Disease persistence on temporal contact networks accounting for heterogeneous infectious periods

Alexandre Darbon, Davide Colombi, Eugenio Valdano, Lara Savini, Armando Giovannini, Vittoria Colizza

Disease persistence on temporal contact networks accounting for heterogeneous infectious periods

Alexandre Darbon[1], Davide Colombi[1], Eugenio Valdano[2],

Lara Savini[3], Armando Giovannini[3] and Vittoria Colizza[1]

[1]INSERM, Sorbonne Université, Institut Pierre Louis d'Épidémiologie et de Santé Publique (IPLESP), 75012 Paris, France
[2]Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Tarragona 43007, Spain
[3]Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise 'G. Caporale', Teramo 64100, Italy

DC, 0000-0001-6901-5022

The infectious period of a transmissible disease is a key factor for disease spread and persistence. Epidemic models on networks typically assume an identical average infectious period for all individuals, thus allowing an analytical treatment. This simplifying assumption is, however, often unrealistic, as hosts may have different infectious periods, due, for instance, to individual host–pathogen interactions or inhomogeneous access to treatment. While previous work accounted for this heterogeneity in static networks, a full theoretical understanding of the interplay of varying infectious periods and time-evolving contacts is still missing. Here, we consider a susceptible-infectious-susceptible epidemic on a temporal network with host-specific average infectious periods, and develop an analytical framework to estimate the epidemic threshold, i.e. the critical transmissibility for disease spread in the host population. Integrating contact data for transmission with outbreak data and epidemiological estimates, we apply our framework to three real-world case studies exploring different epidemic contexts— the persistence of bovine tuberculosis in southern Italy, the spread of nosocomial infections in a hospital, and the diffusion of pandemic influenza in a school. We find that the homogeneous parametrization may cause important biases in the assessment of the epidemic risk of the host population. Our approach is also able to identify groups of hosts mostly responsible for disease diffusion who may be targeted for prevention and control, aiding public health interventions.

1. Introduction

Mathematical modelling of infectious diseases provides an important tool in understanding patterns and determinants of disease spread [1,2]. The foundations of this field rest upon *compartmental models*, in which the population is divided into different classes (compartments) according to the individual health status, and disease progression is modelled through transitions between compartments. This allows the mathematical description of the epidemic outcome, using, for instance, coupled differential equations [2,3]. While usually being tractable, this framework, however, often relies on two important simplifying assumptions. The first is the homogeneous mixing approximation for which the probability of having a contact with an infectious host is identical for each individual of the population, neglecting any underlying social or spatial structure. Such approximation, unrealistic in many circumstances [4–9], was overcome by accounting for host contact heterogeneities in different ways. In this work, we focus on the network epidemiology approach where each individual is represented by a node in a network and edges encode the interactions between nodes. The contact network can be static [10–14] or dynamic [7–9,15–21], representing interactions that evolve in time. In both cases, network structure strongly influences the spreading process, since it depends on the interplay between the infection dynamics and the structural and temporal features of the networks.

The second assumption involves the recovery process, when infected hosts recover from infection either going back to the susceptible state, or acquiring immunity. Typically, models assume that all the infected individuals have identical constant recovery rate [2,3]. This implies that the average duration of the infectious period (the interval during which an infectious host can transmit the disease) is the same for each individual and is exponentially distributed. Recovery is, therefore, described by an identical Markovian process for all individuals in the population that, while being analytically more tractable, can be unrealistic in many contexts. Individuals may indeed be characterized by different genetic [22] and immunogenetic [23] profiles, or may differ regarding age, medical treatment and vaccination [24]. These features, combined with specific epidemiological characteristics, may cause large deviations from the exponential distribution [25–27], and lead to large heterogeneities in the duration of the infectious period at the individual level. An increasing body of the literature started to account for individual heterogeneities, considering mainly individual variations in infectivity [5,28] and susceptibility [29–31]. Preliminary works incorporated this feature only in homogeneously mixed populations through integro-differential equations [10,32], partial differential equations or through the method of stages [27,33–37]. Only recently, the effects of infectious period heterogeneity are analysed on static networks, using message passing approaches [38,39] and heterogeneous mean-field models [40]. These works provide a description of the susceptible-infectious-recovered (SIR) model on static networks and the computation of the basic reproductive number R_0 [2,3]. These approaches, however, neglect the temporal dynamics of the network of contacts between hosts, which becomes particularly relevant if the timescale of the evolution of contacts is comparable to the one of the spreading process [7–9,17,18].

The aim of this work is to understand how heterogeneous durations of the infectious period impact the conditions for the spreading of an epidemic on a time-evolving contact network, in the particularly interesting scenario of the disease timescale being comparable to the one of the underlying network. By extending the infection propagator approach developed in [41,42], we build an analytical framework that allows us to go beyond the two aforementioned assumptions providing an analytical form for the epidemic threshold, i.e. the critical transmission probability below which a pathogen would go extinct in the population. The epidemic threshold is a key epidemiological quantity as it can measure the vulnerability of a system to the introduction of a specific pathogen.

Lastly, we apply this methodology to evaluate the vulnerability of three different real-world systems: bovine tuberculosis in Southern Italy, nosocomial carriage of pathogenic bacteria in hospital facilities and pandemic influenza in closed settings. We show that using compartmental models with homogeneous infectious periods may introduce important biases in the estimate of the epidemic threshold and thus of the epidemic risk of a host population.

2. Methods

2.1. Infection propagator for heterogeneous infectious periods

In this section, we introduce the analytical framework that allows us to compute the epidemic threshold for a spreading process on a time-varying contact network, considering that infected hosts have a constant rate of recovery, but each host is associated with an individual average infectious period.

We start from a susceptible-infectious-susceptible (SIS) disease progression [2,3], and assume a discrete time evolution for both the network and the spreading process. Each node of the network represents a host, and each link a time-resolved contact that is relevant to pathogen transmission. The temporal network is characterized by a finite number of snapshots T (network period) and formally it can be represented by a list of adjacency matrices whose element $W_{ij}^{(t)}$ corresponds to the weight of the link made from node i to node j at time t. At each time step, infected nodes transmit the infection to the susceptible neighbours with a probability that depends both on the weight of the link and on the specific disease transmissibility λ. The weight may correspond to different quantities (e.g. duration or strength of a contact) depending on the context under study, as we will see in the real-world systems addressed in this work. Infectious nodes can spontaneously recover with probability μ. Recovery is a Poisson process with an average infectious period $\tau = 1/\mu$ that in the classic compartmental model formulation is the same for each individual.

Following the approach introduced in [41], the critical behaviour of the epidemic spreading process is completely described by the following quantity, called infection propagator:

$$P(W; \lambda, \mu) = \prod_{t=1}^{T}(1 - \mu + \Lambda^{(T-t)}) \quad \text{with } \Lambda_{ij}^{(t)} = 1 - (1 - \lambda)^{W_{ij}^{(t)}} \quad \text{and} \quad \mu = \mu I. \tag{2.1}$$

The sufficient and necessary condition for the existence of the asymptotically stable disease-free solution is that the spectral radius of P (i.e. its lead eigenvalue) is smaller than 1. Therefore, for a given value of μ, the critical value of the transmissibility λ_c for which the spectral radius is equal to 1 corresponds to the epidemic threshold.

To account for hosts with individual average infectious period τ_i, with i corresponding to the node index, we define μ as a vector where each element μ_i is the inverse of the individual infectious period $\mu_i = 1/\tau_i$. The infection propagator can then be written as

$$P(W; \lambda, \mu) = \prod_{t=1}^{T} M^{(T-t)} \quad \text{with } M_{ij}^{(t)} = \delta_{ij}(1 - \mu_i) + 1 - (1 - \lambda)^{W_{ij}^{(t)}}, \tag{2.2}$$

with δ_{ij} being the Kronecker delta, so that the epidemic threshold expression becomes

$$\rho[P(W; \lambda = \lambda_c, \mu)] = 1, \tag{2.3}$$

where μ is a diagonal matrix with $\vec{\mu}$ on the diagonal.

2.2. Epidemic threshold computation

We compute the spectral radius of the infection propagator P, given by the expression of equation (2.2) using the power iteration method. To compute the threshold, we find the zero for $\rho[P(W; \lambda = \lambda_c, \mu)]^{1/T} - 1 = 0$ using Brent algorithm [43], where λ is a variable and μ is informed from epidemiological data. This procedure returns the value of λ for which the spectral radius is equal to one. A Python library to compute the critical transmissibility is publicly available online for interested researchers [44].

2.3. Epidemic risk assessment in the heterogeneous versus homogeneous cases

We compute the epidemic threshold with the infection propagator approach to assess the epidemic risk of a host population in three real case studies. In each one of them, actual epidemiological data allow us to inform the infection propagator $P(W; \lambda, \mu)$ with empirical estimates for μ_i, corresponding to the actual heterogeneous context. We then compare our findings with the epidemic threshold obtained under the assumption that all hosts have the same infectious period (referred to as the homogeneous case) obtained as a population average, i.e. $\tau_{\text{all}} = \sum_i \tau_i/N$, where N is the total number of hosts (nodes) in the population. This quantity enables a parametrization for the comparison of the two cases.

2.4. Network properties

We define the activity potential of a node in an empirical network as the fraction of the number of time steps during which it makes contact with other nodes over the period of the network T, as defined in [21].

Other quantities used in this study are basic network measures, including the degree, i.e. the number of neighbours a given node has at time t, and the strength, i.e. the total sum of the weights on the

connections that a given node establishes at time t. These quantities are also considered aggregated on a given time interval [45].

2.5. Hellinger distance

To compare two distributions of the epidemic threshold obtained by varying underlying conditions in the host population under study, we use the Hellinger distance [46] that is defined for discrete distributions as follows:

$$H(P_1, P_2) = \left(\frac{1}{2} \sum_{i=1,k} (\sqrt{p_i} - \sqrt{q_i})^2 \right)^{1/2}$$

where P_1 and P_2 are discrete probability distributions over k discrete values: $P_1 = (p_1, \ldots, p_k)$ and $P_2 = (q_1, \ldots, q_k)$. We also considered other measures quantifying the difference between the distributions and obtained results similar to those presented in the next section.

3. Results

3.1. Bovine tuberculosis in Italy

Livestock infectious diseases are of primary interest for animal health and welfare and for the economy of a country, with the potential to produce devastating consequences [47]. Bovine tuberculosis is one of the most widespread zoonotic diseases all over the world [48] and it is very difficult to contain since it can spread while unnoticed, before detection in the animals [49]. Bovine tuberculosis is a notifiable disease in Europe. In Italy, in particular, it has affected cattle population for decades and still circulates in southern regions [50]. Here, we focus on Puglia region in the south of Italy where outbreak data are available and the outbreak duration was shown to vary depending on the production type of the affected premises.

We built the temporal network starting from the dataset of cattle trade movements of the whole population of bovines in Puglia obtained from the Italian National Database for Animal Identification and Registration that is managed by the Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise (IZSAM) on behalf of the Italian Ministry of Health [51,52]. Each time-stamped cattle movement record provides the animal unique identifier and the identifiers of both origin and destination premises. Here, we consider the daily records of bovine trades in Puglia, from 1 January 2006 to 31 December 2012. In this region, 5430 animal holdings displaced 136 206 bovines through 44 272 trade movements in the time period under study. In the network representation, each node corresponds to a farm, and directed links represent the animal movements, weighted with the number of bovines moved [53,54].

Bovine displacements tend to concentrate in the countryside around the three majors cities of the region: Foggia, Bari and Lecce (figure 1a). Outbreak locations correspond to these three geographical clusters, and also appear to be grouped according to the production type of the affected premises (i.e. meat, dairy or mixed production). We define the monthly aggregated average degree of an animal holding as the average number of trading partners it has in a month. We compute this quantity by averaging on all premises and averaging only on the ones trading at least once a month (defined as active, including both sales and purchases). The evolution of these quantities over time shows the low density of the network (figure 1b). Each month active premises represent between 8 and 14% of the total number of premises and their average degree is around 3, leading to a very low overall average degree. Further analysis of the Italian cattle trade network properties was performed in the previous works [9,55,56].

For the outbreak data, we used records of bovine tuberculosis cases in cattle occurring in Puglia between 1983 and 2015. These data were collected by the Italian National Animal Health Managing Information System, an informatics system developed by IZSAM on behalf of the Ministry of Health [51,52]. For each outbreak, we considered data reporting the identifiers of the affected animal holding, its production type, and the start and end of the outbreak based on the date of first infection detection and date of clearing of the infection, respectively. The farm is considered to be cleared from the infection after two consecutive negative tests on all animals that need to be performed each after a minimum of four and a maximum of six weeks from the previous control activity (either the
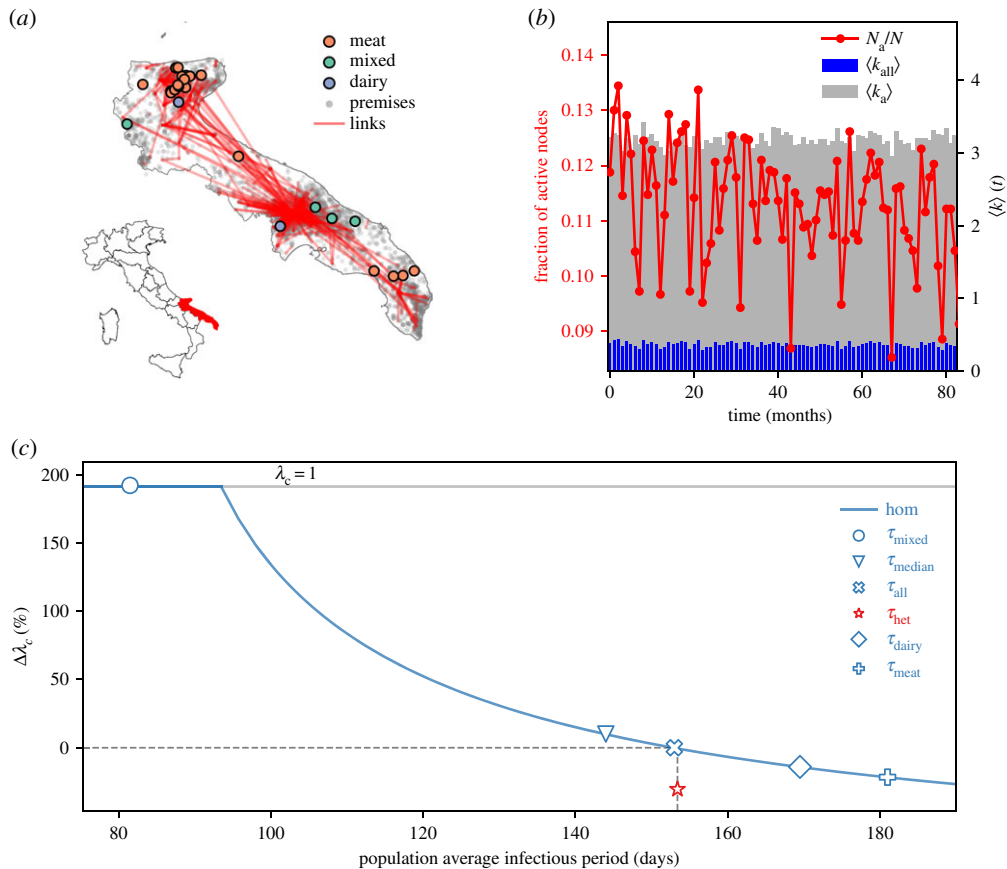
Figure 1. Bovine tuberculosis in Puglia, Italy. (a) Aggregated network of the cattle trade movements in Puglia and location of bovine tuberculosis outbreaks in this region, in the period 1983 – 2015. Only the links with an aggregated strength higher than 30 are represented. Affected farms are identified according to their production type (dairy, meat, mixed). (b) Activity patterns of the cattle trade network in time. The red line shows the proportion of active nodes. Blue bars show the monthly average degree of all nodes and grey bars show the monthly average degree of active nodes. (c) Epidemic threshold estimate comparison between homogeneous and heterogeneous parametrizations. The plot shows the relative variation of the epidemic threshold (expressed in %) as a function of the population average infectious period. The reference for the relative variation is the epidemic threshold obtained in the homogeneous assumption obtained with the population average infectious period τ_{all} computed from the data (table 1). The red star shows the epidemic threshold estimate in the heterogeneous case. Blue markers show particular cases of the threshold estimate in various homogeneous parametrizations. The blue line shows the full dependence of the epidemic threshold estimate in the homogeneous case as a function of the population average infectious period.

slaughtering or the first test). As shown in table 1, outbreak durations in the region vary with production types, as observed in other contexts [57]. We split the premises population in groups corresponding to their production type, and we set to each node the data-driven infectious period corresponding to the associated production type.

Predicted epidemic risks in the homogeneous or heterogeneous case are rather different (figure 1c). The epidemic threshold computed on heterogeneous durations of the outbreak is about 30% smaller than the estimate obtained with the population average. As a result, the homogeneous assumption overestimates the epidemic threshold. Even if we vary the population average infectious period (used in the homogeneous model), for example, by assuming that all premises are either dairy or meat farms, we obtain a higher epidemic threshold, thus systematically underestimating the vulnerability of the livestock system. In more detail, we obtain the largest bias when we parametrize all premises as if they were all of mixed production. Under this assumption, the obtained epidemic threshold is equal to 1, meaning that the pathogen would not be able to persist in the system, in clear contradiction with the current epidemiological status of Puglia with respect to bovine tuberculosis.

Table 1. Estimated outbreak duration of bovine tuberculosis per production type of the affected farm.

production type	outbreak duration (min–max) (days) [51,52]
meat	$\tau_{meat} = 181\ (74-2029)$
dairy	$\tau_{dairy} = 170\ (65-274)$
mixed	$\tau_{mixed} = 82\ (41-292)$
all	$\tau_{all} = 153$

Table 2. Number and proportion of individuals in each class in the hospital network and corresponding estimated carriage duration for *S. Aureus*.

| role	no. of nodes (%)	carriage duration (days) [61–63]
patients	29 (38.67)	$\tau_p = 10$
nurses	27 (36)	$\tau_n = 2.5$
medical doctors	11 (14.66)	$\tau_m = 2.5$
administrative personnel	8 (10.67)	$\tau_a = 2.5$
all	75 (100)	$\tau_{all} = 5.4$

3.2. Nosocomial infection carriage in hospital facilities

The definition of nosocomial infection applies both to infections acquired by patients while being in a healthcare facility, and to occupational infections of medical staff [58]. Their control is a great concern in public health. Several pathogens cause nosocomial infections, such as viruses (e.g. nosocomial influenza, rotaviruses), or pathogenic and non-pathogenic bacteria (e.g. *Mycobacterium tuberculosis*, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae*, *Enterococcus*, *Staphylococcus aureus*) [59]. While host colonization can last for months in the absence of intervention [60], it can be dramatically shortened by detection and intervention protocols, or by the hygiene measures adopted by hospital personnel [61]. In addition, patients are known to take longer to clear nasal colonization than healthcare workers [61,62]. In order to understand how these differences may affect the vulnerability of a hospital setting to nosocomial infections, we focus here on the case of *S. aureus*, for which colonization durations are documented [61–63].

Close proximity interactions are acknowledged to be a route along which nosocomial infections may be transmitted [62,63]. Here, we use a temporal network of contacts within a hospital ward in Lyon, France, collected by the SocioPatterns project [64,65]. The dataset records face-to-face interactions among patients and hospital workers over 4 days, at a 20 s resolution. It includes also individual information that allows us to divide the population into four groups: patients, nurses, medical doctors and administrative personnel (table 2). Spatial constraints appear to be major drivers in network topology, as, for instance, patients are mostly confined to rooms, and administrative personnel to offices. Nurses, moreover, are the most connected, sharing links with all other groups and among themselves (figure 2*a*). The number of contacts follows a daily cycle, featuring peaks during the daytime, and troughs at night (figure 2*b*).

Analogously to the case of bovine tuberculosis, we compare the epidemic threshold obtained under homogeneous and heterogeneous parametrizations of the infection propagator approach. In the heterogeneous case, patients and personnel colonization durations are set to, respectively, 10 and 2.5 days [61–63] while $\tau_{all} = 5.4$ days (table 2). The epidemic risk analysis presented in figure 2*c* shows that the epidemic threshold in the heterogeneous case yields a 100% relative variation with respect to the homogeneous parametrization with τ_{all}. This hints at the presence of a group of potential hosts that are critical to the spread of the pathogen, but have a short infectious period. To obtain the same estimates of the epidemic threshold of the heterogeneous case with the homogeneous parametrization we would need to reduce the average infectious period to $\tau_{all} = 2.8$ days.

To understand the contribution of each class of individuals to the epidemic risk, we computed the epidemic threshold in the heterogeneous parametrization by reducing the carriage duration parameter of a single class at a time for each of the four classes. For the hospital staff, we set the carriage
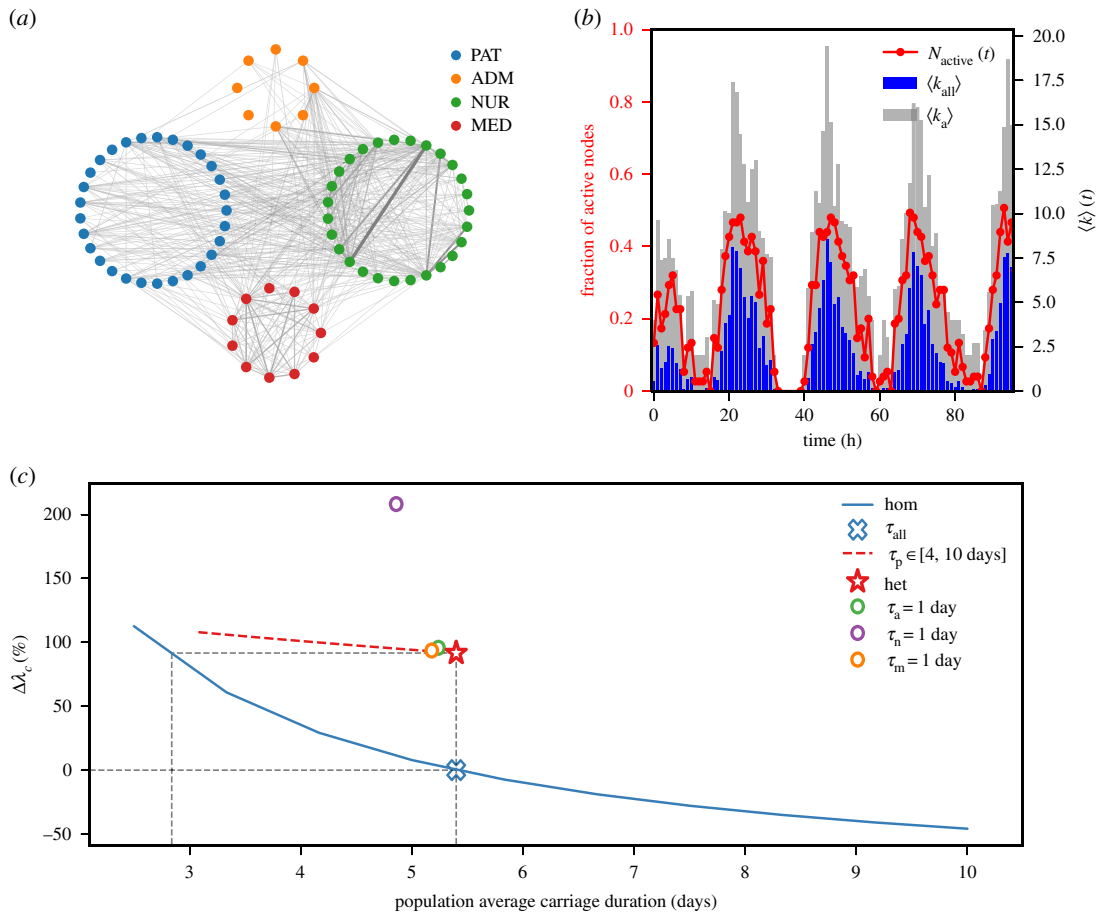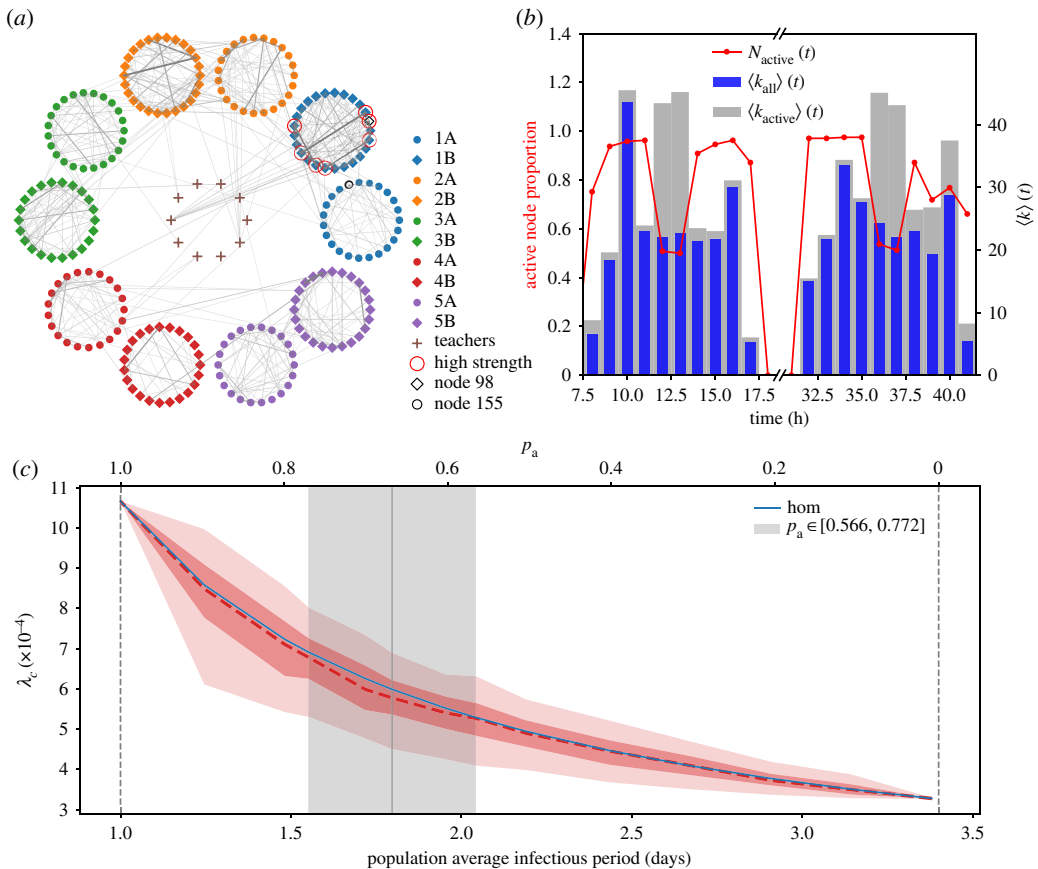
Figure 2. *Staphylococcus aureus* diffusion in a hospital ward. (*a*) Aggregated network of proximity contact in the hospital ward. Nodes are grouped by class of individuals and links' width and colour are proportional to links' weight, measuring the total amount of time associated with each contact. (*b*) Activity patterns of the hospital network in time. The red line shows the proportion of active nodes. Blue bars show the hourly average degree of all nodes and grey bars show the hourly average degree of active nodes. (*c*) Epidemic threshold estimate comparison between homogeneous and heterogeneous parametrizations. The plot shows the relative variation of the epidemic threshold (expressed in %) as a function of the population average infectious period. The reference for the relative variation is the epidemic threshold obtained in the homogeneous assumption (blue cross) obtained with the population average infectious period τ_{all} computed from the data (table 2). The red star shows the epidemic threshold estimate in the heterogeneous case. Circles correspond to the estimates obtained when reducing the carriage duration of each class of hospital personnel to 1 day, in the heterogeneous parametrization. The red dashed line shows the estimates obtained when reducing the carriage duration of patients in the range of 4–10 days, in the heterogeneous parametrization. The blue line shows the full dependence of the epidemic threshold estimate in the homogeneous case as a function of the population average infectious period.

duration to 1 day, and for the patients, we explored a reduction in their carriage duration from 10 to 4 days. This allows us to understand the performance of an intervention aimed at reducing the epidemic risk for the ward, using, for instance, more frequent hand washing for a given class of staff, or enhanced screening and treatment of patients. We observe that while the intervention on administrative personnel or medical doctors would have almost no impact on the epidemic threshold, targeting nurses would dramatically increase the epidemic threshold, yielding a relative variation larger than 200% with respect to the homogeneous assumption. This highlights the important role of nurses in the network of contacts, and their potential of largely facilitating the dispersal of nosocomial infections in the hospital setting [63,66–68].

3.3. Pandemic influenza in closed settings

Influenza is a respiratory infectious disease that can spread through proximity contacts, and affects 10–30% of European population every year [69]. High-risk individuals, such as elderly or immune-deficient,

Figure 3. Pandemic influenza in a school setting. (*a*) Aggregated network of proximity contacts in the school. Nodes are grouped by class of individuals and links' width and colour are proportional to links' weight, measuring the aggregated amount of time associated with each contact. (*b*) Activity patterns of the school network in time. The red line shows the proportion of active nodes. Blue bars show the hourly average degree of all nodes and grey bars show the hourly average degree of active nodes. (*c*) Epidemic threshold estimate comparison between homogeneous and heterogeneous parametrizations. The plot shows the epidemic threshold as a function of the population average infectious period, for both homogeneous (blue) and heterogeneous (red) parametrizations. The red shaded areas show the 95% (lighter red) and 50% (darker red) confidence intervals for the epidemic threshold estimate obtained when randomizing on nodes that are symptomatic or asymptomatic in the heterogeneous case. The grey vertical area indicates the range of population average infectious period corresponding to p_a varying in the empirically estimated 95% confidence interval.

may experience a severe form requiring hospitalization. Particular attention is given to schoolchildren [70,71] since epidemiological evidence suggests that they are critical in the early transmission chains of the disease favouring then the diffusion in the general population. Hosts infected with influenza may develop symptoms or may be asymptomatic, also exhibiting a shorter infectious period in the absence of symptoms [24]. We aim at understanding how the presence of asymptomatic individuals impacts the pandemic risk for a population of students at a school.

We build the temporal network from data reporting time-resolved contacts in a French school [72,73] collected by the SocioPatterns project [64], over a period of 2 days, at a resolution of 20 s. As for hospital facilities, proximity sensors were used to record face-to-face contacts between individuals. The resulting network is composed of $N = 242$ individuals divided into 11 classes, corresponding to 10 classes distributed on five consecutive grades, plus teachers. The network exhibits strong community structure, as children connect more within the same class or the same grade (figure 3*a*). While teachers connect with students from different classes, those links have a lower weight, as they occur less frequently and are shorter in time. The hourly activity timeline shows clearly the three main breaks of the day (figure 3*b*). The proportion of active nodes is very high, showing a high degree of interaction among students [74,75]. It is possible to distinguish the morning and afternoon breaks, characterized by a proportion of active nodes close to 1, from the lunch break where students have lunch or leave the school to eat at home decreasing the network activity. Detailed analysis of the network was reported in the previous work [72,73].

Table 3. Estimated infectious period per symptomatic or asymptomatic influenza infections. The population average τ_{all} depends on the assumed proportion p_a of asymptomatic individuals in the schoolchildren population.

type of infection	infectious period (days) [24]
symptomatic	$\tau_s = 3.4$
asymptomatic	$\tau_a = 1$
all	$\tau_{all} = \tau_{all}(p_a)$

We model the epidemic using infectious period estimates from 2009 influenza pandemic [24,76], yielding 3.4 days for symptomatic infections and 1 day for asymptomatic infections (table 3). We investigate proportions of asymptomatically infected schoolchildren (p_a) in agreement with empirical estimates, i.e. $p_a = 0.669$ (95% CI: 0.556–0.772) [77]. In addition to these values, we also explore the full range from $p_a = 0$ to $p_a = 1$ for a comprehensive analysis. For a given proportion p_a of asymptomatic individuals, we compute the epidemic threshold randomly extracting the asymptomatic nodes $p_a N$ in the population. Repeating this operation 500 times, we obtain a distribution of epidemic threshold values for each value or p_a depending on the assumed symptomatic response of the host population of schoolchildren.

Figure 3c reports the results of the comparison between the homogeneous and heterogeneous parametrizations. We observe that without any asymptomatic individual (i.e. $p_a = 0$ and population average infectious period equal to 3.4 days) the threshold is lower than when considering the presence of asymptomatic, as expected. While it may be natural to think that asymptomatic individuals may create unnoticed paths of infection, their shorter infectious periods indeed play a role in reducing the epidemic vulnerability of the population. When we consider the heterogeneous parametrization, we find that the homogeneous epidemic threshold is almost equal to the median heterogeneous one. This observation is not only true for values of p_a in the confidence interval of the empirical estimates, but also for the whole range from 0 to 1. There are, however, significant fluctuations around the median value, whose width corresponds to approximately 40% of the homogeneous estimate. The epidemic risk assessment assuming homogeneous infectious periods in the population may therefore introduce a large bias leading either to underestimating or overestimating population's risk to infection.

We now want to understand the mechanisms responsible for those fluctuations, specifically, by identifying the hosts that play a key role in shaping the vulnerability. In what follows, we fix the proportion of asymptomatic individuals to the empirical estimate $p_a = 0.669$ [77]. For each node, we determine the two threshold distributions obtained considering the given node as either symptomatically or asymptomatically infected. We then compute the Hellinger distance between these two distributions to quantifies their dissimilarity: the higher the distance is, the more different the two distributions are (see Methods). This quantity will be higher for nodes whose status has a larger impact on the epidemic threshold. The distribution of the Hellinger distance of all nodes, shown in figure 4a, is bimodal and the subset of nodes mainly responsible for the epidemic threshold variation is clearly visible. It is a small group composed of only seven nodes, corresponding to 2.9% of the school population. In figure 4b, we show two examples of the distributions obtained for a node selected in each mode of the distance distribution.

We also observe that these seven nodes belong to the same class (see circled nodes in figure 3a) and do not correspond to hubs in the whole network. They are likely to be a single group of friends. In order to highlight the patterns making these specific nodes more responsible than others for the threshold variation, we investigate various network properties. We find that these nodes exhibit higher strength (figure 4c) and activity potential (not shown).

4. Discussion

Mathematical models have been highly successful in the study and understanding of infectious disease epidemics [2,3]. Some simplifying assumptions, however, have sometimes hindered their applicability to real-world scenarios. Our work helps adding realism to modelling a wide class of highly relevant diseases, as it provides a synthetic and solvable framework for considering individual hosts (or host
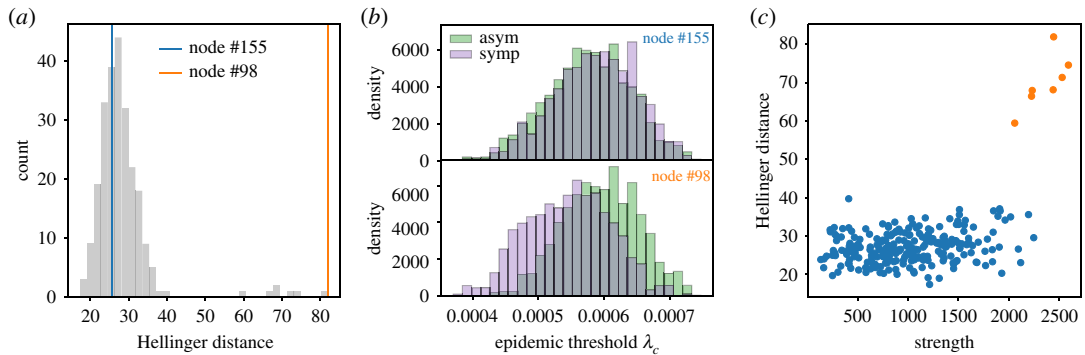
Figure 4. Identification of nodes mostly contributing to the large fluctuations obtained with the heterogeneous parametrization compared to the homogeneous case. (a) Distribution of the Hellinger distance computed on the epidemic threshold distributions obtained by considering one at a time each node in the symptomatic or asymptomatic classes. The blue and the orange lines show examples of distance values obtained for two nodes, one close to the peak of the distribution (blue) and the other being an outlier (orange). These nodes are identified in the network representation of figure 3. (b) Epidemic threshold distributions depending on symptomatic status following infection (purple for symptomatic infection, green for asymptomatic infection), for the two nodes highlighted in (a). (c) Hellinger distance as a function of the node strength. Orange points correspond to nodes with outlier Hellinger distance values. These nodes are identified in the network representation of figure 3.

classes) that have different characteristic infectious periods. By building on the methodology described in [41,42], we have defined and computed the epidemic threshold with arbitrarily heterogeneous recovery probabilities, assuming them to be constant. We then applied our analytical framework to three relevant case studies, showing that we can successfully measure how such heterogeneity impacts the vulnerability of a particular population to disease introduction, considering bovine tuberculosis in southern Italy, nosocomial infections in hospital wards, and influenza-like epidemics in schools.

We stress that, while an increasing body of the literature has already demonstrated the importance of including contact structure [4–16] and host-specific heterogeneities [25–31,37,78], few studies have considered the combined impact of these two factors in the estimation of the epidemic risk. Theoretical works that include host-specific heterogeneities were previously carried out while assuming homogeneously mixed population [10,27,32,33,37] or using a static contact network [38–40], neglecting the possible interplay between the spreading process timescale and the contact evolution.

Our findings suggest that obtaining an accurate, individual or class-specific, estimate of the infectious period is an important step for building realistic spreading models. Indeed we have shown that if we set the same recovery probability to all hosts, we may greatly bias the estimate of the vulnerability, with the sign of the error depending on the specific setting. For bovine tuberculosis in Puglia, any homogeneous parametrization of the outbreak duration predicts a lower epidemic risk with respect to the heterogeneous case. In particular, when the shortest recorded outbreak duration is considered in the homogeneous approximation ($\tau = \tau_{mixed}$), the disease is predicted not to be able to persist in the system, in evident contrast with the documented endemic presence of bovine tuberculosis in the region. On the contrary, in the case of the *S. Aureus* infection within a hospital, the classic homogeneous assumption causes to substantially overestimate the epidemic risk, with respect to the heterogeneous case. The presence of medical personnel with a shorter carriage duration is able to counteract the spreading potential of patients assumed to be carrying the pathogen for a longer time, and to further reduce the risk with respect to a homogeneous parametrization with the population average. Finally, for pandemic influenza in school settings, even though the median epidemic threshold obtained from the heterogeneous assumption matches the homogeneous result, we observe important fluctuations that may lead to either underestimating or overestimating vulnerability.

All these findings highlight the importance of properly accounting for heterogeneous infectious periods, as ignoring this feature may lead to a biased estimate of the epidemic risk, and consequently inefficient control strategies [79].

In addition, despite the fact that the responsibility of more active nodes in the epidemic risk has already been proven [45,56], our approach was able to highlight them and show their contribution to population vulnerability, depending on their individual infectious period and their network properties. This can have clear implications in devising targeted interventions. For example, we have shown that in the hospital ward hygiene practices that reduce carriage duration in nurses can lead to

a strong reduction in nosocomial risk. No impact is instead observed reducing the carriage duration of the administrative personnel or medical doctors. These findings are explained by the highly connected role of nurses within the hospital facility, given they interact considerably with both patients and other hospital personnel. As such, they represent an ideal target group for prevention measures aimed at lowering the risk for pathogen diffusion [65,80–82]. Also in the case of pandemic influenza in school, our approach allows the identification of those individuals who mostly contribute to disease circulation and persistence.

Our findings provide useful insights for the understanding of host heterogeneities in disease spread and can be used to build more realistic data-driven mathematical approaches for real case scenarios and targeted control measures. However, several key theoretical and practical issues are still to be addressed. First, empirical evidence shows that asymptomatically infected individuals with influenza tend to be less infectious than those who develop symptoms [24]. Our approach, however, assumes that disease transmissibility would not vary depending on host symptoms status, as our study was restricted to the role of heterogeneity of infectious periods only. Such variation can be taken into account by introducing a variation of the weights of the contacts established by an asymptomatic infectious individual while infected. The framework with varying transmissibility may also apply to other epidemic contexts, such as e.g. the spread of nosocomial infections. Nurses and doctors indeed are expected to adopt hygiene measures (e.g. use of disposable gloves, hand washing, etc.) that can reduce their transmissibility if infected. Another host-specific factor that is disregarded by our approach, for the sake of simplicity, is individual susceptibility that may play an important role in the dynamics of many infectious diseases (e.g. influenza).

Second, our study considered a basic intervention expressed with the reduction of the carriage duration for medical personnel in the hospital. More realistic interventions may, however, be considered, as e.g. pharmaceutical treatments aimed not only at reducing the infectious period but also the probability of transmission of the disease (e.g. antiviral treatments for influenza infections [24], or the use of antibiotics for bacterial infections). In addition, recent work highlighted the role of heterogeneous carriage duration of bacterial strains in ruling the relative abundance of each strain under antibiotic treatments [83]. Accounting for these aspects would allow a more realistic design of targeted interventions aiming at raising the epidemic threshold in the most efficient way.

Finally, here we considered a limited set of infectious period values instead of full distributions, for the sake of simplicity (e.g. to highlight the role of productive types in the spread of bovine tuberculosis, instead of focusing on farm-specific fluctuations) and because of lack of data (e.g. in the contexts of pandemic flu or *S. aureus* spread). Nonetheless, the infectious propagator approach is able to integrate host-specific recovery rates, as presented in the Methods section. Going beyond this, more realistic and pathogen-specific distributions of infectious periods should be further considered, differently from the exponential distributions with heterogeneous individual infectious periods assumed here. This would probably require a radical redesign of the infection propagator approach to address the inclusion of such distributions. These various future directions would help the theoretical understanding of disease spreading processes with increasingly realistic epidemic models.

References

1. Kermack WO, McKendrick AG. 1927 A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A* **115**, 700–721. (doi:10.1098/rspa.1927.0118)

2. Anderson RM, May RM. 1992 *Infectious diseases of humans: dynamics and control*. Oxford, New York: Oxford University Press.

3. Keeling MJ, Rohani P. 2011 *Modeling infectious diseases in humans and animals*. Princeton, NJ: Princeton University Press.

4. Woolhouse ME *et al.* 1997 Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Natl Acad. Sci. USA* **94**, 338–342. (doi:10.1073/pnas.94.1.338)

5. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. 2005 Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359. (doi:10.1038/nature04153)

6. Bansal S, Read J, Pourbohloul B, Meyers LA. 2010 The dynamic nature of contact networks in infectious disease epidemiology. *J. Biol. Dyn.* **4**, 478–489. (doi:10.1080/17513758.2010.503376)

7. Isella L, Stehlé J, Barrat A, Cattuto C, Pinton J-F, Van den Broeck W. 2011 What's in a crowd? Analysis of face-to-face behavioral networks. *J. Theor. Biol.* **271**, 166–180. (doi:10.1016/j.jtbi.2010.11.033)

8. Rocha LEC, Liljeros F, Holme P. 2010 Information dynamics shape the sexual networks of internet-mediated prostitution. *Proc. Natl Acad. Sci. USA* **107**, 5706–5711. (doi:10.1073/pnas.0914080107)

9. Bajardi P, Barrat A, Natale F, Savini L, Colizza V. 2011 Dynamical patterns of cattle trade movements. *PLoS ONE* **6**, e19869. (doi:10.1371/journal.pone.0019869)

10. Keeling MJ, Rand DA, Morris AJ. 1997 Correlation models for childhood epidemics. *Proc. R. Soc. Lond. B* **264**, 1149–1156. (doi:10.1098/rspb.1997.0159)

11. Pastor-Satorras R, Vespignani A. 2001 Epidemic dynamics and endemic states in complex networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **63**, 066117. (doi:10.1103/PhysRevE.63.066117)

12. Pastor-Satorras R, Vespignani A. 2001 Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200–3203. (doi:10.1103/PhysRevLett.86.3200)

13. Moreno Y, Pastor-Satorras R, Vespignani A. 2002 Epidemic outbreaks in complex heterogeneous networks. *Eur. Phys. J. B* **26**, 521–529. (doi:10.1140/epjb/e20020122)

14. Barthélemy M, Barrat A, Pastor-Satorras R, Vespignani A. 2004 Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Phys. Rev. Lett.* **92**, 178701. (doi:10.1103/PhysRevLett.92.178701)

15. Gross T, D'Lima CJD, Blasius B. 2006 Epidemic dynamics on an adaptive network. *Phys. Rev. Lett.* **96**, 208701. (doi:10.1103/PhysRevLett.96.208701)

16. Read JM, Eames KTD, Edmunds WJ. 2008 Dynamic social networks and the implications for the spread of infectious disease. *J. R. Soc. Interface* **5**, 1001–1007. (doi:10.1098/rsif.2008.0013)

17. Miritello G, Moro E, Lara R. 2011 Dynamical strength of social ties in information spreading. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **83**, 045102. (doi:10.1103/PhysRevE.83.045102)

18. Karsai M, Kivelä M, Pan RK, Kaski K, Kertész J, Barabási A-L, Saramäki J. 2011 Small but slow world: how network topology and burstiness slow down spreading. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **83**, 025102. (doi:10.1103/PhysRevE.83.025102)

19. Ferreri L, Bajardi P, Giacobini M, Perazzo S, Venturino E. 2014 Interplay of network dynamics and heterogeneity of ties on spreading dynamics. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **90**, 012812. (doi:10.1103/PhysRevE.90.012812)

20. Iribarren JL, Moro E. 2009 Impact of human activity patterns on the dynamics of information diffusion. *Phys. Rev. Lett.* **103**, 038702. (doi:10.1103/PhysRevLett.103.038702)

21. Perra N, Gonçalves B, Pastor-Satorras R, Vespignani A. 2012 Activity driven modeling of time varying networks. *Sci. Rep.* **2**, 469. (doi:10.1038/srep00469)

22. Segal S, Hill AVS. 2003 Genetic susceptibility to infectious disease. *Trends Microbiol.* **11**, 445–448. (doi:10.1016/S0966-842X(03)00207-5)

23. Fryer HR, Frater J, Duda A, Roberts MG, 2010 Modelling the evolution and spread of HIV immune escape mutants. *PLoS Pathog.* **6**, e1001196. (doi:10.1371/journal.ppat.1001196)

24. Dorjee S, Poljak Z, Revie CW, Bridgland J, McNab B, Leger E, Sanchez J. 2013 A review of simulation modelling approaches used for the spread of zoonotic influenza viruses in animal and human populations. *Zoonoses Public Health* **60**, 383–411. (doi:10.1111/zph.2013.60.issue-6)

25. Bailey NT. 1975 *The mathematical theory of infectious diseases and its applications*. High Wycombe, UK: Charles Griffin & Company Ltd.

26. Simpson REH. 1952 Infectiousness of communicable diseases in the household. *Lancet* **2**, 549–554. (doi:10.1016/S0140-6736(52)91357-3)

27. Wearing HJ, Rohani P, Keeling MJ. 2005 Appropriate models for the management of infectious diseases. *PLoS Med.* **2**, e174. (doi:10.1371/journal.pmed.0020174)

28. Riley S *et al.* 2003 Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science* **300**, 1961–1966. (doi:10.1126/science.1086478)

29. Li X, Cao L, Cao GF. 2010 Epidemic prevalence on random mobile dynamical networks: individual heterogeneity and correlation. *Eur. Phys. J. B* **75**, 319–326. (doi:10.1140/epjb/e2010-00090-9)

30. Zhang H-F, Xie J-R, Chen H-S, Liu C, Small M. 2016 Impact of asymptomatic infection on coupled disease-behavior dynamics in complex networks. *EPL (Europhysics Letters)* **114**, 38004. (doi:10.1209/0295-5075/114/38004)

31. Hardie R-A, Luo M, Bruneau B, Knight E, Nagelkerke NJD, Kimani J, Wachihi C, Ngugi EN, Plummer FA. 2008 Human leukocyte antigen-DQ alleles and haplotypes and their associations with resistance and susceptibility to HIV-1 infection. *AIDS* **22**, 807–816. (doi:10.1097/QAD.0b013e3282f51b71)

32. Hethcote HW, Tudor DW. 1980 Integral equation models for endemic infectious diseases. *J. Math. Biol.* **9**, 37–47. (doi:10.1007/BF00276034)

33. Anderson D, Watson R. 1980 On the spread of a disease with gamma distributed latent and infectious periods. *Biometrika* **67**, 191–198. (doi:10.1093/biomet/67.1.191)

34. Lloyd AL, May RM. 1996 Spatial heterogeneity in epidemic models. *J. Theor. Biol.* **179**, 1–11. (doi:10.1006/jtbi.1996.0042)

35. Anderson H, Britton T. 2000 Stochastic epidemics in dynamic populations: quasi-stationarity and extinction. *J. Math. Biol.* **41**, 559–580. (doi:10.1007/s002850000060)

36. Cox DR, Miller HD. 2017 *The theory of stochastic processes*. Boca Raton, FL: CRC Press.

37. Krylova O, Earn DJ. 2013 Effects of the infectious period distribution on predicted transitions in childhood disease dynamics. *J. R. Soc. Interface* **10**, 20130098. (doi:10.1098/rsif.2013.0098)

38. Karrer B, Newman ME. 2010 Message passing approach for general epidemic models. *Phys.*

Rev. E **82**, 016101. (doi:10.1103/PhysRevE.82.016101)

39. Sherborne N, Miller JC, Blyuss KB, Kiss IZ. 2018 Mean-field models for non-Markovian epidemics on networks. *J. Math. Biol.* **76**, 755–778. (doi:10.1007/s00285-017-1155-0)

40. Gou W, Jin Z. 2017 How heterogeneous susceptibility and recovery rates affect the spread of epidemics on networks. *Infect. Dis. Model.* **2**, 353–367. (doi:10.1016/j.idm.2017.07.001)

41. Valdano E, Ferreri L, Poletto C, Colizza V. 2015 Analytical computation of the epidemic threshold on temporal networks. *Phys. Rev. X* **5**, 021005. (doi:10.1103/physrevx.5.021005)

42. Valdano E, Poletto C, Colizza V. 2015 Infection propagator approach to compute epidemic thresholds on temporal networks: impact of immunity and of limited temporal resolution. *Eur. Phys. J. B* **88**, 341. (doi:10.1140/epjb/e2015-60620-5)

43. Brent RP. 1973 *Algorithms for minimization without derivatives*. Englewood Cliffs, NJ: Prentice-Hall.

44. Valdano E, Ferreri L, Poletto C, Colizza V. 2015 GitHub: computing the epidemic threshold on temporal networks. See http://github.com/eugenio-valdano/threshold.

45. Holme P, Saramäki J. 2013 *Temporal Networks*. Berlin, Germany: Springer.

46. Hellinger E. 1909 Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen. *Journal für die reine und angewandte Mathematik* **136**, 210–271. (doi:10.1515/crll.1909.136.210)

47. Tomley FM, Shirley MW. 2009 Livestock infectious diseases and zoonoses. *Phil. Trans. R. Soc. B* **364**, 2637–2642. (doi:10.1098/rstb.2009.0133)

48. Humblet M-F, Boschiroli ML, Saegerman C. 2009 Classification of worldwide bovine tuberculosis risk factors in cattle: a stratified approach. *Vet. Res.* **40**, 1–24. (doi:10.1051/vetres/2009033)

49. Etter E, Donado P, Jori F, Caron A, Goutard F, Roger F. 2006 Risk analysis and bovine tuberculosis, a re-emerging zoonosis. *Ann. NY Acad. Sci.* **1081**, 61–73. (doi:10.1196/annals.1373.006)

50. Italian Veterinary Epidemiological Bulletin. 2018 Officially Free Territories. See http://www.izs.it/BENV.NEW/Engine/RAServePG.php/P/251210040610/T/Territori-ufficialmente-indenni.

51. Italian Ministry of Health. 2012 Ordinanza Ministero della Salute 9 agosto 2012: 'Misure straordinarie di polizia veterinaria in materia di tubercolosi, brucellosi bovina e bufalina, brucellosi ovi-caprina, leucosi'.

52. Italian Ministry of Health. 2018 Sistema Informativo Veterinario. See https://www.vetinfo.sanita.it/sso_portale/accesso_home.pl.

53. Dubé C, Ribble C, Kelton D, McNab B. 2009 A review of network analysis terminology and its application to foot-and-mouth disease modelling and policy development. *Transbound. Emerg. Dis.* **56**, 73–85. (doi:10.1111/jva.2009.56.issue-3)

54. Martínez-López B, Perez AM, Sánchez-Vizcaíno JM. 2009 Social network analysis: review of general concepts and use in preventive

veterinary medicine. *Transbound. Emerg. Dis.* **56**, 109–120. (doi:10.1111/jva.2009.56.issue-4)

55. Natale F, Giovannini A, Savini L, Palma D, Possenti L, Fiore G, Calistri P. 2009 Network analysis of Italian cattle trade patterns and evaluation of risks for potential disease spread. *Prev. Vet. Med.* **92**, 341–350. (doi:10.1016/j.prevetmed.2009.08.026)

56. Valdano E, Poletto C, Giovannini A, Palma D, Savini L, Colizza V. 2015 Predicting epidemic risk from past temporal contact data. *PLoS Comput. Biol.* **11**, e1004152. (doi:10.1371/journal.pcbi.1004152)

57. Alvarez J et al. 2012 Eradication of bovine tuberculosis at a herd-level in Madrid, Spain: study of within-herd transmission dynamics over a 12-year period. *BMC Vet. Res.* **8**, 100. (doi:10.1186/1746-6148-8-100)

58. World Health Organization. 2011 *World Health Organization 2011 report*. Geneva, Switzerland: World Health Organization.

59. Kluytmans J, van Belkum A, Verbrugh H. 1997 Nasal carriage of *Staphylococcus aureus*: epidemiology, underlying mechanisms, and associated risks.. *Clin. Microbiol. Rev.* **10**, 505–520. (doi:10.1128/CMR.10.3.505)

60. Marschall J, Mühlemann K. 2006 Duration of methicillin-resistant *Staphylococcus aureus* carriage, according to risk factors for acquisition. *Infect. Control Hosp. Epidemiol.* **27**, 1206–1212. (doi:10.1086/507917)

61. Temime L, Opatowski L, Pannet Y, Brun-Buisson C, Boëlle PY, Guillemot D. 2009 Peripatetic health-care workers as potential superspreaders. *Proc. Natl Acad. Sci. USA* **106**, 18 420–18 425. (doi:10.1073/pnas.0900974106)

62. Hornbeck T, Naylor D, Segre AM, Thomas G, Herman T, Polgreen PM. 2012 Using sensor networks to study the effect of peripatetic healthcare workers on the spread of hospital-associated infections. *J. Infect. Dis.* **206**, 1549–1557. (doi:10.1093/infdis/jis542)

63. Obadia T et al. 2015 Detailed contact data and the dissemination of *Staphylococcus aureus* in hospitals. *PLoS Comput. Biol.* **11**, e1004170. (doi:10.1371/journal.pcbi.1004170)

64. 2008 *SocioPatterns project*. See www.sociopatterns.org.

65. Vanhems P, Barrat A, Cattuto C, Pinton J-F, Khanafer N, Régis C, Kim B-A, Comte B, Voirin N. 2013 Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLoS ONE* **8**, e73970. (doi:10.1371/journal.pone.0073970)

66. Boyce JM, Pittet D, 2002 Guideline for hand hygiene in health-care settings: recommendations of the healthcare infection control practices advisory committee and the HICPAC/SHEA/APIC/IDSA hand hygiene task force. *Infect. Control Hosp. Epidemiol.* **23**, 3–40. (doi:10.1086/503164)

67. Blok HEM, Troelstra A, Kamp-Hopmans TEM, Gigengack-Baars ACM, Vandenbroucke-Grauls CMJE, Weersink AJL, Verhoef J, Mascini EM. 2003 Role of healthcare workers in outbreaks of methicillin-resistant *Staphylococcus aureus*: a 10-year evaluation from a Dutch University Hospital. *Infect. Control Hosp. Epidemiol.* **24**, 679–685. (doi:10.1086/502275)

68. Pittet D, Allegranzi B, Sax H, Dharan S, Pessoa-Silva CL, Donaldson L, Boyce JM, 2006 Evidence-based model for hand transmission during patient care and the role of improved practices. *Lancet Infect. Dis.* **6**, 641–652. (doi:10.1016/S1473-3099(06)70600-4)

69. ECDC and WHO. 2018 Europe weekly influenza update. See http://flunewseurope.org.

70. World Health Organization. 2009 Pandemic H1N1 2009, WHO Regional Office for South-East Asia technical report. See http://www.who.int/iris/handle/10665/205605.

71. Lipsitch M, Riley S, Cauchemez S, Ghani AC, Ferguson NM. 2009 Managing and reducing uncertainty in an emerging influenza pandemic. *New England Journal of Medicine* **361**, 112–115. (doi:10.1056/NEJMp0904380)

72. Gemmetto V, Barrat A, Cattuto C. 2014 Mitigation of infectious disease at school: targeted class closure vs school closure. *BMC Infect. Dis.* **14**, 695. (doi:10.1186/s12879-014-0695-9)

73. Stehlé J et al. 2011 High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS ONE* **6**, e23176. (doi:10.1371/journal.pone.0023176)

74. Longini IM, Koopman JS, Monto AS, Fox JP. 1982 Estimating household and community

transmission parameters for influenza. *Am. J. Epidemiol.* **115**, 736–751. (doi:10.1093/oxfordjournals.aje.a113356)

75. Viboud C, Böelle P-Y, Cauchemez S, Lavenu A, Valleron A-J, Flahault A, Carrat F. 2004 Risk factors of influenza transmission in households. *Br. J. Gen. Pract.* **54**, 684–689.

76. Tuite AR et al. 2010 Estimated epidemiologic parameters and morbidity associated with pandemic H1N1 influenza. *CMAJ* **182**, 131–136. (doi:10.1503/cmaj.091807)

77. Hsieh Y-H, Tsai C-A, Lin C-Y, Chen J-H, King C-C, Chao D-Y, Cheng K-F, 2014 Asymptomatic ratio for seasonal H1N1 influenza infection among schoolchildren in Taiwan. *BMC Infect. Dis.* **14**, 80. (doi:10.1186/1471-2334-14-80)

78. Keeling MJ, Grenfell BT. 1997 Disease extinction and community size: modeling the persistence of measles. *Science* **275**, 65–67. (doi:10.1126/science.275.5296.65)

79. Darbon A, Valdano E, Poletto C, Giovannini A, Savini L, Candeloro L, Colizza V. 2018 Network-based assessment of the vulnerability of Italian regions to bovine brucellosis. *Prev. Vet. Med.* **158**, 25–34. (doi:10.1016/j.prevetmed.2018.07.004)

80. Isella L et al. 2011 Close encounters in a pediatric ward: measuring face-to-face proximity and mixing patterns with wearable sensors. *PLoS ONE* **6**, e17144. (doi:10.1371/journal.pone.0017144)

81. Polgreen PM, Tassier TL, Pemmaraju SV, Segre AM. 2010 Prioritizing healthcare worker vaccinations on the basis of social network analysis. *Infect. Control Hosp. Epidemiol.* **31**, 893–900. (doi:10.1086/655466)

82. Bernard H, Fischer R, Mikolajczyk RT, Kretzschmar M, Wildner M. 2009 Nurses' contacts and potential for infectious disease transmission. *Emerg. Infect. Dis.* **15**, 1438–1444. (doi:10.3201/eid1509.081475)

83. Lehtinen S, Blanquart F, Croucher NJ, Turner P, Lipsitch M, Fraser C. 2017 Evolution of antibiotic resistance is linked to any genetic mechanism affecting bacterial duration of carriage. *Proc. Natl Acad. Sci. USA* **114**, 1075–1080. (doi:10.1073/pnas.1617849114)