Supplementary Information
Heterogeneous length of stay of hosts' movements and spatial epidemic spread

C. Poletto[1], M. Tizzoni[1,2], V. Colizza[3,4,5]

1 Empirical data on length of stay

Length of stay is a key factor in the management and development of the tourism industry and represents an important issue of debate for the economic literature [1, 2, 3]. Due to its importance for the economy, data on the average length of stay of travellers are collected by public and private agencies in several countries of the world, at different geographical resolutions.

When considering official tourism statistics, it is important to point out that the statistical definition of tourism is broader than the common, everyday definition, because it encompasses not only private travel but also business travel. This is the case, for instance, of the definition of tourism given by Eurostat [4]. Similarly, tourism surveys of several countries cover a wide range of travel purposes, including holiday, but not limited to it: people travelling to visit relatives and friends, or for educational purposes, are also included in the statistics. This is primarily because the agencies view tourism from an economic perspective.

Many of the statistical datasets are publicly available and can be downloaded from the internet. They usually provide the average length of stay of travellers in a defined country, along with the travellers' country of origin, purpose of visit and local destination. Even if data do not provide information at the individual level, all the available statistical sources show that the length of stay of travellers around the world is broadly distributed, usually spanning several orders of magnitude in time. Moreover, the length of stay varies between different geographic locations within the same country. In the following subsections we list the main sources of our empirical findings, organized by reporting country.

[1]Computational Epidemiology Laboratory, Institute for Scientific Interchange (ISI), Torino, Italy
[2]Mathematical Modeling Services, BC Centre for Disease Control, 655 West 12th Ave., Vancouver BC V5Z 4R4, Canada
[3]INSERM, U707, Paris, France
[4]UPMC Université Paris 06, Faculté de Médecine Pierre et Marie Curie, UMR S 707, Paris, France
[5]Institute for Scientific Interchange (ISI), Torino, Italy

1.1 European Union data

Europe is a major tourist destination, and six of the member states of the EU are among the world's top ten destinations for holiday-makers. The Institute of Statistics of the European Commission, Eurostat, has been collecting data on the development and structure of tourism since 1995 in all the countries of the European Union. All datasets are available from the Eurostat website (http://epp.eurostat.ec.europa.eu/) and include the average length of stay of tourists along with several other indicators.

Eurostat statistics define as short trips those that lasted less than 4 days, and long trips those that lasted 4 days or more. Business trips tend to be short, but they represent only about 20% of the total trips made in Europe in 2010 (http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Seasonality_in_tourism_demand). Holiday trips, accounting for the remaining 80%, show a broad distribution of the length of stay. Even if less than a half of the holiday trips are long, their duration can extend up to a few months and the observed distribution of the length of stay is very similar for several European countries, as shown in Figure 1d of the main text.

1.2 United Kingdom data

The UK Office for National Statistics issues every year a detailed report of travel statistics titled "Travel trends" and freely available at: http://www.statistics.gov.uk/. Travel Trends presents the main results from the International Passenger Survey (IPS) which collects information on travel to and from the United Kingdom. The survey is based on face-to-face interviews with a random sample of passengers as they enter or leave the UK by the principal air, sea and tunnel routes.

Length of stay of travellers departing from the UK or arriving to the UK is one of the several information collected by the IPS. Figure S1 shows the distribution of the length of stay of travellers, measured as number of nights, reported by the issue of "Travel trends" in the year 2007. All travel purposes are considered and the two panels show different distributions corresponding to two different trip origin or destination: United States and European countries. Figure 1c of the main text reports the same quantity, considering all countries of origin and destination.

In general all distributions indicate a vast heterogeneity of the length of stay, with a considerable fraction of individuals whose sojourn lasted more than 7 days and up to 1 year.

Furthermore, based on the UK Office for National Statistics data it is possible to extract the geographical distribution of the length of stay within the country. Travel Trends reports the average length of stay in the top 50 towns where foreign travellers have spent their sojourn. Figure S2 shows the average length of stay in 22 towns of the UK, for travellers from North America, Europe and other countries. In order to be consistent with our modeling framework, we selected from the original report the towns that can be univocally associated to an airport.

Overall, the length of stay varies of about one order of magnitude from one town to the other, depending on the origin of travellers, indicating a strong geographic dependence of this quantity. This confirms our modelling assumption that assigns different length of stays to different locations. At the same time, the number of towns for which data are available is too small to perform a meaningful statistical analysis that could confirm the assumed dependence of the length of stay with the subpopulation degree (see main text, equation 1).

1.3 Australia data

Tourism Research Australia (TRA) is a branch of the Australian Department of Resources, Energy and Tourism that provides statistics, research and analysis to support the Australian tourism industry. Data published by TRA are available from its website (`http://www.tra.australia.com`) and include the length of stay of travellers who visited Australia for all purposes.

Figure S3 reports the mean and median length of stay of travellers who visited Australia in 2008 from different origin countries and for different purposes: holiday, business, educational and employment.

It is clear that business trips tend to be shorter, usually lasting less than 15 days, but on the other side educational and employment trips are usually very long and can easily exceed the 3 months duration, on average. Holiday trips show large fluctuations in their duration and can last from few days up to 50 days, on average.

2 Epidemic model with non-Markovian mobility

2.1 Non-Markovian mobility model

Hosts' recurrent mobility patterns can be described by defining a non-Markovian travelling process coupling the subpopulations of the metapopulation system [5, 6]. Consider a subpopulation i connected to a set of other subpopulations $\upsilon(i)$. The hosts belonging to i, N_i, are divided in subclasses according to the location where they are present: N_{ii} indicates the individuals of i not travelling and thus present in i, and for each subpopulation j belonging to $\upsilon(i)$, N_{ij} indicates the subclass of individuals of i travelling to j. The mobility process is described by the rate equations governing the dynamics of the classes N_{ii} and N_{ij}. By introducing the parameters σ_{ij} and τ_{ij}, that represent the leaving rate from the origin subpopulation i to the destination j and the average length of stay, respectively, we can write [6] the equations:

$$d_t N_{ii}(t) = -\sigma_i N_{ii}(t) + \sum_{l \in \upsilon(i)} \tau_{il}^{-1} N_{il}(t) \quad \text{(S1)}$$

$$d_t N_{ij}(t) = \sigma_{ij} N_{ii}(t) - \tau_{ij}^{-1} N_{ij}(t). \quad \text{(S2)}$$

The first equation describes the time evolution of the class N_{ii}, which is determined by the balance between the flux of people leaving the subpopulation, regulated by the total leaving rate $\sigma_i = \sum_{j \in \upsilon(i)} \sigma_{ij}$ and the flux of people coming back from all the possible destinations. The second equation describes the behaviour of the class N_{ij} of people travelling to the destination j. The average length of stay τ_{ij} accounts for the time scale characteristic of the trip. We assume that this time scale depends on the destination only, $\tau_{ij} \equiv \tau_j$, and that it is heterogeneously distributed among the subpopulations. The number of individuals present in the subpopulation i, N_i^* and the total traffic volume along the connection ij, T_{ij} can be expressed as a combination of the rate parameters and the occupation numbers N_{ii} and N_{ij}. N_i^* is the sum of people resident in i staying at home and people resident in the subpopulations connected to i and traveling in i, thus it reads

$$N_i^*(t) \quad = \quad N_{ii}(t) + \sum_{l \in \upsilon(i)} N_{li}(t). \tag{S3}$$

T_{ij} is the sum of people departing from i to j, $\sigma_{ij} N_{ii}(t)$, and people coming back from i to j, $\tau_i^{-1} N_{ji}(t)$, therefore its expression is given by

$$T_{ij}(t) \quad = \quad \sigma_{ij} N_{ii}(t) + \tau_i^{-1} N_{ji}(t). \tag{S4}$$

The dynamics described by Eq. (S1) and (S2) conserves the number of individuals resident in each population, $N_i = N_{ii}(t) + \sum_{l \in \upsilon(i)} N_{il}(t)$. All the other quantities are functions of time and evolve till the equilibrium state is asymptotically reached. Under the assumption of $\sigma_i \ll \tau_j^{-1}$ for all i and j the relaxation time is given by $\tau_{max} = \max_j \tau_j$ [5]. At the equilibrium the occupation numbers of the classes N_{ii} and N_{ij} can be computed from Eqs. (S1) and (S2) by placing the derivatives $d_t N_{ij}(t)$ and $d_t N_{ii}(t)$ equal to zero. From Eq. (S2) we obtain $N_{ij}(t) = \tau_{ij}\, \sigma_{ij} N_{ii}(t)$. We plug this expression it into the relation for the total number of individuals of the subpopulation i, $N_i = N_{ii}(t) + \sum_{l \in \upsilon(i)} N_{il}(t)$ and we obtain the expression for N_{ii}. The explicit forms of N_{ii} and N_{ij} then reads:

$$N_{ii} \quad = \quad \frac{N_i}{1 + \sum_{l \in \upsilon(i)} \sigma_{il} \tau_l} \tag{S5}$$

$$N_{ij} \quad = \quad \frac{\sigma_{ij} \tau_j N_i}{1 + \sum_{l \in \upsilon(i)} \sigma_{il} \tau_l}; \tag{S6}$$

The quantities T_{ij} and N_i^* at the equilibrium can then be computed from Eq. (S4) and (S3).

2.2 Degree-block approximation

In the case of a large number of subpopulations and a complex connectivity pattern a network based approach can be employed to provide a coarse grained description of the mobility dynamics and obtain approximate explicit forms for Eqs. (S5) and (S6). In particular a mean-field description of the system is possible by means of the degree-block approximation that allows to properly account for the topological heterogeneities and the empirical scaling relations that relate demographic and mobility aspects to the subpopulation degrees.

The degree-bock approximation assumes that nodes with the same degree are statistically equivalent [7, 8, 9, 10, 11]. The subpopulations are divided in classes according to their degree and are described by average quantities over the degree-block,

$$X_k = \frac{1}{V_k} \sum_{i | k_i = k} X_i, \tag{S7}$$

where V_k is the number of nodes with degree k, X_i the quantity of interest, and the sum runs over all nodes having degree k_i equal to k. The non-Markovian travel dynamics introduced here can be reformulated within the degree-block description by averaging each quantity within the degree class. Mean field differential equations of the mobility process described by Eq. (S1) and (S2) can be written in the degree block notation by accounting for the connectivity pattern encoded in the degree distribution and in the degree correlation function. The resulting equations are described in the Methods section of the main paper.

In the following we provide some further details on the derivation of the equilibrium relations expressed by the Eqs. (2) and (3) of the main paper. We start from the equilibrium condition

$$\sigma_{kk'}N_{kk} - \tau_{k'}^{-1}N_{kk'} = 0, \tag{S8}$$

which yields

$$N_{kk'} = \tau_{k'}\,\sigma_{kk'}N_{kk}. \tag{S9}$$

We consider at this point the relation for the number of individuals resident in a subpopulation of degree k, that in the degree block notation reads

$$N_k \equiv \frac{\bar{N}}{\langle k^{\phi}\rangle}k^{\phi} = N_{kk} + k\sum_{k'}\frac{k'P(k')}{\langle k\rangle}N_{kk'}, \tag{S10}$$

where we are assuming an uncorelated network, i. e. $P(k'|k) = k'P(k')/\langle k\rangle$. We plug Eq. (S9) into Eq. (S10) and we remember that

$$\begin{aligned}\sigma_{kk'} &= \frac{w_{kk'}}{N_k} = \sigma k^{\theta-\phi}k'^{\theta} &\text{(S11)}\\ \tau_{k'} &= \frac{\bar{\tau}}{\langle k^{\chi}\rangle}k'^{\chi}. &\text{(S12)}\end{aligned}$$

We obtain in this way

$$\frac{\bar{N}}{\langle k^{\phi}\rangle}k^{\phi} = N_{kk}\left(1 + \frac{\sigma\bar{\tau}}{\langle k^{\chi}\rangle}\,k^{\theta-\phi+1}\sum_{k'}\frac{P(k')}{\langle k\rangle}k'^{\theta+\chi+1}\right). \tag{S13}$$

The term within round bracket defines the denominator of ν_k. The sum on the left gives the moment of order $\theta+\chi+1$ thus recovering the expression given in the main text

$$\nu_k = \left(1 + \sigma\bar{\tau}\,\frac{\langle k^{\theta+\chi+1}\rangle}{\langle k^{\chi}\rangle\,\langle k\rangle}\,k^{\theta-\phi+1}\right)^{-1} \tag{S14}$$

Eq. (S13) provides us the expression of N_{kk} given in Eq. (2) of the main text, Eq. (3) follows directly by recovering the equilibrium relation (S9)

$$\begin{aligned}N_{kk} &= \frac{\bar{N}}{\langle k^{\phi}\rangle}\nu_k k^{\phi} &\text{(S15)}\\ N_{kk'} &= \frac{\sigma\bar{N}}{\langle k^{\phi}\rangle}\frac{\bar{\tau}}{\langle k^{\chi}\rangle}\nu_k k^{\theta}k'^{\theta+\chi}. &\text{(S16)}\end{aligned}$$

The latter equations allow computing $T_{kk'}$, the total volume of people traveling from a subpopulation of degree k to a subpopulation of degree k' at the equilibrium. Indeed Eq. (S4) for the traffic rewritten in the degree block notation reads

$$T_{kk'} = \sigma_{kk'}N_{kk} + \tau_k^{-1}N_{k'k} \tag{S17}$$

By substituting the expressions of $\sigma_{kk'}$ and τ_k and considering the equilibrium relations for N_{kk} and $N_{kk'}$ we recover the expression reported in the section Methods of the main text

$$T_{kk'} = \frac{\sigma\bar{N}}{\langle k^{\phi}\rangle}(kk')^{\theta}\left(\nu_k + \nu_{k'}\right). \tag{S18}$$

3 Invasion threshold for the homogeneous network

With the aim of understanding the role played by the topological heterogeneity, in the main paper we address the comparison between the heterogeneous and the homogenous network. Here, we derive the global invasion threshold parameter R_* for the case of a homogeneous network.

All nodes have the same degree $\bar{k}$ and the same population $\bar{N}$, with these values being equal to the average values of the heterogenous case. The mobility fluxes along the links are homogeneously distributed and are determined by the leaving rate $\sigma_{kk'} = \sigma\bar{k}^{2\theta}$ and the length of stay $\tau_k = \bar{\tau}$. In particular we notice that, due to the absence of topological fluctuations, the length of stay becomes now a homogenous quantity, and the exponent χ disappears from the equations. The invasion process is described by the dynamics of the diseased subpopulations at generation n, D^n, that is governed by the equation

$$D^n = D^{n-1}(\bar{k}-1)\left(1 - R_0^{-\lambda_{\bar{k}\bar{k}}}\right)\left(1 - \frac{D^{n-1}}{V}\right), \tag{S19}$$

that is the analogous of Eq. (4) of the main paper. The number of seeds, $\lambda_{kk'}$, is given by the expression

$$\lambda_{kk'} = 2\alpha\,\sigma\bar{N}\bar{\tau}\;\bar{k}^{2\theta}\nu_{\bar{k}}, \tag{S20}$$

where α is the attack rate of the SIR epidemic and $\nu_{\bar{k}} = \left(1 + \sigma\bar{\tau}\;\bar{k}^{2\theta+1}\right)^{-1}$.

We plug Eq. (S20) into Eq. (S19), and we assume that the epidemic is at the early stage (thus $1 - \frac{D^{n-1}}{V} \simeq 1$), and that R_0 is close to unit (thus $\alpha \simeq \frac{2(R_0-1)}{R_0^2}$ [12]). We recover then the explicit form for Eq. (S19)

$$D^n = 4\sigma\bar{N}\bar{\tau}\;\frac{(R_0-1)^2}{R_0^2}\;\bar{k}^{2\theta}(\bar{k}-1)\nu_{\bar{k}}D^{n-1}, \tag{S21}$$

Eq. (S21) provides the expression of R_*. Indeed, D^n will grow exponentially and the epidemic will invade a non-infinitesimal fraction of the network if the following threshold condition is satisfied

$$R_* \equiv 4\sigma\bar{N}\bar{\tau}\;\frac{(R_0-1)^2}{R_0^2}\;\bar{k}^{2\theta}(\bar{k}-1)\nu_{\bar{k}} > 1. \tag{S22}$$

Figure S4 shows the analytical surface $R_*(R_0, \chi)$ as given by Eq. (S22). Parameters are the same as those considered for the case of the heterogenous network as shown in Figure 2 of the main paper. The z-axis range and the color code are chosen in order to allow the comparison between the two cases. With respect to the heterogenous case, R_* reaches significantly smaller values.

4 Dependence of the global invasion threshold on the model parameters

Different values of the parameters governing mobility network structure and traffic distribution can be chosen in order to reproduce specific demographic settings and kinds of mobility. In this section we provide additional examples of the invasion region $R_*(R_0, \chi) > 1$ in varying the parameters θ, γ and σ. Figure S5 shows the curves $R_*(R_0, \chi) = 1$ on three panels each of them considering variations in one of the three parameters, all the others kept the same as in Figure 3. For each parameter two different cases are compared for both the heterogeneous and the homogenous network. The invasion region $R_*(R_0, \chi) > 1$, not colored for the sake of visualization, is always located above the curve, being R_* a monotonous increasing function of R_0. The behavior shown in Figure 3 of the main text is recovered for all the cases analyzed: the curve $R_0(\chi)$ of the critical basic reproductive number for the heterogenous network is a decreasing function, and for low negative values of χ the curve approaches the homogenous one, sometimes intersecting it, as discussed in the main paper. In varying the scenario considered we observe quantitative differences. In particular, as shown in Figure S5(a), larger values of θ corresponding to stronger fluctuations in the traffic distribution enhance the invasion potential of the disease yielding smaller values of the critical basic reproductive number. On the opposite side a more uniform traffic distribution obtained with $\theta = 0.3$ reduces the spreading capability especially in the case of low negative values of χ. For the parameter γ (Figure S5(b)) we consider the values 2 and 2.5 corresponding to networks characterized by a higher degree of topological heterogeneity than the one considered in the main paper. Consistently, we find smaller critical values of the basic reproductive number. Finally, the increasing of the diffusion rate σ (Figure S5(c)) determines an increase in the value of R_* and a consequently decrease of the threshold value of the basic reproductive number.

References

[1] Bull, A. (1995) The economy of travel and tourism (2nd ed.). Australia: Longman.

[2] Decrop, A., Snelders, D. (2004) Planning the summer vacation: an adaptable and opportunistic process. Annals of Tourism Research 31(4), 1008-1030.

[3] Gokovali, U., Bahar, O., Kozak, M. (2007) Determinants of length of stay: a practical use of survival analysis. Tourism Management 28, 736-746.

[4] Eurostat (2007). Tourism Statistics. Data Transmission Compendium.

[5] Keeling, M.J., Rohani, P. (2002). Estimating spatial coupling in epidemiological systems: a mechanistic approach. Ecol. Lett. 5, 20–29.

[6] Sattenspiel, L., Dietz, K. (1995). A structured epidemic model incorporating geographic mobility among regions. Math. Biosci. 128, 71–91.

[7] Pastor-Satorras, R., Vespignani, A. (2001). Epidemic spreading in scale-free networks. Phys. Rev. Lett. 86, 3200–3203.

[8] Pastor-Satorras, R., Vespignani, A. (2001). Epidemic dynamics and endemic states in complex networks. Phys. Rev. E 63, 066117.

[9] Colizza, V., Pastor-Satorras, R., Vespignani, A. (2007). Reaction-diffusion processes and metapopulation models in heterogeneous networks. Nature Physics 3, 276-282.

[10] Colizza, V., Vespignani, A., (2007). Invasion threshold in heterogeneous metapopulation networks. Phys. Rev. Lett. 99, 148701.

[11] Colizza, V., Vespignani A., (2008). Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. J. Theor. Biol. 251, 450-467.

[12] Murray, J. D. (2005). Mathematical Biology. 3rd edition Berlin: Springer Verlag.

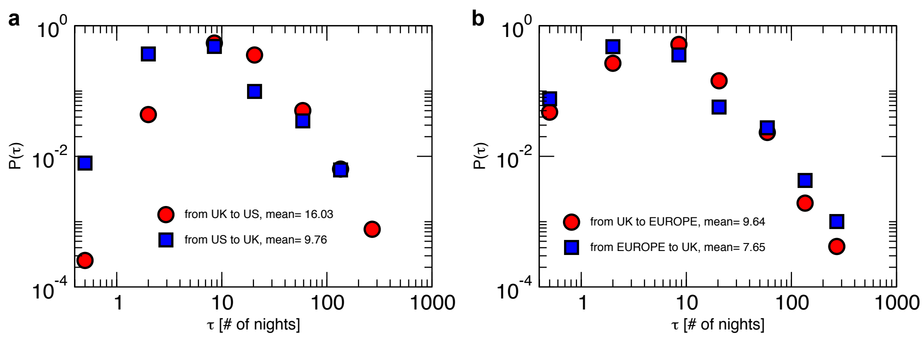

Figure S1: Distribution of the length of stay of trips made by foreign travellers visiting the UK (in blue) and by British travelling abroad (red). Data are aggregated over 7 time periods: nil nights, 1 to 3, 4 to 13, 14 to 27, 28 to 90 nights, 3 to 6 month, 6 month to 1 year. Each panel reports different statistics according to the country of origin and destination: USA (**a**), European countries (**b**).

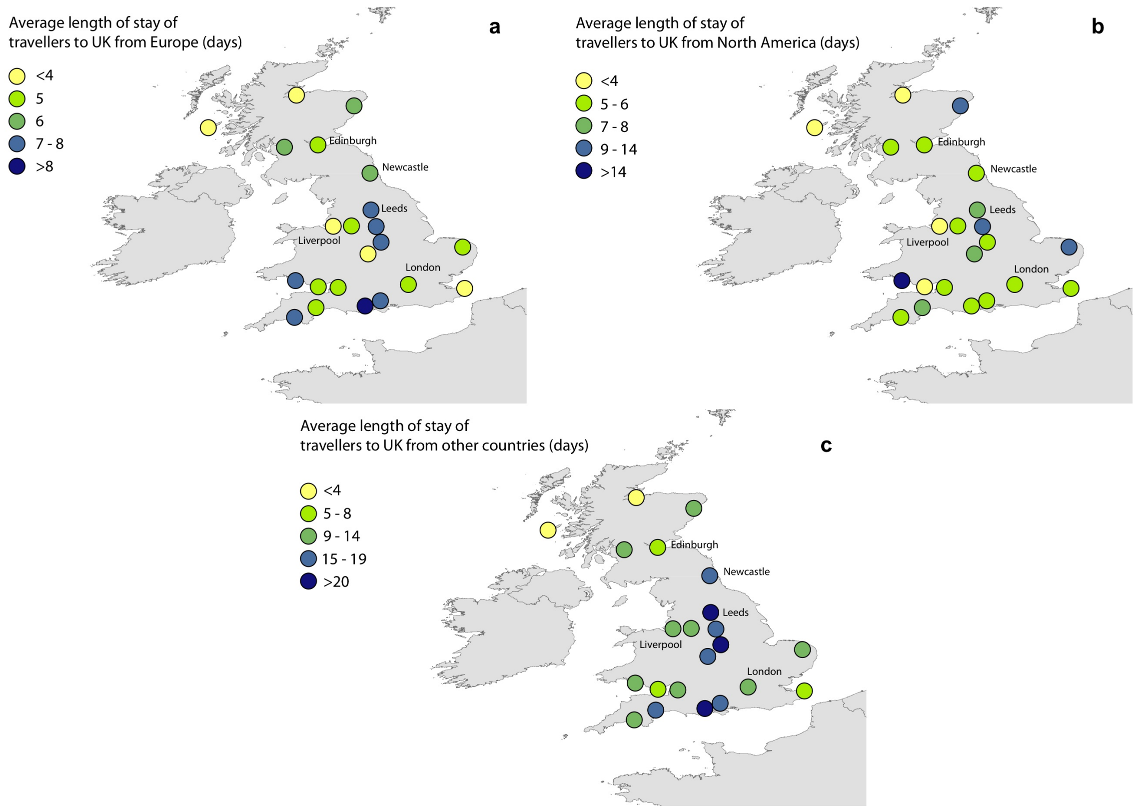

Figure S2: Geographical variations in the average length of stay of foreign travellers in different cities of the UK. Each panel reports different statistics according to the country of origin: European countries (**a**), North America (**b**), other countries **c**). Data are reported by Travel Trends 2009.

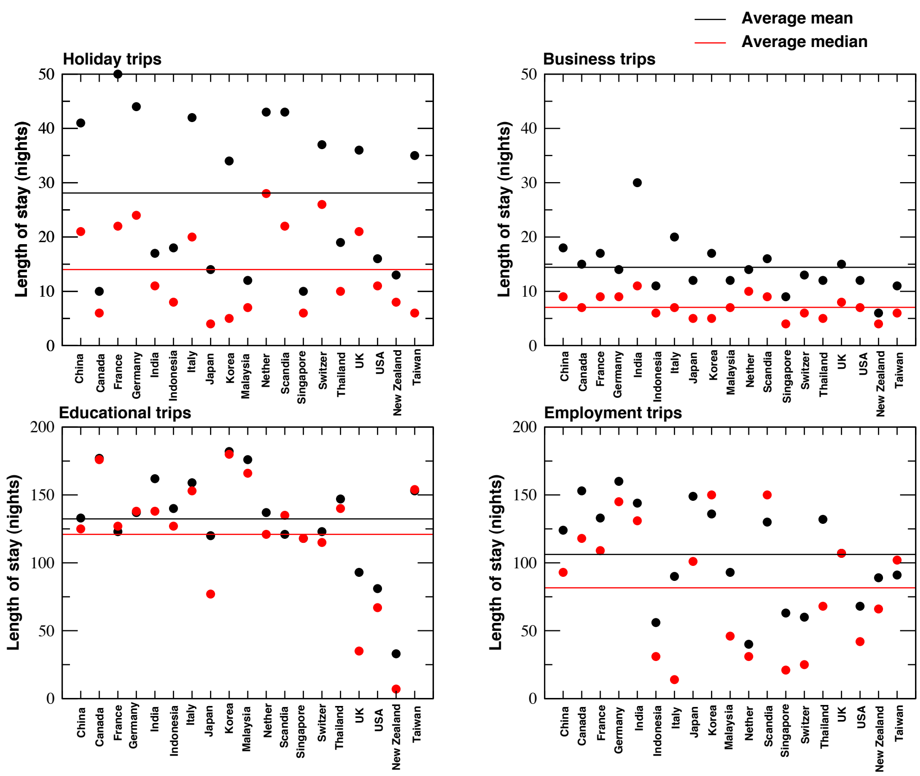

Figure S3: Mean (black dots) and median (red dots) length of stay, measured as number of nights, of travellers who visited Australia in 2008 from different origin countries, according to the purpose of their visit: holiday, business, educational and employment. Black (red) lines indicate the average mean (median) length of stay over all countries of origin.

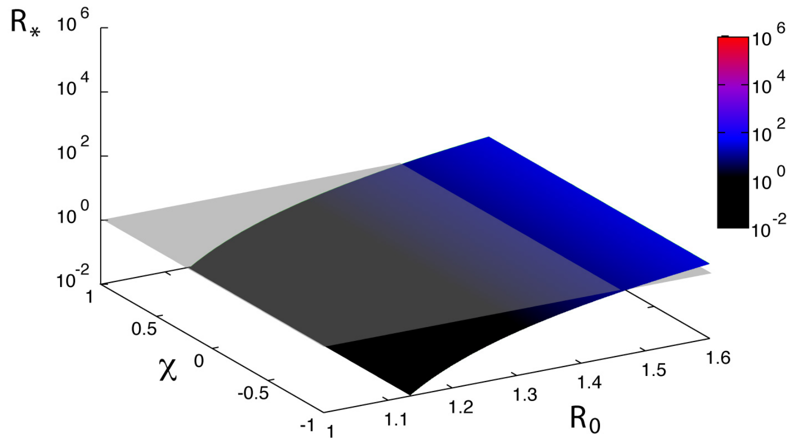

Figure S4: Analytical surface $R_*(R_0, \chi)$ for a homogeneous network. The range of the z-axis and the color code are chosen in order to allow the comparison with the surface obtained for the heterogeneous case, shown in Figure 3 of the main paper

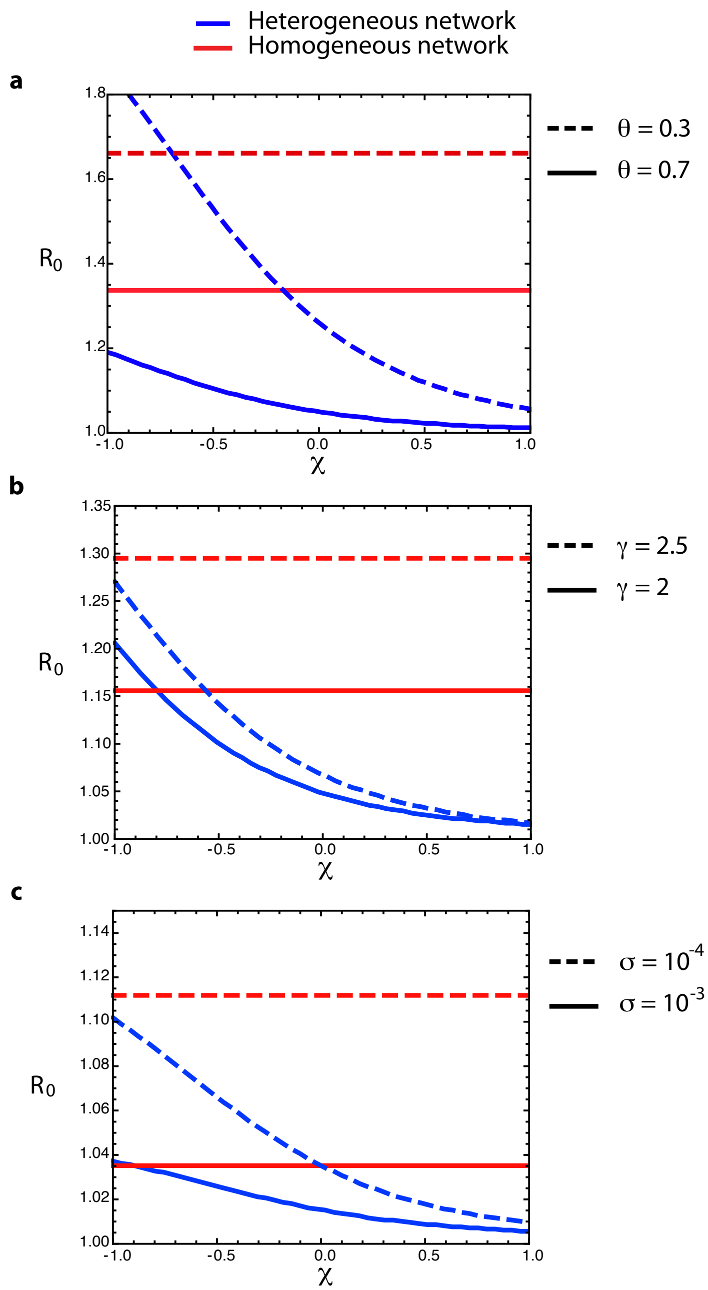

Figure S5: Global invasion thresholod $R_*(R_0, \chi) = 1$. Each panel displays the effects of variations in the parameter indicated in the legend keeping the others the same as in Figure 3 of the main paper. In each panel two different values of the parameter are compared (continuous and dashed lines) for both the heterogenous and the homogenous network (blue and red respectively).