanzaro, M., Bogunã, M. & Pastor-Satorras, R. Generation of uncorrelated random scale-free networks. *Phys. Rev. E* **71,** 027103 (2005).

11. U.S. Census Bureau, http://www.census.gov/.