Journal of Theoretical Biology 367 (2015) 203–221

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

# Global epidemic invasion thresholds in directed cattle subpopulation networks having source, sink, and transit nodes

Phillip Schumm [a,*], Caterina Scoglio [b], Qian Zhang [c], Duygu Balcan [d,1]

[a] USDA-ARS, Center for Grain and Animal Health Research, Arthropod Borne Animal Diseases Research Unit, 1515 College Avenue, Manhattan, KS 66502, USA
[b] Kansas State University, Electrical and Computer Engineering Department, Epicenter, USA
[c] Northeastern University, College of Computer and Information Science, Laboratory for the Modeling of Biological and Socio-technical Systems, USA
[d] Istanbul Technical University, Department of Physics, Faculty of Sciences and Letters, Maslak 34469, Istanbul, Turkey

## HIGHLIGHTS

- We model demographics and disease in cattle systems using directed networks.
- We uncover a pair of global epidemic invasion thresholds for directed networks.
- We determine the critical movement rates of the invasion thresholds.
- The thresholds' ordering depends on the network composition and node degrees.
- A larger fraction of transitable nodes increase the risk of epidemic outbreaks.

## ARTICLE INFO

*Article history:*
Received 20 December 2013
Received in revised form
4 November 2014
Accepted 8 December 2014
Available online 16 December 2014

*Keywords:*
Simulation
Susceptible
Infected
Recovered

## ABSTRACT

Through the characterization of a metapopulation cattle disease model on a directed network having source, transit, and sink nodes, we derive two global epidemic invasion thresholds. The first threshold defines the conditions necessary for an epidemic to successfully spread at the global scale. The second threshold defines the criteria that permit an epidemic to move out of the giant strongly connected component and to invade the populations of the sink nodes. As each sink node represents a final waypoint for cattle before slaughter, the existence of an epidemic among the sink nodes is a serious threat to food security. We find that the relationship between these two thresholds depends on the relative proportions of transit and sink nodes in the system and the distributions of the in-degrees of both node types. These analytic results are verified through numerical realizations of the metapopulation cattle model.

Published by Elsevier Ltd.

## 1. Introduction

Today's computing and technological resources have enabled the compilation of large-scale reliable data sets. Computational and network sciences have provided substantial advances in many fields through these data (Vespignani, 2012; Barrat et al., 2008; Castellano et al., 2009; Schweitzer et al., 2009; Battiston et al., 2012). Not the least among these fields is the study of disease spread through human populations (Lloyd and May, 2001; Pastor-Satorras and Vespignani, 2001). The spatial characterization of a spreading disease can be captured through a division of the host population into relatively distinct and discrete patches of subpopulations (Taylor et al., 2010; Parikh et al., 2013). The disease spread then occurs within subpopulations possessing infectious individuals, and the movements of individuals between pairs of locations allow the process to spread through the system (Hanski and Gilpin, 1997; Tilman and Kareiva, 1997; Bascompte and Solé, 1998; Hanski and Gaggiotti, 2004). For spatially structured populations of well-defined social groups interfaced through individuals' movements, epidemics have been successfully characterized using metapopulation dynamics (Anderson and May, 1984; May and Anderson, 1984, 1979; Bolker et al., 1993; Bolker and Grenfell, 1995; Sattenspiel and Dietz, 1995; Lloyd and May, 1996; Keeling and Rohani, 2002; Hethcote, 1978; Grenfell and Harwood, 1997;

* Corresponding author.
*E-mail address:* phillip.schumm@ars.usda.gov (P. Schumm).

[1] Dr. Duygu Balcan has passed away during the completion of this work. Dr. Balcan has co-designed the study and significantly contributed to the solutions of the quasi-equilibrium population levels. Furthermore, she has closely mentored Q.Z. and P.S. through this project. P.S. and C.S. have finalized the derivations, experiments, and manuscripts and are fully responsible for the content.

http://dx.doi.org/10.1016/j.jtbi.2014.12.007
0022-5193/Published by Elsevier Ltd.

Grenfell and Bolker, 1998; Ferguson et al., 2003; Riley, 2007). Mobility dynamics have been used to study systems coupled by both individual movements and aggregations of movements as effective forces of infection between subpopulations (Bolker and Grenfell, 1995; Lloyd and May, 1996; Keeling and Rohani, 2002; Rvachev and Longini, 1985; Grais et al., 2003; Longini, 1998; Flahault and Valleron, 1992; Ruan et al., 2006; Danon et al., 2011; Earn et al., 1998; Rohani et al., 1999; Keeling, 2000; Park et al., 2002; Vazquez, 2007). Recently, data-driven computational models have employed such metapopulation approaches to describe the large-scale geographical spread of infectious diseases (Riley, 2007; Grais et al., 2004; Cooper et al., 2006; Hollingsworth et al., 2006; Hufnagel et al., 2004; Colizza et al., 2007, 2006).

The study of metapopulation systems has been adapted to complex networks where each node hosts a subpopulation of individuals. These individuals flow between nodes through mobility processes. This has led to the exploration of contagion processes taking place on complex network structures with complex mobility models (Colizza et al., 2007; Colizza and Vespignani, 2007; Balcan and Vespignani, 2011; Poletto et al., 2012; Gonçalves et al., 2013). A global epidemic invasion threshold has been defined that describes whether a contagion process will spread to a notable fraction of the other subpopulations or will die out in a finite amount of time in the thermodynamic limit (Colizza and Vespignani, 2007, 2008; Balcan and Vespignani, 2011, 2012; Poletto et al., 2012, 2013; Ball et al., 1997; Cross et al., 2005, 2007; Liu et al., 2013). This threshold parameter can be determined from the characteristics of the contagion process, the mobility model, and the subpopulation network structure.

A successful metapopulation approach is based on detailed knowledge of the disease, the spatial network structure, and the movement of individuals (Chowell et al., 2003; Barrat et al., 2004; Guimerà et al., 2005; Brockmann et al., 2006; Patuelli et al., 2007; González et al., 2008; Balcan et al., 2009; Vespignani, 2009; Liu et al., 2014). In the study of human infectious disease outbreaks, the mobility arcs of the subpopulation network can and have been considered undirected and symmetrical (Colizza et al., 2006; Balcan et al., 2009). When livestock systems are modeled in subpopulation networks, very different processes control the movements of individual cattle within the livestock industry. Recent collection, compilation, and analyses of detailed data describing individual livestock movements have enabled the construction of livestock movement networks in various European countries (Kao et al., 2006; Rautureau et al., 2011; Bigras-Poulin et al., 2006, 2007; Robinson and Christley, 2007; Baptista et al., 2008; Natale et al., 2009). These studies have found that the arcs of livestock networks are highly unidirectional due to specializations of subpopulations (dairy farms, markets, slaughterhouses, and more) and that an animal rarely returns to its origin (Buhnerkempe et al., 2013). From livestock movement databases, data-driven computational models of geographical spread of livestock disease have increased in popularity in the literature (Kao et al., 2007; Green et al., 2006; Bajardi et al., 2012; Xue et al., 2013). The conventional models have implemented mass-action laws and distance-based kernels and have not been driven by the movement databases (Ferguson et al., 2001; Keeling et al., 2001; Hsu et al., 2011; Xue et al., 2012). Apart from some computational models using mechanistic approaches in the metapopulation structure, these models consider livestock premises as their basic units and miss the resolution level of individual animals (Keeling et al., 2010; Lentz et al., 2012).

In this paper, we design a metapopulation network model and investigate the existence of the novel global epidemic invasion threshold for directed networks. We illustrate these directed system results on a basic livestock industry model. Our model system includes directed cattle movements leading to a natural division of source, transit, and sink subpopulations. Within cattle systems, source nodes may represent cow-calf or dairy farms that rarely purchase cattle, transit nodes might represent grazers or backgrounders, and sink nodes can represent large feed/finishing operations (final waypoints prior to slaughter). In Section 2, our metapopulation livestock system is defined on directed networks connected through cattle movements. Death, birth, and importation processes are described within each subpopulation. With the use of degree-block variables, we find the quasi-equilibrium population levels across the network that result from this combination of population flows (Balcan and Vespignani, 2012). In Section 3, we add a susceptible-infected-recovered model within each subpopulation and determine the directed system version of the undirected global epidemic invasion threshold $R_*$ (Colizza and Vespignani, 2007; Balcan and Vespignani, 2012). This threshold defines whether a disease can be expected to spread through the system or to quickly die out. Beyond the global invasion threshold, for directed metapopulation systems, we discover a second invasion threshold that defines whether a disease outbreak is likely to circulate within transit subpopulations or to spill over into the sink subpopulations where it may threaten food safety. We utilize computational resources to extensively simulate this model, demonstrating the analytical observations through numerical exploration in Section 4. Section 5 summarizes this paper with conclusions and discussions.

## 2. Metapopulation model of livestock industry

We describe a generic livestock system through a directed subpopulation network. Let each node $i$ represent a premises or a small, well-mixed region with a homogeneous population of cattle $N_i(t)$. Cattle flow between nodes on directed arcs, with each arc described by its origin node $i$ and its destination node $j$. In directed networks, each node may have a set of arcs coming into the node from an in-neighborhood and a set of arcs leaving the node to its out-neighborhood. The number of arcs arriving to node $i$ is the in-degree of node $i$ and is represented by $k_i^{in}$. Likewise the number of arcs departing node $i$ is the out-degree $k_i^{out}$. Each node $i$ has a pair of node degrees which we represent by a degree vector $\overrightarrow{k_i} \equiv (k_i^{in}, k_i^{out})$. We classify each of $V$ nodes based on their respective node degrees. For every node $i$ having $k_i^{in} = 0$, we refer to node $i$ as a "source" node. "Sink" nodes are all nodes such that $k_i^{in} \neq 0$ and $k_i^{out} = 0$. We describe the remaining nodes as "transit" nodes. Cattle are permitted to enter and to leave each transit node. We denote the fractions of sources and sinks in the network respectively by $\eta_{out}$ and $\eta_{in}$.

The life cycle of our cattle begins as they are either born or imported into the set of source nodes. After arriving at a sink node, cattle may terminate their life cycle when they move off to a slaughter house at a rate of $\delta$ per individual. The journey of each individual from its source (or the node at which the individual is initiated) can lead to either a sink node or perpetual wandering among the transit nodes. These movements occur inward to each transit or sink node $j$ from its in-neighborhood and outward from each source or transit node $i$ to its out-neighborhood. For each source or transit node $i$, each individual moves out on one of the $k_i^{out}$ arcs departing the node with a uniform movement rate:

$$d_{ij} = \frac{p}{k_i^{out}}, \tag{1}$$

where $p$ represents the total per capita diffusion rate from all nodes having $k_i^{out} \neq 0$. Let $\overline{N}$ be the system average number of cattle per node. We maintain the total system population of

cattle $\overline{N}V$ by recycling the slaughtered cattle back into the system through the birth and importation processes as new susceptible cattle in the source nodes. The fraction of recycled cattle that is selected for birth rather than importation is $p_\beta$. Further details are available in Appendix C. The first step in characterizing these three classes of nodes (source, transit, and sink) is to determine the equilibrium distributions of their populations. We explore a mean-field approximation to the livestock system through assuming the statistical equivalence of each group of nodes (and their subpopulations) that share the same joint-degree vector $\vec{k}_i$. This allows us to introduce "degree-block" variables which are indexed by $\vec{k}_i$ rather than individual node $i$. Such an approximation is not novel and has been successfully demonstrated in several dynamical systems (Colizza et al., 2007; Colizza and Vespignani, 2007, 2008; Balcan and Vespignani, 2011, 2012). The degree-block variables are then defined from the node variables. Statistical equivalence is assumed for every node within the same degree-block, that is, sharing the same degree $\vec{k}_i$. The average population size of nodes at time $t$ in degree-block $\vec{k}_i$ is then

$$N_{\vec{k}}(t) \equiv \frac{1}{V_{\vec{k}}} \sum_{i \mid \vec{k}_i = \vec{k}} N_i(t), \tag{2}$$

where $V_{\vec{k}}$ is the number of nodes having degree $\vec{k}$. The arc-based variables can also be defined between degree-blocks rather than between individuals. The average diffusion rate on arcs from nodes with degree $\vec{k}$ to nodes with degree $\vec{k}'$ is defined to be

$$d_{\vec{k}\,\vec{k}'} \equiv \frac{1}{E_{\vec{k}\,\vec{k}'}} \sum_{i \mid \vec{k}_i = \vec{k}} \sum_{j \mid \vec{k}_j = \vec{k}'} d_{ij} = \frac{p}{k_i^{out}} \quad \forall k_i^{out} \neq 0, \tag{3}$$

where $E_{\vec{k}\,\vec{k}'}$ is the number of arcs originating from nodes with degree $\vec{k}$ and terminating at nodes with degree $\vec{k}'$. Since the movement as described in Eq. (1) depends only on the degree, the average degree-block expression for movement is equivalent. We present a detailed discussion and further derivations regarding these degree-block variables in Appendix A. Next we derive the average population sizes of nodes in each class by degree $\vec{k}$ at the quasi-equilibrium of the collective process of cattle movement, birth, death, and importation. Let $P_a$ be the pairwise node degree distribution, specifically, let $P_a(\vec{j}, \vec{k})$ be the probability of finding an arc with a degree of $\vec{j}$ at its origin and $\vec{k}$ at its destination. The rate equation of average population size $N_{\vec{k}}(t)$ in a subpopulation with joint-degree $\vec{k}$ is given by

$$\begin{aligned} \partial_t N_{\vec{k}}(t) = & -\delta_{\vec{k}} N_{\vec{k}}(t) \delta_{k^{out},0} - d_{\vec{k}} N_{\vec{k}}(t) \left(1 - \delta_{k^{out},0}\right) \\ & + \left[\beta_{\vec{k}}(t) N_{\vec{k}}(t) + \epsilon_{\vec{k}}(t)\right] \delta_{k^{in},0} \\ & + k^{in} \sum_{\vec{j}} P_a\left(\vec{j}, \vec{k} \mid \vec{k}\right) d_{\vec{j}\,\vec{k}} N_{\vec{j}}(t) \left(1 - \delta_{j^{out},0}\right). \end{aligned} \tag{4}$$

The birth rate $\beta_{\vec{k}}(t)$ and the importation rate $\epsilon_{\vec{k}}(t)$ are only observed in the source nodes. Likewise the death rate $\delta_{\vec{k}}$ is only observed in the sink nodes and is identical across all in-degrees. The compositions of $\beta_{\vec{k}}(t)$ and $\epsilon_{\vec{k}}(t)$ are found in Appendix A. Finding the equilibria of Eq. (37) for each class of node is our desired objective. This can be accomplished by determining the solutions of Eq. (37) that cause $\partial_t N_{\vec{k}}(t)$ to be equal to zero. Let us introduce a notation to distinguish the degree vectors of each class of nodes. We denote the degrees with the source class, transit class, and sink class as $\vec{k}^{(1)}$ $\left(k_i^{in} = 0 \ \forall i\right)$, $\vec{k}^{(2)}$ $\left(k_i^{in} \neq 0 \text{ and } k_i^{out} \neq 0 \ \forall i\right)$, and $\vec{k}^{(3)}$ $\left(k_i^{in} \neq 0 \text{ and } k_i^{out} = 0 \ \forall i\right)$, respectively. We examine each of the three node classes separately and derive the equilibrium distribution for each. For $0 < p_\beta < 1$, we find the equilibrium configuration of the source class of nodes to be

$$N^*_{\vec{k}^{(1)}}(t) = \frac{k^{out(1)}}{\left\langle k^{out} \right\rangle} \frac{\eta_{in}(1-\eta_{in})\delta}{\eta_{out}\left[(1-\eta_{out})\delta + \eta_{in} p\right]} \overline{N}, \tag{5}$$

where $\left\langle k^{out} \right\rangle$ is the average out-degree of the node in the network. The quasi-equilibrium populations of the source nodes are distributed proportional to their respective out-degrees. This dependence arises from the assumption that the importations to source nodes depend on the source nodes' out-degrees as described in Appendix A. The transit and sink nodes likewise depend primarily on their node degrees, but for these two classes, it is the in-degrees that determine their equilibrium population distributions. Respectively, for transit nodes and sink nodes we find

$$N^*_{\vec{k}^{(2)}}(t) = \frac{k^{in(2)}}{\left\langle k^{out} \right\rangle} \frac{(1-\eta_{out})\delta}{\left[(1-\eta_{out})\delta + \eta_{in} p\right]} \overline{N} \quad \text{and} \tag{6}$$

$$N^*_{\vec{k}^{(3)}}(t) = \frac{k^{in(3)}}{\left\langle k^{out} \right\rangle} \frac{(1-\eta_{out})p}{\left[(1-\eta_{out})\delta + \eta_{in} p\right]} \overline{N}. \tag{7}$$

This dependence of the equilibrium populations on the in-degrees of each node is reminiscent of a directed random walk process and implies a stronger dependence of the equilibrium configurations on the cattle movement process rather than the cattle recycling processes. The derivations of these results are included in Appendix A as well as their characteristic relaxation times and the boundary situations of $p_\beta = 0$ and $p_\beta = 1$. It is worth remembering that these derivations depend on the assumption of no correlations between the degrees of connected nodes. In the following section, we explore a disease process on top of these mobility dynamics, assuming that the system has reached the quasi-equilibrium configurations before the introduction of the disease.

## 3. Global epidemic invasion thresholds of livestock model

Let us consider the addition of disease to our livestock model for the purpose of exploring the global invasion threshold of epidemics in the livestock system (Colizza and Vespignani, 2007, 2008; Balcan and Vespignani, 2012). Following the classic Susceptible–Infected–Recovered (SIR) disease model, for each node $i$ at time $t$, we define the number of susceptible, infected, and recovered individuals respectively as $S_i(t)$, $I_i(t)$, and $R_i(t)$ (Keeling and Rohani, 2011). We assume that the initial livestock populations are fully susceptible and that, within each node, the subpopulations are well-mixed. Inside each node, each infected individual transmits infection to all susceptible individuals at rate $\beta$. The model assumes a mass-action law to describe the force on infection acting on each susceptible individual that is proportional to the prevalence of infected individuals in the node. This force of infection is equivalently the rate at which each susceptible individual becomes infected. The recovery rate $\mu$ is the rate at which each infected individual recovers from the disease. Once recovered, an individual remains in the recovered state until and unless it happens to be recycled through the slaughter processes.

The SIR model can be summarized as

$$N_i(t) = S_i(t) + I_i(t) + R_i(t), \tag{8}$$

$$\frac{dS_i}{dt} = -\lambda_i(t) S_i(t), \tag{9}$$

$$\frac{dI_i}{dt} = \lambda_i(t) S_i(t) - \mu I_i(t), \tag{10}$$

$$\frac{dR_i}{dt} = \mu I_i(t), \tag{11}$$

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

where $\lambda_i(t)$ is the per capita force of infection. This SIR model can be characterized by the basic reproductive number $R_0 = \beta/\mu$ which is the average number of secondary infections generated by a typical infected individual in its first time step of being infectious (Keeling and Rohani, 2011). The reproductive number $R_0$ serves as a threshold parameter at the local population level. Only if $R_0 > 1$, then the disease can spread to a finite proportion of the population. If $R_0 < 1$, the disease will die out in a finite amount of time and only impact a minuscule fraction of the susceptible population (zero in the thermodynamic limit of $N \to \infty$). In our case, this threshold determines the growth of the disease within the subpopulation of each node. For the disease to spread beyond the infected subpopulation, it must be carried by infected cattle moving through the network. Thus a threshold to determine system-wide disease invasion must also consider cattle movement parameters.

A new global invasion threshold for metapopulation models has been introduced (Colizza and Vespignani, 2007, 2008; Balcan and Vespignani, 2011, 2012; Ball et al., 1997; Harris, 1989; Vazquez, 2006). This threshold considers the generations of infected nodes through a branching process during the initial stages of a disease outbreak. For the branching process model, the approximated growth rate of the number of infected locations is defined as $R_*$. $R_*$ is a subpopulation reproductive number which is the average number of infected subpopulations generated from a typical infected subpopulation in a fully susceptible, structured metapopulation system (Colizza and Vespignani, 2007, 2008; Balcan and Vespignani, 2011, 2012; Ball et al., 1997; Cross et al., 2005, 2007). The popular title for $R_*$ is the global epidemic invasion threshold because, if $R_* > 1$, the epidemic will impact a finite fraction of the subpopulations of the system (in the thermodynamic limit of $V \to \infty$). Similarly, if $R_* < 1$, the epidemic will reach only a minuscule fraction of nodes and will die out in finite time. $R_* = 1$ describes the system-level epidemic invasion threshold for metapopulation models.

Let us consider the characterization of the three classes of nodes in this directed system through the $R_*$ branching model. The branching model assumes that the epidemic is in its early stages with a $R_0$ value just above 1 and that the majority of the subpopulations are susceptible. Let $D^n_{\overrightarrow{k}^{(x)}}$ count the $n$th generation number of diseased subpopulations with degree $\overrightarrow{k}^{(x)}$ from node class $x$. This can be estimated as a function of the three sets $\{D_{\overrightarrow{k}^{(1)}}{}^{n-1}\}$, $\{D_{\overrightarrow{k}^{(2)}}{}^{n-1}\}$, and $\{D_{\overrightarrow{k}^{(3)}}{}^{n-1}\}$ of the $(n-1)$th generation. The "infection" of a node occurs when infected cattle move from one node into another fully susceptible node, wherein they serve as a seed for the SIR process. The expression for the branching process as expressed by Colizza and Vespignani (2008) and Balcan and Vespignani (2012) is written as

$$D^n_{\overrightarrow{k}} = \sum_{\overrightarrow{j}} D^{n-1}_{\overrightarrow{j}} P_a\left(\overrightarrow{j}, \overrightarrow{k} \mid \overrightarrow{j}\right) j^{out} p(\overrightarrow{j}, \overrightarrow{k}) \prod_{m=0}^{n-1} \left(1 - \frac{D^m_{\overrightarrow{k}}}{V_{\overrightarrow{k}}}\right). \tag{13}$$

This branching equation models the $n$th generation of newly infected nodes by considering that each subpopulation in a node with an out-degree of $j$ has the potential to infect $j$ other nodes by the intersection of three events. The three events are that a node with degree $\overrightarrow{k}$ exists in the out neighborhood of a node with degree $\overrightarrow{j}$, $P_a\left(\overrightarrow{j}, \overrightarrow{k} \mid \overrightarrow{j}\right)$, that the neighbor with degree $k$ has not been infected in previous generations, $\prod_{m=0}^{n-1}\left(1 - D^m_{\overrightarrow{k}}/V_{\overrightarrow{k}}\right)$, and that the disease will spread from the node with degree $j$ to the node with degree $k$, $p(\overrightarrow{j}, \overrightarrow{k})$.

Considering the source nodes, we note that they possess no incoming degrees and thus will not receive infections from inward moving cattle. Source nodes may contain disease among their subpopulations from an initialization of the disease within the source nodes, but the number of source nodes possessing infected cattle is a non-increasing number. Within the branching model, $D^n_{\overrightarrow{k}^{(x)}}$ is specifically the number of newly infected nodes having degree $\overrightarrow{k}^{(x)}$ and it does not count pre-existing outbreaks. The branching process estimates for the $n$th generation source nodes, $D^n_{\overrightarrow{k}^{(1)}}$, are identically equal to 0 for every generation. The source nodes do not contribute to the global epidemic invasion threshold. Incoming arcs to the transit nodes can arrive from both source nodes and transit nodes. Likewise for the sink nodes, their incoming arcs arrive from both source and transit nodes. However, as $D^n_{\overrightarrow{k}^{(1)}} = 0 \ \forall n$, only the newly infected nodes of the transit nodes $\{D_{\overrightarrow{k}^{(2)}}{}^{n}\}$ $\left(\text{and not } \{D_{\overrightarrow{k}^{(1)}}{}^{n}\}\right)$ are considered in the branching model as sources of infection. Seeing that the diseased subpopulations of the sink nodes depend only on the transit nodes and the transit nodes depend only on themselves, we proceed to derive the global epidemic invasion threshold $R_*$ for the directed metapopulation system from the system-level outbreak among the transit nodes: $D^n_{\overrightarrow{k}^{(2)}}$ as a function of the set $\{D_{\overrightarrow{k}^{(2)}}{}^{n-1}\}$. With details kept aside in Appendix B, we reach

$$D^n_{\overrightarrow{k}^{(2)}} = \frac{2p}{\mu}\left(1 - R_0^{-1}\right)^2 \sum_{\overrightarrow{j}^{(2)}} D^{n-1}_{\overrightarrow{j}^{(2)}} P_a\left(\overrightarrow{j}^{(2)}, \overrightarrow{k}^{(2)} \mid \overrightarrow{j}^{(2)}\right) N_{\overrightarrow{j}^{(2)}}, \quad n > 1. \tag{14}$$

To uncover the global invasion threshold, the point at which the branching process grows or decreases must be determined. Therefore we estimate the growth factor $R_*$ of this process such that, when the factor is equal to unity, the process is on the edge between growth and decline. With full details in Appendix B, we derive the global invasion threshold for our directed metapopulation system as

$$R_* = \frac{2p\delta(1-\eta_{out})(1-\eta_{out}-\eta_{in})\overline{N}}{\mu\left[(1-\eta_{out})\delta + \eta_{in} p\right]}\left(1 - \frac{1}{R_0}\right)^2 \frac{\left\langle \left(k^{in(2)}\right)^2 \right\rangle}{\left\langle k^{in} \right\rangle^2}. \tag{15}$$

The most notable trait of this system's $R_*$ is its dependence on the topology degree distributions, namely the moment $\left\langle k^{in} \right\rangle$ of the full network and the moment $\left\langle \left(k^{in(2)}\right)^2 \right\rangle$ of the transit nodes'

in-degree distribution. Several dynamical processes on undirected networks have demonstrated that heterogeneity in node degrees lowers the threshold value, encouraging the spread of disease (Barrat et al., 2008). For directed subpopulation networks, this heterogeneity is particularly that of the in-degrees of the transit nodes. These nodes compose the strongly connected component of the networks. Greater diversity among the in-degrees of the transit nodes produces a larger value of $R_*$. As $R_*$ increases beyond 1, the disease under consideration will break out across the metapopulation system. The restriction of the movements of individuals is a common disease control strategy. We can explore the restriction of movement by deriving the critical value $p_c$ of the movement rate $p$ such that the global epidemic invasion threshold is equal to 1. Appendix B also includes the derivation of $p_c$ with the result of

$$p_c = \frac{\mu\delta(1-\eta_{out})\left\langle k^{in}\right\rangle^2}{2\delta\overline{N}(1-\eta_{out})(1-\eta_{out}-\eta_{in})\left(1-\frac{1}{R_0}\right)^2\left\langle\left(k^{in(2)}\right)^2\right\rangle-\eta_{in}\mu\left\langle k^{in}\right\rangle^2}. \tag{16}$$

If the movement parameter $p$ of the livestock system is greater (less) than $p_c$, then the value of $R_*$ will be greater (less) than 1 and the epidemic can be expected to spread (die out). The dependence of $p_c$ on the heterogeneity of the transit nodes' in-degrees is similar to that of $R_*$. If the heterogeneity is greater, $p_c$ will be lower, increasing the opportunities for a disease to spread among the transit nodes. The consideration of the three classes of nodes (source, transit, sink) permits the existence of a second global invasion threshold. This new threshold would determine whether the disease will remain within the transit nodes or spread into the sink nodes, creating larger epidemics. Let us consider the dependence of this process on the successful arrival of infected cattle from the transit nodes. Similar to Eq. (14), we approximate the branching process from the transit to the sink nodes as

$$D^n_{\overrightarrow{k}^{(3)}} = \frac{2p}{\mu}\left(1-R_0^{-1}\right)^2 \sum_{\overrightarrow{j}^{(2)}} D^{n-1}_{\overrightarrow{j}^{(2)}} P_a\left(\overrightarrow{j}^{(2)}, \overrightarrow{k}^{(3)} \mid \overrightarrow{j}^{(2)}\right) N_{\overrightarrow{j}^{(2)}}, \quad n>1. \tag{17}$$

Eq. (17) is similar in structure to Eq. (14), with the difference being found in the conditional pairwise degree probabilities. This difference relates the growth process of this second branching process to the in-degree distribution of the sink nodes. With details in Appendix B, we derive the transit-to-sink invasion threshold $R_*^{TS}$ as

$$R_*^{TS} = \frac{2p\delta(1-\eta_{out})\eta_{in}\overline{N}}{\mu\left[(1-\eta_{out})\delta+\eta_{in}p\right]}\left(1-\frac{1}{R_0}\right)^2\frac{\left\langle\left(k^{in(3)}\right)^2\right\rangle}{\left\langle k^{in}\right\rangle^2}. \tag{18}$$

This $R_*^{TS}$ describes the tipping point for a disease to move from the primary class of disease-sustaining nodes into a second class of nodes. Similar to $R_*$, $R_*^{TS}$ depends on the heterogeneity of the in-degree distribution of a class of nodes, but that dependence is on the in-degrees of sink nodes rather than transit nodes. The amount of similarity between the in-degree distribution of the transit nodes and the in-degree distribution of the sink nodes determines the similarity of these two global invasion thresholds. We can explore this relationship further through the respective critical movement rates. The critical movement rate $p_c^{TS}$ for this threshold is given by

$$p_c^{TS} = \frac{\mu\delta(1-\eta_{out})\left\langle k^{in}\right\rangle^2}{2\delta\overline{N}(1-\eta_{out})\eta_{in}\left(1-\frac{1}{R_0}\right)^2\left\langle\left(k^{in(3)}\right)^2\right\rangle-\eta_{in}\mu\left\langle k^{in}\right\rangle^2}. \tag{19}$$

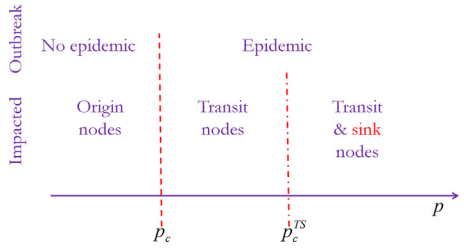


**Fig. 1.** Two transitions are observed as the cattle movement rate $p$ increases past the pair of critical movement rates, $p_c$ and $p_c^{TS}$. When the movement rate is less than $p_c$, the disease only impacts the nodes where the disease has originated (or very close by) and it is unable to spread to a significant fraction of the nodes. When the movement rate is greater than $p_c$ and less than $p_c^{TS}$, the disease will impact a significant fraction of the transit nodes with no significant impact on the sink nodes. When the movement rate is greater than both critical movement rates, the disease will spread to notable fractions of both transit and sink nodes.

The derivation of Eq. (19) is included in Appendix B. If the movement parameter $p$ of the livestock system is greater (less) than $p_c^{TS}$, then the value of $R_*^{TS}$ will be greater (less) than 1 and the epidemic can be expected to (not) spread from the transit nodes to the sink nodes. For higher levels of heterogeneity among the in-degrees of the sink nodes, $p_c^{TS}$ will be lower, increasing the opportunities for a disease to spread out of the transit nodes into the sink nodes through the movements of infected cattle.

Let us examine the relationship between $p_c^{TS}$ and $p_c$. If the thresholds described by $p_c$ and $p_c^{TS}$ are ordered such that $p_c^{TS}<p_c$, then scenarios having movement rates above $p_c$ will initiate disease outbreaks within the transit nodes as well as within the sink nodes. However, if a network's parameters are such that $p_c^{TS}>p_c$, then there will be a second transition in the size of the epidemic as the movement rate is increased. This two-transition situation is depicted in Fig. 1. Through derivations shown in Appendix B, we find that $p_c^{TS}<p_c$ when the ratio of the second moments of the two classes' in-degree distributions $\left\langle\left(k^{in(2)}\right)^2\right\rangle/\left\langle\left(k^{in(3)}\right)^2\right\rangle$ is less than the ratio of the sink nodes to the transit nodes $\eta_{in}/(1-\eta_{out}-\eta_{in})$. The reverse is also true, that is

$$\frac{\left\langle\left(k^{in(2)}\right)^2\right\rangle}{\left\langle\left(k^{in(3)}\right)^2\right\rangle} > \frac{\eta_{in}}{(1-\eta_{out}-\eta_{in})} \Rightarrow p_c^{TS}>p_c. \tag{20}$$

The reverse case shown in Eq. (20) is the more interesting of the two scenarios as it implies the existence of two sequential movement thresholds. These thresholds describe a multi-step process of disease first breaking out in the transit nodes and secondly, as the movement rate increases, spreading successfully into the sink nodes. If a network of directed cattle movements meets the criteria of Eq. (20), there exists an opportunity for an outbreak to spread freely within the strongly connected component of the network, while the sink nodes remain nearly disease-free. The results of this section depend on the assumption of statistical equivalence within each degree block and the assumption that there are no correlations between the degree vectors of connected nodes. We have considered a simplistic set of outgoing movement rates to be uniformly $p$ and time independent. Within each node, only a single species of livestock and a single, well-mixed type of livestock population have been considered. We have assumed a relatively closed population within each node in the SIR model. These assumptions limit the applicability of the model, but the global invasion thresholds may yet provide a reasonable estimate of the risk of an epidemic.

## 4. Stochastic simulation of livestock model

We conducted numerical simulations of the described livestock model to verify our analytical results. We track individual cattle through time as births, importations, disease dynamics, and movements take place. In these Monte Carlo simulations, we explore the impacts of the diffusion rate, the movement rate of cattle to slaughter, and the network topology on the resulting disease outbreaks.

### 4.1. Dynamical processes

We implemented the movement model of Section 2 to study the directed system of cattle movements. Synthetic livestock networks were generated using a modified uncorrelated configuration model (Zhou et al., 2010). The populations of each node were initiated through a multinomial distribution, allotting $\overline{N}V$ animals across $V$ nodes, with probabilities proportional to the equilibrium populations described in Eqs. (5)–(7). We divide time into discrete intervals of length $\Delta t$. Within nodes having a non-zero out-degree, individual animals move to neighboring populations at the rate $d_{ij}$ of Eq. (1). The probability that an individual in location $i$ moves to location $j$ within its out-neighborhood within $\Delta t$ is $d_{ij}\Delta t$. During each time step, an individual within a source node will replicate itself with probability $\beta_k(t)\Delta t$. Also within the source nodes, individuals arrive through importation with a probability of $\epsilon_k(t)\Delta t$. Each animal in a sink node leaves for a slaughter house with probability $\delta\Delta t$ within a time step.

On top of this system of movement, birth, and death, we simulate a Susceptible–Infected–Recovered (SIR) disease model. We assume independence between the demographic processes and the disease process. Each subpopulation is divided into three states, assigning each individual to one of $S_i(t)$, $I_i(t)$, or $R_i(t)$. Susceptible (S) individuals are infected by infected (I) animals that share their same node at the same time step. Assuming a homogeneous mixing of the subpopulation of node $i$, we compute the probability of a susceptible (S) individual transitioning to the infected (I) state as

$$\lambda_i(t)\Delta t = \beta\frac{I_i(t)}{N_i(t)}\Delta t, \tag{21}$$

where $\beta$ is the transmission rate of infection. The prevalence of infectious cattle in the subpopulation is captured as $I_i(t)/N_i(t)$. The infected (I) individual recovers with probability $\mu\Delta t$ and remains in the recovered (R) state until it passes out of a sink node on its way to a slaughter house. The new cattle arriving through the birth and importation processes are added to the susceptible (S) portion of their node's subpopulation. This fresh flow of susceptible individuals permits the existence of endemic outbreaks. We demonstrate and discuss these in Section 4.2.

### 4.2. Numerical results

The directed subpopulation network structure has been designed with $V = 10^4$ nodes, having fractions of sources, transits, and sinks of 0.45, 0.45, and 0.10, respectively. This allotment of nodes to classes is similar to the observed distribution in the Italian network of cattle premises (Bajardi et al., 2011). Both the in- and out-degree distributions (denoted by $P_{v-o}(k)$ and $P_{v-i}(k)$ in Appendix A) follow a power-law distribution. The exponent of the in-degree distribution is $-2.1$ and all degrees are no greater than $k_{max} = 100$. The maximum degree is set to allow uncorrelated interconnections of the nodes, then the connections are established according a directed uncorrelated configuration model (Zhou et al., 2010). As the network structures are generated from a stochastic model, we conduct the following experiments on 10 realizations of the network configuration. More than 500 transit nodes consistently compose the giant strongly connected component of the network, while the sources and sinks rely on the transit nodes for connectivity. We consider an average population size of $\overline{N} = 10^5$ and thus a system population of $V\overline{N} = 10^9$ head. We study the livestock system with a time step of $\Delta t = 1$ day. These simulations are designed to characterize the global epidemic invasion thresholds through their dependence on the movement rate $p$ of the sources and sinks as well as their relative independence from the slaughter rate $\delta$ of the sink nodes. For all simulations, we have fixed the birth re-introduction fraction at $p_\beta = 0.8732$ and the disease parameters, assigning the basic reproduction number to $R_0 = 1.20$ and the infectious period to $\mu^{-1} = 7$ days. The infection rate follows as $\beta = \mu R_0$.

In each simulation, we initiate the populations through a multinomial distribution with the probabilities computed according to Eqs. (5)–(7). We initiate the disease with $I_i(0) = 10$ infected individuals within a subpopulation of a single node $i$ chosen at random from the nodes having minimal out-degree in the giant strongly connected component of the network. This limits the choice of the seed subpopulation to a transit node, and thus there is no initial infection in the source nodes or the sink nodes. The nature of the population recycling processes allows a potential endemic state to occur in the system. Therefore we let the disease progress in the metapopulation system until it either dies out or satisfies our criteria as endemic. We classify an epidemic simulation as one having a duration of 100 simulation years. If a trial reaches that length, we stop the trial and collect the data at the 100-year mark. In the results we present here, all realizations producing a successful outbreak in the initially seeded subpopulation have been included in our results. A successful outbreak is considered as an outbreak resulting in at least 1% of the node's population contracting the disease. We conduct simulations of each set of parameters until we collect 5000 successful outbreaks for that set, 500 per each realization of the livestock network.

We vary the movement rate $p$ and describe the resulting outbreaks through four variables: the duration of the epidemic, the number of total cases (global attack rate), the number of subpopulations having a secondary case (epidemic size), and the fraction of outbreaks classified as endemic. Fig. 2 demonstrates the outbreaks that take place as the movement rate $p$ is increased. The two critical movement rates are plotted on each sub-figure with vertical lines marking their values. The transit-to-sink invasion threshold is higher than the within-transit invasion threshold, thus a two-part transition can be expected. The average total epidemic size shown in Fig. 2A increases smoothly, yielding no hint of the two-stage transition. When the infected sink nodes are counted explicitly, the second transition can be observed in the inverted red triangles in Fig. 2A. In Fig. 2B and C, the median attack rate and the epidemic duration increase slowly between the two critical movement rates and faster after crossing both. In Fig. 2D, we numerically examine the endemic outcomes as a fraction of the total trials. As defined by our numerical criteria, the endemic cases do not begin until the movement rate is significantly greater than both critical rates. Once the endemic cases first occur, they rise very suddenly to occur in nearly every trial. It must be noted that in a real endemic situation, the epidemic duration no longer contains much significance as a metric. The duration of an endemic is expected to be much greater than that of a non-endemic SIR outbreak. With further simulations, we explored the impact of the death rate $\delta$ on the epidemics in Fig. 3. We repeated the experiment of Fig. 2 over a spectrum of death rates. The results are depicted in a three-dimensional manner, using a spectrum of color. Although there was no visible dependence of either of the critical movement rates on $\delta$, we did observe variations in the epidemic behavior. On the logarithm scale, the epidemics remain

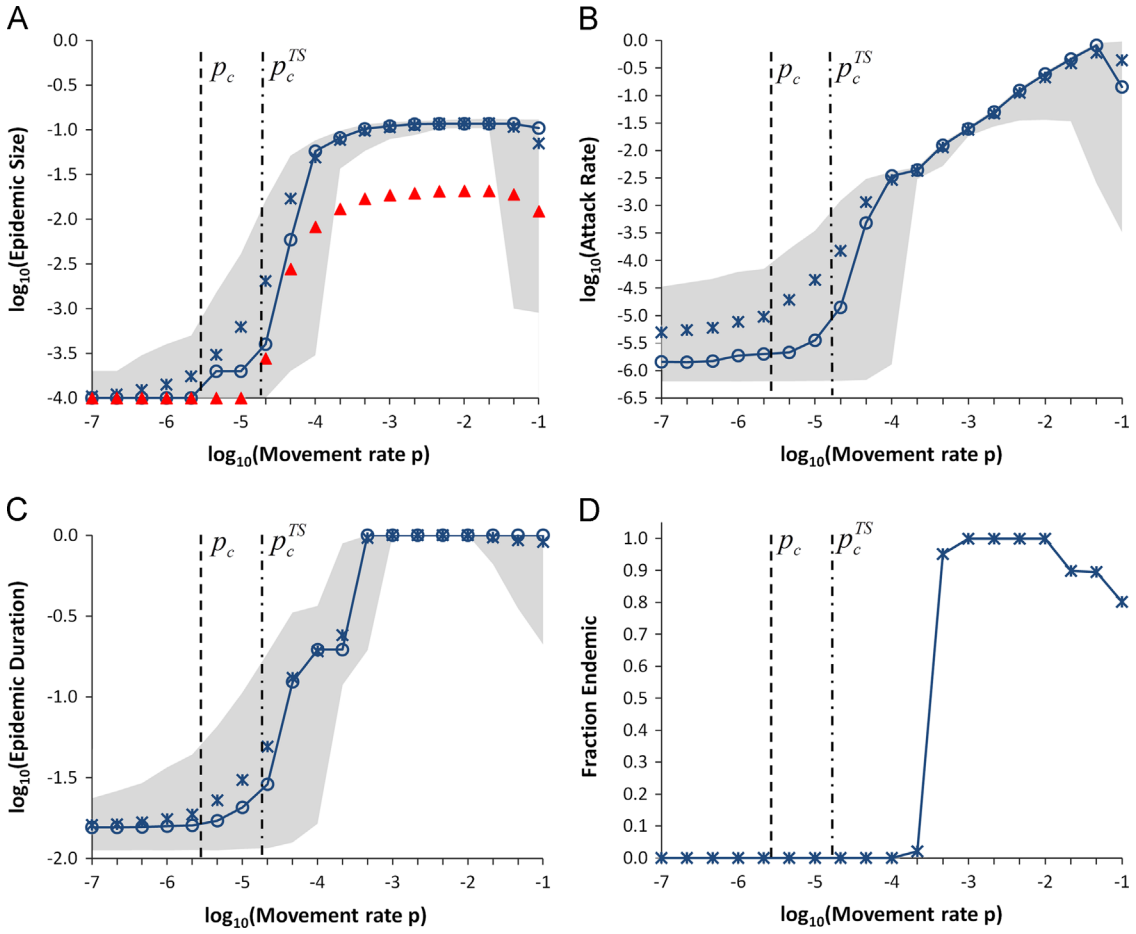


**Fig. 2.** (A) The distribution of the epidemic size as a function of the movement rate $p$ (with the sample averages denoted by stars, the sample medians denoted by circles connected on the line, and the 95% confidence interval denoted by the shaded region) rises as it passes the two invasion thresholds where $p_c$ for global invasion is marked with a vertical dashed line and $p_c^{TS}$ for invasion of sink nodes is marked with a vertical dash-dotted line. The epidemic size is normalized by the network size $V$. The inverted red triangles represent the average number of infected sink nodes as a fraction of the system size $V$. When the average number of infected sink nodes was less than one, it was rounded up to one for an artificial floor of $\log_{10}(1/V) = -4$. (B) The distribution of the global attack rate follows the same notations as A. The global attack rate is scaled by a factor of 30 times the total system population, $30V\overline{N}$. (C) The distribution of the epidemic duration in days follows the same notations as A, normalized by 100 years. (D) The fraction of successful outbreaks that resulted in endemic situations is plotted against the movement rate $p$. All sub-figures represent realizations that consider a per capita per day death rate of 0.02. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

relatively small until crossing both invasion thresholds. The average epidemic size shows very little dependence on the death (slaughter) rate in Fig. 3A. For high movement rates and low death rates in Fig. 3B, the attack rates are smaller, perhaps limited by the reduced supply of susceptible individuals. In Fig. 3C and D, peninsulas appear in the highest values of the epidemic durations and the endemic fractions, respectively. This peculiar region of long-lasting epidemics may suggest a balance between the system movement and death rates that favors endemics. The numerical exploration of the endemic potential in Fig. 3D is not equivalent to an analytic study of endemic equilibria, but serves as a quantitative proxy. An analytical exploration of the endemic situation has not been undertaken.

The global invasion thresholds $R_*$ and $R_*^{TS}$ for directed networks add complexity to the $R_*$ of undirected systems (Balcan and Vespignani, 2012). To better characterize the compositions of these thresholds, we conducted 3 analytic explorations of the relationships between various metrics and the thresholds. First we varied the movement rate $p$ and the basic reproduction number $R_0$ and computed the values of $R_*$ and $R_*^{TS}$ in Fig. 4. The remaining parameters of the thresholds were chosen to match the network, demographic, and disease parameters of the simulations shown in Figs. 2 and 3. Fig. 4A and B respectively show the base-10 logarithm values of $R_*$ and $R_*^{TS}$ as a function of $p$ and $R_0$. Values of the thresholds which are below 1 are shown in shades of blue and those above are shown in shades of red. The thresholds are below 1 (0 on the logarithmic scale) for low movement rates and low values of $R_0$. Due to the network parameters used, $R_*$ crosses 1 before $R_*^{TS}$ does as $p$ and $R_0$ are increased. This ordering of $R_* > R_*^{TS}$ is in agreement with Eq. (20) when we consider the synthetic network compositions and degree distributions. We next examined the dependence of $R_*$ and $R_*^{TS}$ on both the movement rate $p$ and the death rate $\delta$. Fig. 5 displays the results. The data presented in Fig. 5A is comparable to the simulation results shown in Fig. 3A, with the exception of the difference in the ranges of $\delta$ explored. Fig. 5A adds to the story of Fig. 3A by suggesting that even smaller values of $\delta$would have further reduced the epidemic attack rates. The almost square shapes of Fig. 5A and B suggest that the roles of $p$ and $\delta$ on the thresholds are nearly independent of each other. Again we notice that $R_* > R_*^{TS}$ as in Fig. 4. This relationship of the two thresholds depends on the fractional compositions of the network nodes as well as the network degree distributions. Although there are many dimensions to explore in the directed network degree distributions, the impacts of the fractional compositions of the network on the thresholds are easier to estimate. We study the thresholds' dependence on $\eta_{out}$

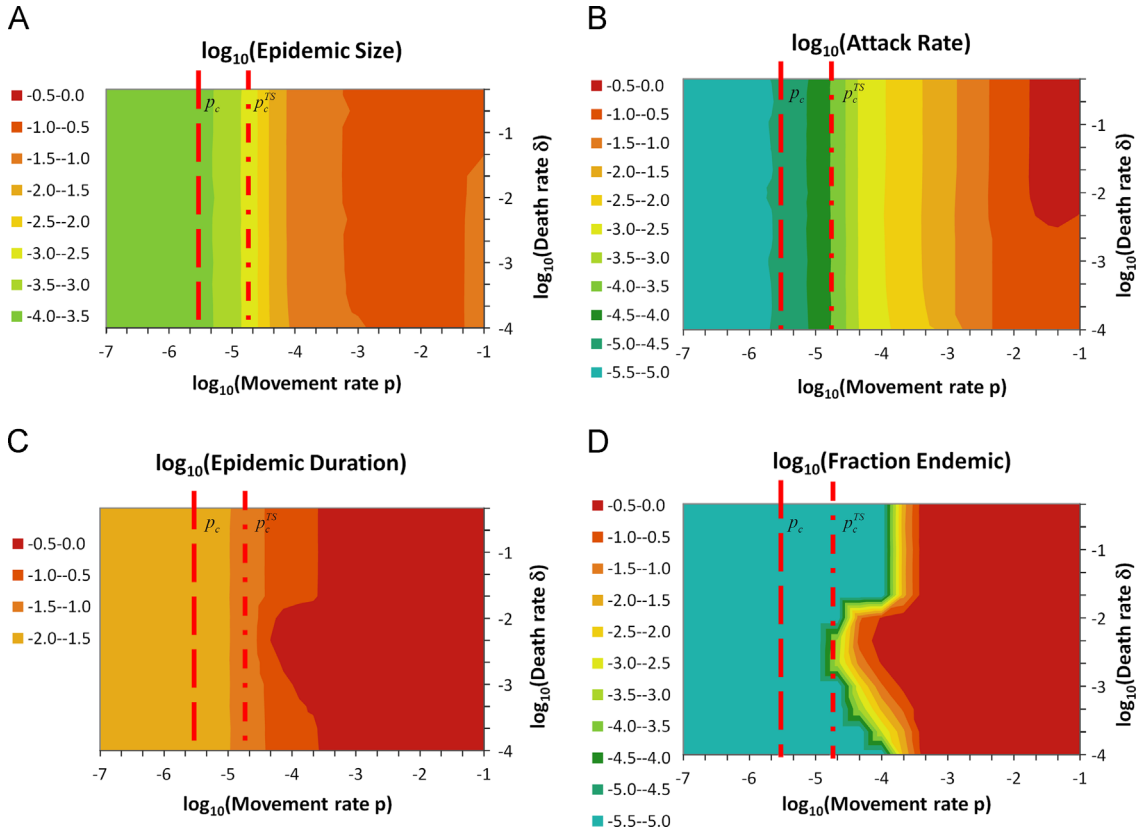


**Fig. 3.** (A) The logarithm of the average of the epidemic size as a function of the movement rate $p$ and the death rate $\delta$ is colored in the third dimension. The color scheme is displayed on the left side of the plot. The critical movement rate $p_c$ is marked with a red dashed line and similarly $p_c^{TS}$ is marked with a red dash-dotted line. The average epidemic size is normalized by the network size $V$. (B) The logarithm of the average of the global attack rate follows the same notations as A. The average global attack rate is scaled by a factor of 30 times the total system population, $30V\overline{N}$. (C) The averages of the logarithm of the epidemic duration are normalized by 100 years and follow the same notations as A. (D) The logarithm of the fraction of successful outbreaks that resulted in endemic situations is plotted against the movement rate $p$ and death rate $\delta$. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

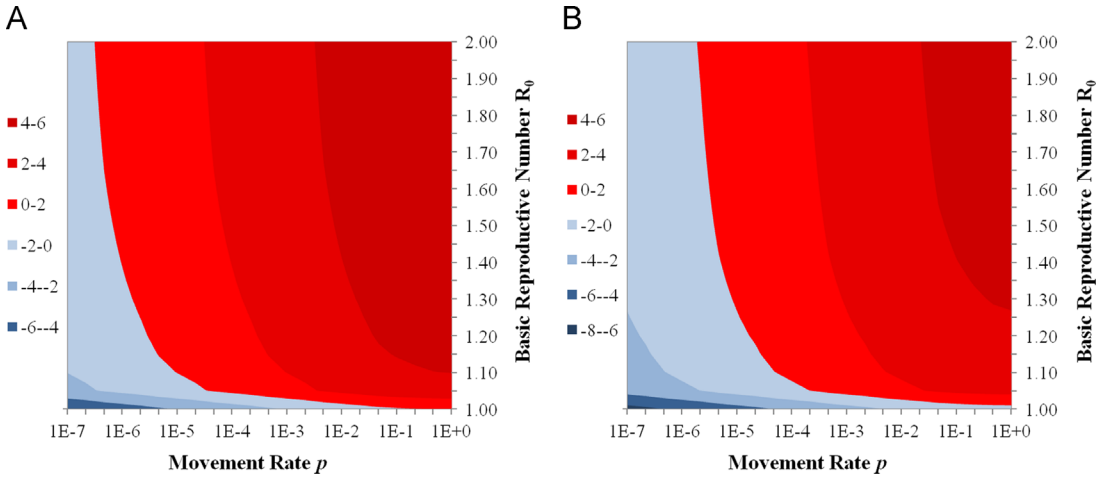


**Fig. 4.** (A) The dependence of the global epidemic invasion threshold for directed networks $R_*$ on the movement rate $p$ and the basic reproduction number $R_0$ is plotted using the base-10 logarithmic values of the threshold on a color map. (B) The dependence of the global transit-to-sink epidemic invasion threshold for directed networks $R_*^{TS}$ on the movement rate $p$ and the basic reproduction number $R_0$ is plotted using the base-10 logarithmic values of the threshold on the same color map. The lowest values of the thresholds are shown in darker shades of blue that become lighter shades of blue as the thresholds approach unity or 0 on the logarithmic scale. Values of the thresholds that are greater than unity are shown in red, with darker shades of red depicting higher values of the thresholds. The parameters used in computing these thresholds are the movement rate $p$, ranging from $10^{-7}$ to 1.0, the death rate $\delta = 0.02$, the fraction of nodes that are sources $\eta_{out} = 0.45$, the fraction of nodes that are sinks $\eta_{in} = 0.10$, the average population per node $\overline{N} = 10^5$, the recovery rate $\mu^{-1} = 7.0$, the basic reproduction number $R_0$, ranging from 1.001 to 2.000, the average node in-degree $\langle k^{in} \rangle = 1.52$, the average of the squares of the transit node in-degrees $\left\langle \left(k^{in(2)}\right)^2 \right\rangle = 47.68$, and the average of the squares of the sink node in-degrees $\left\langle \left(k^{in(3)}\right)^2 \right\rangle = 37.68$. All network parameters were chosen to represent the synthetic networks used in Figs. 2 and 3. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

and $\eta_{in}$ in Fig. 6. The infeasible pairs of $\eta_{out}$ and $\eta_{in}$ are depicted in white. Fig. 6A shows a single trend, where $R_*$ increases as the fraction of transit nodes increases. Or equivalently, $R_*$ decreases as the sum $\eta_{out} + \eta_{in}$ increases. The variation of $R_*$ that occurs almost perpendicular to this trend suggests that there is no significant difference between the contributions of $\eta_{out}$ and $\eta_{in}$. This might

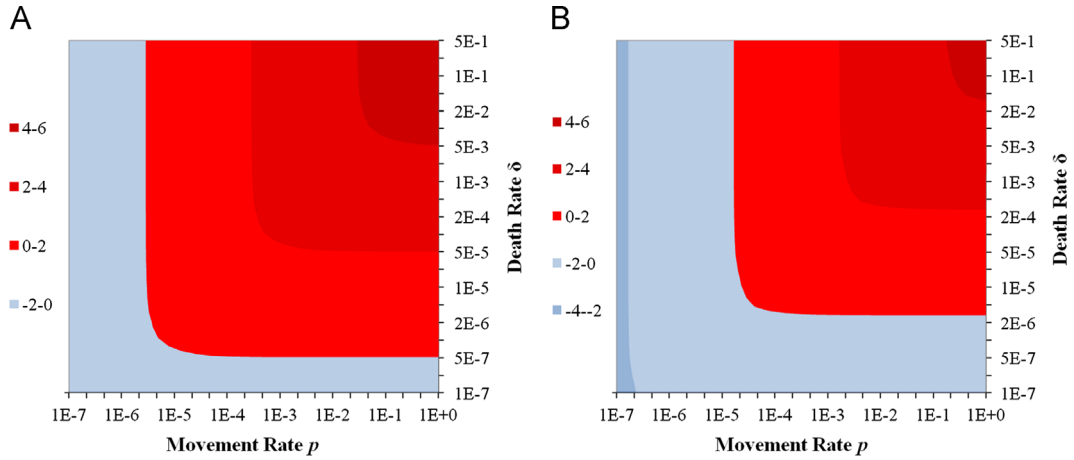


**Fig. 5.** (A) The dependence of the global epidemic invasion threshold for directed networks $R_*$ on the movement rate $p$ and the death rate $\delta$ is plotted using the base-10 logarithmic values of the threshold on a color map. (B) The dependence of the global transit-to-sink epidemic invasion threshold for directed networks $R_*^{TS}$ on the movement rate $p$ and the death rate $\delta$ is plotted using the base-10 logarithmic values of the threshold on the same color map. The lowest values of the thresholds are shown in darker shades of blue that become lighter shades of blue as the thresholds approach unity or 0 on the logarithmic scale. Values of the thresholds that are greater than unity are shown in red, with darker shades of red depicting higher values of the thresholds. The parameters used in computing these thresholds are the movement rate $p$, ranging from $10^{-7}$ to 1.0, the death rate $\delta$, ranging from $10^{-7}$ to 0.50, the fraction of nodes that are sources $\eta_{out}=0.45$, the fraction of nodes that are sinks $\eta_{in}=0.10$, the average population per node $\overline{N}=10^5$, the recovery rate $\mu^{-1}=7.0$, the basic reproduction number $R_0=1.20$, the average node in-degree $\langle k^{in}\rangle=1.52$, the average of the squares of the transit node in-degrees $\left\langle \left(k^{in(2)}\right)^2\right\rangle=47.68$, and the average of the squares of the sink node in-degrees $\left\langle \left(k^{in(3)}\right)^2\right\rangle=37.68$. All network parameters were chosen to represent the synthetic networks used in Figs. 2 and 3. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

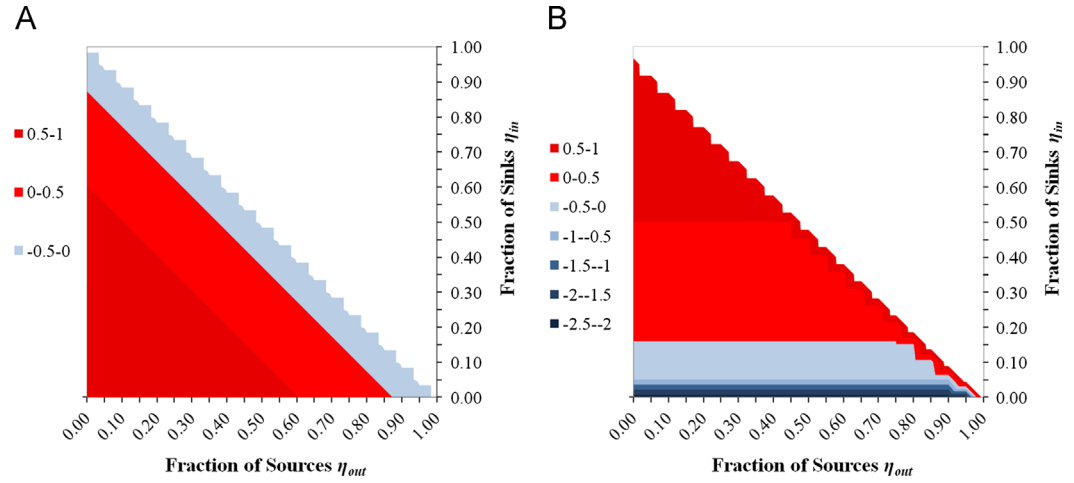


**Fig. 6.** (A) The dependence of the global epidemic invasion threshold for directed networks $R_*$ on the fraction of source nodes $\eta_{out}$ and the fraction of sink nodes $\eta_{in}$ is plotted using the base-10 logarithmic values of the threshold on a color map. (B) The dependence of the global transit-to-sink epidemic invasion threshold for directed networks $R_*^{TS}$ on the fraction of source nodes $\eta_{out}$ and the fraction of sink nodes $\eta_{in}$ is plotted using the base-10 logarithmic values of the threshold on the same color map. The lowest values of the thresholds are shown in darker shades of blue that become lighter shades of blue as the thresholds approach unity or 0 on the logarithmic scale. Values of the thresholds that are greater than unity are shown in red, with darker shades of red depicting higher values of the thresholds. The infeasible combinations of $\eta_{out}$ and $\eta_{in}$ are depicted in white. Note that the faction of transit nodes can be determined as $1-\eta_{out}-\eta_{in}$. The parameters used in computing these thresholds are the movement rate $p=10^{-5}$, the death rate $\delta=0.02$, the fraction of nodes that are sources $\eta_{out}$, ranging from 0.00 to 1.00, the fraction of nodes that are sinks $\eta_{in}$, ranging from 0.001 to 1.00, the average population per node $\overline{N}=10^5$, the recovery rate $\mu^{-1}=7.0$, the basic reproduction number $R_0=1.20$, the average node in-degree $\langle k^{in}\rangle=1.52$, the average of the squares of the transit node in-degrees $\left\langle \left(k^{in(2)}\right)^2\right\rangle=47.68$, and the average of the squares of the sink node in-degrees $\left\langle \left(k^{in(3)}\right)^2\right\rangle=37.68$. All network parameters were chosen to represent the synthetic networks used in Figs. 2 and 3. Not all combinations of $\eta_{out}$ and $\eta_{in}$ represent feasible network configurations when considering the values used for the in-degree averages. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

seem interesting as $\eta_{out}$ and $\eta_{in}$ appear in different places in the expression of $R_*$. However, as $R_*$ captures the epidemic invasion that occurs primarily in the transit nodes, the parallel dependence of $\eta_{out}$ and $\eta_{in}$ seems reasonable. The results for $R_*^{TS}$, shown in Fig. 6B, present a two-part relationship of the threshold with the network fractional compositions. The primary trend in Fig. 6B is a dependence of $R_*^{TS}$ on $\eta_{in}$, the fraction of sink nodes. As $R_*$ increases with the fraction of transit nodes, so also $R_*^{TS}$ increases with the fraction of sink nodes. The second trend in Fig. 6 Bappears when the fraction of transit nodes approaches zero. As the fraction of transit nodes nears zero, the values of $R_*^{TS}$ increase. Given a specific in-degree distribution for the sink nodes, as there are fewer and fewer transit nodes, the average out-degree of the transit nodes will be higher and higher. Further, these larger out-degrees will exhibit greater connectivity to the sink nodes as the number of transit nodes decreases, due to degree balancing. These degree-based limitations would increase the likelihood that an epidemic spreads from the transit nodes into the sink nodes as the fraction of transit nodes decreases. Only in networks with smaller fractions of transit nodes, will this secondary dependence arise for $R_*^{TS}$. The results of this section have not explored the impact of the network size, the average population size, or the degree distributions of the directed livestock networks. To fathom out the limitations of these results, further analysis is needed.

## 5. Conclusions

In this paper, we have designed a metapopulation model of a simplified livestock system for the purpose of exploring global invasion thresholds on directed metapopulation networks that contain source and sink nodes. With death, birth, importation, and movement processes, we have derived degree-block, mean-field solutions to the quasi-equilibria of the node populations. We added an SIR model to the system and have derived the directed-network version of the global epidemic invasion threshold $R_*$, describing the conditions necessary for a global outbreak. Furthermore, our analysis of the sink node populations has produced a second, novel global invasion threshold, $R_*^{TS}$, that describes the conditions necessary for an epidemic to break out of the giant strongly connected component and into the populations of the sink nodes. For these two thresholds, we have derived the critical movement rates. We determined the dependence of the critical movement rates' relationship on the ratios of the second moments of the node-type in-degree distributions and the fractions of the transit and sink nodes. The existence of this second global invasion threshold creates two transitions in the potential threat of emerging epidemics. The first transition enables a potential epidemic to invade the cattle populations. The second transition permits the epidemic to move with cattle to the slaughter facilities and to pose a risk to human populations. We have demonstrated these results numerically using synthetically generated, directed cattle networks.

We have further explored the composition of $R_*$ and $R_*^{TS}$ by examining interactions between the thresholds and three pairs of parameters. The results in Figs. 4 and 5 suggest that epidemics may be avoided by maintaining very low rates of movement and/or slaughter; however, these ideal rates are likely to represent unrealistically low values for a large-scale cattle industry. Although Fig. 6 is only an approximation, the analysis presented in Fig. 6B suggests that a danger exists in having too many sink nodes or too few transit nodes. Having only a small number of transit nodes may increase the chances of a disease breaking out of the transit nodes and into the sink nodes. In terms of a cattle industry, this would suggest that premises which regularly bring cattle in and out such as livestock markets and backgrounders should not represent too small of a fraction of the premises in the complete livestock network. If they do, they may pose a higher risk to the finishing feedlots (with a higher $R_*^{TS}$) and further to the supply of beef for human consumption. However, Fig. 6A suggests a concern with having too many transit nodes in the computation of $R_*$. $R_*$ is the primary threshold; $R_*^{TS}$ depends $R_*$ being greater than unity to even suggest a threat to the sink nodes. Therefore the smaller fraction of transit nodes, and thus less network connectivity, is the ideal direction for the network composition. Secondarily, a smaller fraction of sink nodes may be desirable. A mandated removal of transit nodes is not feasible for most countries and governments. A further focusing of disease-related surveillance and regulations on transit nodes may be a viable option for the prevention of epidemics among directed livestock systems.

The analyses and results of this paper have come with assumptions, and opportunities remain for more general results to be derived. The most critical assumptions have been made regarding the movements of livestock. We have considered a constant rate of movement $p$ that occurs uniformly to any node in the originating node's out-neighborhood. This has ignored both the heterogeneity and the strong seasonal components of actual livestock systems. We have only considered static metapopulation networks interconnected through fixed routes of cattle flows. Such assumptions would only be valid in the situation where the disease dynamics occur on a time scale that is significantly shorter than the time scale of the seasonal variations in the movements. Also related to the structures of cattle movements, we have only considered networks which possess no correlations between the degrees of connected nodes. All of the above assumptions could be relaxed if there exists data characterizing the livestock trade network under consideration. Further restrictions in this analysis include the importation of only susceptible individuals (see Appendix C) and the study of only a single disease and single livestock species and type within the system. Analytic derivations of endemic equilibria have not yet been determined. The lack of investigation into the impacts of the average population size, the number of nodes, and the node degree distributions are further limitations of this work.

These assumptions and limitations motivate future work. A deeper analytic understanding of these directed systems is necessary, but the most interesting future steps may be those that construct the livestock movement networks and traffic functions. The significance of these inputs for the system model and respective analyses drives the need for real-world data to characterize these movement inputs. A data-driven livestock metapopulation movement system, when combined with the results presented here, would enable the evaluation of different livestock disease control strategies, especially movement-restriction based methods. Despite the simplistic livestock movements which we have considered, our work has produced the global epidemic invasion threshold for directed metapopulation networks as well as a second novel global invasion threshold. The consideration of epidemic control strategies through these thresholds enables an immediate assessment of the strategies' effectiveness.

## Acknowledgments

We would like to dedicate this paper to Dr. Duygu Balcan for her invaluable friendship and contributions to this work. Dr. Balcan has recently passed away in an unfortunate accident in August, 2013. We wish to convey our deepest condolences to all the lives she has touched. We would like to acknowledge Dr. Vespignani for the many insights and inputs from him and his research team regarding this work. A portion of this work was completed while PS was on sponsored visits to the Computational Epidemiology Laboratory at the Institute for Scientific Interchange, in Turin, Italy. The computing for this project was performed on the Beocat Research Cluster at Kansas State University, which is funded in part by NSF Grants CNS-1006860, EPS-1006860, and EPS-0919443. Partial funding for this project was provided by the National Agricultural Biosecurity Center at Kansas State University. This research was partially funded by US Department of Agriculture, the Agricultural Research Service with Specific Cooperative Agreement 58-5430-1-0356.

## Appendix A. Solutions of model demographics

In the following we consider an ensemble of directed subpopulation networks that contain “source” and “sink” nodes. We propose population dynamics on the network ensemble, which considers birth, death, and migration processes while preserving the total population of the system.

### A.1. Subpopulation networks

Imagine to have a directed subpopulation network made of $V$ nodes. Each node can assume one of the three types or classes: source, sink, or transit. While source and sink nodes are connected to the rest of the system via out-going and in-coming links only, respectively, transit nodes have both out- and in-neighborhoods.

The probability of finding a node of type source (sink) when a node is chosen at random is $\eta_{out}$ ($\eta_{in}$), whereas $1-\eta_{out}-\eta_{in}$ is the probability that we will pick a transit node if a node is selected at random.

For the sake of simplicity, we assume that all the nodes but sources obey the same in-degree distribution $P^{in}_{v-o}\left(k^{in}\right)$ while all the nodes except sinks follow the same out-degree distribution $P^{out}_{v-i}\left(k^{out}\right)$. Moreover, we consider the case that there are no single-node degree correlations between the in- and out-degrees of transit nodes. However, the formulation below can be generalized to cover all the cases in which different types of nodes assume different degree distributions and also that there are single-node degree correlations. The joint-degree distribution $P_v(\vec{k})$ of the subpopulation network under the simplifying assumptions above is thus given by

$$P_v(\vec{k}) = \eta_{out}\delta_{k^{in},0}P^{out}_{v-i}\left(k^{out}\right) + \eta_{in}\delta_{k^{out},0}P^{in}_{v-o}\left(k^{in}\right) + (1-\eta_{out}-\eta_{in})P^{in}_{v-o}\left(k^{in}\right)P^{out}_{v-i}\left(k^{out}\right), \quad (22)$$

where

$$P^{in}_{v-o}(0)=0 \quad \text{and} \quad P^{out}_{v-i}(0)=0.$$

Note that $\delta_{x,y}$ here represents the Kronecker delta. We focus on the degree distribution in order to obtain the relationships which are to be used in the following derivations. The in-degree distribution $P^{in}_v\left(k^{in}\right)$ of the subpopulation network is

$$P^{in}_v\left(k^{in}\right) = \sum_{k^{out}} P_v\left(k^{in},k^{out}\right) = \eta_{out}\delta_{k^{in},0} + (1-\eta_{out})P^{in}_{v-o}\left(k^{in}\right). \quad (23)$$

We can obtain the relationship between the average in-degree of all the nodes excluding sources and the average in-degree $\left\langle k^{in}\right\rangle$ of the whole system by

$$\left\langle k^{in}\right\rangle = \sum_{k^{in}} k^{in}P^{in}_v\left(k^{in}\right) = (1-\eta_{out}) \sum_{k^{in}\geq 1} k^{in}P^{in}_{v-o}\left(k^{in}\right)$$

$$\Rightarrow \sum_{k^{in}\geq 1} k^{in}P^{in}_{v-o}\left(k^{in}\right) = \frac{\left\langle k^{in}\right\rangle}{(1-\eta_{out})}. \quad (24)$$

The probability that if we pick a node at random we will hit a sink with in-degree $k^{in}$ is $P_v\left(k^{in},0\right) = \eta_{in}P^{in}_{v-o}\left(k^{in}\right)$. Then it follows that

$$\sum_{k^{in}} k^{in}P^{in}_v\left(k^{in},0\right) = \eta_{in}\sum_{k^{in}} k^{in}P^{in}_{v-o}\left(k^{in}\right) = \eta_{in}\frac{\left\langle k^{in}\right\rangle}{(1-\eta_{out})}, \quad (25)$$

where we have plugged in Eq. (24). These last two equations are going to prove their worth when we consider population dynamics. The out-degree distribution of the subpopulation network is given by

$$P^{out}_v\left(k^{out}\right) = \sum_{k^{in}} P_v\left(k^{in},k^{out}\right) = \eta_{in}\delta_{k^{out},0} + (1-\eta_{in})P^{out}_{v-i}\left(k^{out}\right). \quad (26)$$

We obtain the relationship between the average out-degree $\left\langle k^{out}\right\rangle$ of the whole network and that of all the nodes excluding sinks by

$$\left\langle k^{out}\right\rangle = \sum_{k^{out}} k^{out}P^{out}_v\left(k^{out}\right) = (1-\eta_{in}) \sum_{k^{out}\geq 1} k^{out}P^{out}_{v-i}\left(k^{out}\right)$$

$$\Rightarrow \sum_{k^{out}\geq 1} k^{out}P^{out}_{v-i}\left(k^{out}\right) = \frac{\left\langle k^{out}\right\rangle}{(1-\eta_{in})}. \quad (27)$$

The probability of encountering a source with out-degree $k^{out}$ when a node is chosen at random is $P_v\left(0,k^{out}\right) = \eta_{out}P^{out}_{v-i}\left(k^{out}\right)$. It then follows that

$$\sum_{k^{out}} k^{out}P^{out}_v\left(0,k^{out}\right) = \eta_{out}\sum_{k^{out}} k^{out}P^{out}_{v-i}\left(k^{out}\right) = \eta_{out}\frac{\left\langle k^{out}\right\rangle}{(1-\eta_{in})}, \quad (28)$$

where we have plugged in Eq. (27). Similarly, the last two equations are going to be useful in the following where we consider population dynamics on the subpopulation networks described here. Let us denote the pairwise (or joint) node degree distribution by $P_a\left(\vec{k},\vec{k}'\right)$ which is the probability that an arc selected at random has its origin from a node with degree $\vec{k}$ and its termination at a node with degree $\vec{k}'$. The probability to find a node with degree $\vec{k}$ at the origin of a randomly chosen arc is

$$P^{out}_a(\vec{k}) = \sum_{\vec{k}'} P_a\left(\vec{k},\vec{k}'\right) = \frac{k^{out}P_v(\vec{k})}{\left\langle k^{out}\right\rangle}. \quad (29)$$

Likewise, the probability to find a node with degree $\vec{k}$ at the destination of a randomly chosen arc is

$$P^{in}_a(\vec{k}) = \sum_{\vec{k}'} P_a\left(\vec{k}',\vec{k}\right) = \frac{k^{in}P_v(\vec{k})}{\left\langle k^{in}\right\rangle}. \quad (30)$$

With an assumption of no correlations between the degrees of connected nodes, the joint node degree distribution becomes $P_a\left(\vec{k},\vec{k}'\right) = P^{out}_a(\vec{k})P^{in}_a\left(\vec{k}'\right)$. The locations of arcs in the subpopulation network as described by $P_a\left(\vec{k},\vec{k}'\right)$ play a significant role in the dynamic processes of the system.

### A.2. Demography in subpopulation networks

Consider now that each subpopulation of joint-degree $\vec{k}$ is occupied by $N_{\vec{k}}(t)$ individuals at time $t$. Each individual in a subpopulation of joint-degree $\vec{k}$ moves to a subpopulation with joint-degree $\vec{k}'$ in its out-neighborhood at rate

$$d_{\vec{k}\vec{k}'} = \frac{p}{k^{out}} \quad \forall k^{out}\neq 0, \quad (31)$$

yielding a total per capita diffusion rate $d_{\vec{k}}$ of

$$d_{\vec{k}} = p \quad \forall k^{out}\neq 0. \quad (32)$$

The above diffusion process is obviously defined for all the subpopulations except sinks. Each individual in a sink subpopulation with joint-degree $\vec{k}$, on the other hand, leaves the subpopulation network or dies at rate

$$\delta_{\vec{k}} = \delta \quad \forall k^{out}=0. \quad (33)$$

Each individual in a source subpopulation $\vec{k}$ replicates herself or gives birth to an offspring at rate

$$\beta_{\vec{k}}(t) = p_\beta\delta\frac{\sum_{\vec{j}\,|j^{out}=0} P_v\left(\vec{j}\right)N_{\vec{j}}(t)}{\sum_{\vec{l}\,|l^{in}=0} P_v\left(\vec{l}\right)N_{\vec{l}}(t)} \quad \forall k^{in}=0, \quad (34)$$

where $p_\beta$, $0\leq p_\beta\leq 1$, represents the fraction of individuals "recycled" from the death process into the system through the birth process. The remaining fraction, $\left(1-p_\beta\right)$, re-enters the network through an importation process to the source nodes.

Notice that the fraction term of Eq. (34) is the ratio of the total population in the sink nodes to the total population in the source nodes at time $t$. Each source subpopulation $\vec{k}$ imports individuals from an external source at rate

$$\epsilon_{\vec{k}}(t) = \left(1-p_\beta\right)\delta \frac{k^{out}}{\sum_{\vec{l}\,|\,l^{in}=0} P_v\left(\vec{l}\right) l^{out}} \sum_{\vec{j}\,|\,j^{out}=0} P_v\left(\vec{j}\right) N_{\vec{j}}(t) \quad \forall k^{in}=0, \tag{35}$$

that is proportional to the out-degree of the source node. Remembering Eq. (28) simplifies the importation rate to

$$\epsilon_{\vec{k}}(t) = \left(1-p_\beta\right)\delta \frac{(1-\eta_{in})k^{out}}{\eta_{out}\left\langle k^{out}\right\rangle} \sum_{\vec{j}\,|\,j^{out}=0} P_v\left(\vec{j}\right) N_{\vec{j}}(t) \quad \forall k^{in}=0, \tag{36}$$

The rate equation of average population size $N_{\vec{k}}(t)$ in a subpopulation with joint-degree $\vec{k}$ is then

$$\begin{aligned}\partial_t N_{\vec{k}}(t) = &-\delta_{\vec{k}} N_{\vec{k}}(t)\delta_{k^{out},0} - d_{\vec{k}} N_{\vec{k}}(t)\left(1-\delta_{k^{out},0}\right)\\ &+\left[\beta_{\vec{k}}(t) N_{\vec{k}}(t) + \epsilon_{\vec{k}}(t)\right]\delta_{k^{in},0}\\ &+k^{in}\sum_{\vec{j}} P_a\left(\vec{j},\vec{k}\,|\,\vec{k}\right) d_{\vec{j}}\, {}_{\vec{k}} N_{\vec{j}}(t)\left(1-\delta_{j^{out},0}\right).\end{aligned} \tag{37}$$

If we insert all the rates in Eq. (37) and also assume that there are no correlations between the degrees of connected nodes, we get

$$\begin{aligned}\partial_t N_{\vec{k}}(t) = &-\delta N_{\vec{k}}(t)\delta_{k^{out},0} - p N_{\vec{k}}(t)\left(1-\delta_{k^{out},0}\right)\\ &+\left[p_\beta\delta\frac{N_{\vec{k}}(t)}{\sum_{\vec{l}\,|\,l^{in}=0} P_v\left(\vec{l}\right) N_{\vec{l}}(t)} + \left(1-p_\beta\right)\delta\frac{(1-\eta_{in})k^{out}}{\eta_{out}\left\langle k^{out}\right\rangle}\right]\\ &\times \sum_{\vec{j}\,|\,j^{out}=0} P_v\left(\vec{j}\right) N_{\vec{j}}(t)\delta_{k^{in},0}\\ &+p\frac{k^{in}}{\left\langle k^{out}\right\rangle}\sum_{\vec{j}} P_v\left(\vec{j}\right) N_{\vec{j}}(t)\left(1-\delta_{j^{out},0}\right).\end{aligned} \tag{38}$$

The first thing to consider is that the average population per node $\overline{N} = \sum_{\vec{k}} P_v(\vec{k}) N_{\vec{k}}(t)$ is kept invariant over time by the above dynamical process, i.e., $\sum_{\vec{k}} P_v(\vec{k})\partial_t N_{\vec{k}}(t) = 0$, which leads to

$$\begin{aligned}\partial_t N_{\vec{k}}(t) = &-\delta N_{\vec{k}}(t)\delta_{k^{out},0} - p N_{\vec{k}}(t)\left(1-\delta_{k^{out},0}\right)\\ &+\left[p_\beta\delta\frac{N_{\vec{k}}(t)}{\sum_{\vec{l}\,|\,l^{in}=0} P_v\left(\vec{l}\right) N_{\vec{l}}(t)} + \left(1-p_\beta\right)\delta\frac{(1-\eta_{in})k^{out}}{\eta_{out}\left\langle k^{out}\right\rangle}\right]\\ &\times \sum_{\vec{j}\,|\,j^{out}=0} P_v\left(\vec{j}\right) N_{\vec{j}}(t)\delta_{k^{in},0}\\ &+p\frac{k^{in}}{\left\langle k^{out}\right\rangle}\left[\overline{N} - \sum_{\vec{j}\,|\,j^{out}=0} P_v\left(\vec{j}\right) N_{\vec{j}}(t)\right].\end{aligned} \tag{39}$$

Now let us simplify the equations at the expense of more definitions, which are going to be helpful in the rest of the derivations. Let us denote the total population in all source nodes per all the nodes at time $t$ by

$$\psi_{out}(t) \equiv \sum_{\vec{k}\,|\,k^{in}=0} P_v(\vec{k}) N_{\vec{k}}(t). \tag{40}$$

Similarly we denote the total population in all sink nodes per all the nodes at time $t$ by

$$\psi_{in}(t) \equiv \sum_{\vec{k}\,|\,k^{out}=0} P_v(\vec{k}) N_{\vec{k}}(t). \tag{41}$$

Substituting Eqs. (40) and (41) into the rate equation of $N_{\vec{k}}$, we obtain

$$\begin{aligned}\partial_t N_{\vec{k}}(t) = &-\delta N_{\vec{k}}(t)\delta_{k^{out},0} - p N_{\vec{k}}(t)\left(1-\delta_{k^{out},0}\right)\\ &+\left(p_\beta\delta\frac{N_{\vec{k}}(t)}{\psi_{out}(t)} + \left(1-p_\beta\right)\delta\frac{(1-\eta_{in})k^{out}}{\eta_{out}\left\langle k^{out}\right\rangle}\right)\psi_{in}(t)\delta_{k^{in},0}\\ &+p\frac{k^{in}}{\left\langle k^{out}\right\rangle}(\overline{N}-\psi_{in}(t)).\end{aligned} \tag{42}$$

It is hard to analyze the above closed system all at once. We thus consider the rate equation for each type of node separately. To clearly distinguish the results from this point onward, we adopt the following notation for the respective node degrees of each type of node: $\vec{k}^{(1)}$ for all source nodes $\left(k^{in}=0\right)$, $\vec{k}^{(2)}$ for all transit nodes $\left(k^{in}\neq 0 \text{ and } k^{out}\neq 0\right)$, and $\vec{k}^{(3)}$ for all sink nodes $\left(k^{in}\neq 0 \text{ and } k^{out}=0\right)$.

(1) *Sink nodes*: The rate equation of the average population size in a sink node with joint-degree $\vec{k}^{(3)} = \left(k^{in(3)}, 0\right)$ is

$$\partial_t N_{\vec{k}^{(3)}}(t) = -\delta N_{\vec{k}^{(3)}}(t) + p\frac{k^{in(3)}}{\left\langle k^{out}\right\rangle}(\overline{N}-\psi_{in}(t)). \tag{43}$$

If we multiply both sides of the above equation by $P_v\left(\vec{k}^{(3)}\right)$ and then sum over all the joint-degrees $\vec{k}^{(1)} = \left(k^{in(3)}, 0\right)$, we obtain

$$\partial_t\psi_{in}(t) = -\left(\delta + \frac{\eta_{in}p}{1-\eta_{out}}\right)\psi_{in}(t) + \frac{\eta_{in}p}{1-\eta_{out}}\overline{N}, \tag{44}$$

where we have substituted in Eq. (25). This first order differential equation can be solved as reported in Appendix A.3, yielding the equilibrium configuration

$$\psi_{in}^{*} = \frac{\eta_{in}p}{(1-\eta_{out})\delta + \eta_{in}p}\overline{N}, \tag{45}$$

and a characteristic relaxation time of $(\delta + \eta_{in}p/(1-\eta_{out}))^{-1}$. Using the solution for $\psi_{in}(t)$, we similarly solve the differential equation for $N_{\vec{k}^{(3)}}(t)$. The solution, in particular, leads to the equilibrium configuration

$$N^{*}_{\vec{k}^{(3)}} = \frac{k^{in(3)}}{\left\langle k^{out}\right\rangle}\frac{(1-\eta_{out})p}{(1-\eta_{out})\delta + \eta_{in}p}\overline{N}, \tag{46}$$

and the characteristic relaxation time $\delta^{-1}$ (see Appendix A.3).

(2) *Transit nodes*: The rate equation of the average population size in a transit node subpopulation with joint-degree $\vec{k}^{(2)}$ follows very closely the rate equation for sink nodes:

$$\partial_t N_{\vec{k}^{(2)}}(t) = -pN_{\vec{k}^{(2)}}(t) + p\frac{k^{in(2)}}{\left\langle k^{out}\right\rangle}(\overline{N}-\psi_{in}(t)), \tag{47}$$

whose solution is reported in Appendix A.3. The solution of $N_{\vec{k}^{(2)}}(t)$ assumes the characteristic relaxation time $\max\left(p^{-1},(\delta+\eta_{in}p/(1-\eta_{out}))^{-1}\right)$ to the equilibrium configuration of the populations of the transit nodes:

$$N^*_{\vec{k}^{(2)}}=\frac{k^{in(2)}}{\langle k^{out}\rangle}\frac{(1-\eta_{out})\delta}{(1-\eta_{out})\delta+\eta_{in}p}\overline{N}. \tag{48}$$

(3) *Source nodes*: The rate equation of the average population in a source node with joint-degree $\vec{k}^{(1)}=\left(0,k^{out(1)}\right)$ is

$$\partial_t N_{\vec{k}^{(1)}}(t)=-pN_{\vec{k}^{(1)}}(t)+\left(p_\beta\delta\frac{N_{\vec{k}^{(1)}}(t)}{\psi_{out}(t)}+\left(1-p_\beta\right)\delta\frac{(1-\eta_{in})k^{out(1)}}{\eta_{out}\langle k^{out}\rangle}\right)\psi_{in}(t). \tag{49}$$

If we multiply both sides of the above equation by $P_v\left(\vec{k}^{(1)}\right)$ and then sum over all the joint-degrees $\vec{k}^{(1)}=\left(0,k^{out(1)}\right)$, we have

$$\partial_t\psi_{out}(t)=-p\psi_{out}(t)+\delta\psi_{in}(t), \tag{50}$$

where we have substituted in Eq. (28). As derived in Appendix A.3, the solution $\psi_{out}(t)$ admits the equilibrium configuration

$$\psi^*_{out}=\frac{\delta}{p}\psi^*_{in}=\frac{\eta_{in}\delta}{(1-\eta_{out})\delta+\eta_{in}p}\overline{N}, \tag{51}$$

with a characteristic time scale of $\max\left(p^{-1},(\delta+\eta_{in}p/(1-\eta_{out}))^{-1}\right)$. Particular attention is devoted to the rate equation of the degree-block variable $N_{\vec{k}^{(1)}}(t)$ in that we consider three cases of $p_\beta$ separately:

- If $p_\beta=1$, meaning that we only consider birth, then the rate equation is given by

$$\partial_t N_{\vec{k}^{(1)}}(t)=\left(-p+\delta\frac{\psi_{in}(t)}{\psi_{out}(t)}\right)N_{\vec{k}^{(1)}}(t). \tag{52}$$

The expression in brackets is equal to $\psi^{-1}_{out}\partial\psi_{out}$ (see Eq. (50)), yielding

$$\frac{\partial_t N_{\vec{k}^{(1)}}(t)}{N_{\vec{k}^{(1)}}}(t)=\frac{\partial\psi_{out}(t)}{\psi_{out}(t)}. \tag{53}$$

Using the definition of $\psi_{in}(t)$ in Eq. (41), we obtain the solution

$$N_{\vec{k}^{(1)}}(t)=\frac{1}{\eta_{out}}\psi_{out}(t). \tag{54}$$

This means that the characteristic relaxation time is $\max\left(p^{-1},(\delta+\eta_{in}p/(1-\eta_{out}))^{-1}\right)$ and the equilibrium configuration is

$$N^*_{\vec{k}^{(1)}}=\frac{1}{\eta_{out}}\psi^*_{out}=\frac{\eta_{in}\delta}{\eta_{out}\left[(1-\eta_{out})\delta+\eta_{in}p\right]}\overline{N}. \tag{55}$$

- Now let us look at the other extreme of $p_\beta=0$, meaning that we only consider the importation process. In this case, the rate equation is

$$\partial_t N_{\vec{k}^{(1)}}(t)=-pN_{\vec{k}^{(1)}}(t)+\delta\frac{(1-\eta_{in})k^{out(1)}}{\eta_{out}\langle k^{out}\rangle}\psi_{in}(t). \tag{56}$$

Using the solution for $\psi_{in}(t)$, we can solve the differential equation (see Appendix A.3) which yields the equilibrium configuration

$$N^*_{\vec{k}^{(1)}}=\frac{(1-\eta_{in})\delta\psi^*_{in}}{\eta_{out}p}\frac{k^{out(1)}}{\langle k^{out}\rangle}=\frac{k^{out(1)}}{\langle k^{out}\rangle}\frac{\eta_{in}(1-\eta_{in})\delta}{\eta_{out}\left[(1-\eta_{out})\delta+\eta_{in}p\right]}\overline{N} \tag{57}$$

and the characteristic relaxation time $\max\left(p^{-1},(\delta+\eta_{in}p/(1-\eta_{out}))^{-1}\right)$.

- On the other hand, if $0<p_\beta<1$, then we have all the terms of the rate equation, that is, Eq. (49). In this case, the differential equation does not seem to be easy to solve. However, the equilibrium configuration can easily be evaluated, yielding

$$N^*_{\vec{k}^{(1)}}=\frac{(1-\eta_{in})\delta\psi^*_{in}}{\eta_{out}p}\frac{k^{out(1)}}{\langle k^{out}\rangle}=\frac{k^{out(1)}}{\langle k^{out}\rangle}\frac{\eta_{in}(1-\eta_{in})\delta}{\eta_{out}\left[(1-\eta_{out})\delta+\eta_{in}p\right]}\overline{N}, \tag{58}$$

that is exactly the same as the solution for the case of $p_\beta=0$ (i.e., no birth).

### A.3. Solutions of rate equations

In the following, we solve the differential equations for the average population sizes in subpopulations of different types of nodes; sink, transit, and source.

(1) *Sink nodes*: Remember the rate equation $\partial_t\psi_{in}(t)$ in Eq. (44) which expresses the change of the total number of individuals in sink nodes. The solution of this first order differential equation is

$$\psi_{in}(t)=e^{-(\delta+\eta_{in}p/(1-\eta_{out}))t}\left(C_{in}+\overline{N}\frac{\eta_{in}p}{1-\eta_{out}}\int e^{(\delta+\eta_{in}p/(1-\eta_{out}))t}\,dt\right), \tag{59}$$

$$\psi_{in}(t)=C_{in}e^{-(\delta+\eta_{in}p/(1-\eta_{out}))t}+\overline{N}\frac{\eta_{in}p}{(1-\eta_{out})\delta+\eta_{in}p}, \tag{60}$$

where $C_{in}$ is a time-independent variable determined by initial conditions. If we assume that all the subpopulations have the same number of individuals initially, i.e., $N_{\vec{k}^{(1)}}(0)=N_{\vec{k}^{(2)}}(0)=N_{\vec{k}^{(3)}}(0)=\overline{N}$, then we have that $\psi_{in}(0)=\eta_{in}\overline{N}$, leading to

$$\psi_{in}(t)=\eta_{in}\overline{N}\frac{(1-\eta_{out})\delta-(1-\eta_{in})p}{(1-\eta_{out})\delta+\eta_{in}p}e^{-(\delta+\eta_{in}p/(1-\eta_{out}))t}+\eta_{in}\overline{N}\frac{p}{(1-\eta_{out})\delta+\eta_{in}p}. \tag{61}$$

Now we turn our attention to the rate equation of the joint-degree block variable $N_{\vec{k}^{(3)}}(t)$ in Eq. (43). The solution is given by

$$N_{\vec{k}^{(3)}}(t)=e^{-\delta t}\left(C_{\vec{k}^{(3)}}+p\frac{k^{in(3)}}{\langle k^{out}\rangle}\int e^{\delta t}\left(\overline{N}-\psi_{in}(t)\right)\,dt\right),$$

$$N_{\vec{k}^{(3)}}(t)=C_{\vec{k}^{(3)}}e^{-\delta t}+(1-\eta_{out})\overline{N}\frac{p}{(1-\eta_{out})\delta+\eta_{in}p}\frac{k^{in(3)}}{\langle k^{out}\rangle}+(1-\eta_{out})\overline{N}\frac{(1-\eta_{out})\delta-(1-\eta_{in})p}{(1-\eta_{out})\delta+\eta_{in}p}\frac{k^{in(3)}}{\langle k^{out}\rangle}e^{-(\delta+\eta_{in}p/(1-\eta_{out}))t}, \tag{62}$$

where $C_{\vec{k}^{(3)}}$ is a constant fixed by the initial condition $N_{\vec{k}^{(3)}}(0)=\overline{N}$. Inserting the initial condition yields

$$N_{\vec{k}^{(3)}}(t)=\overline{N}\left(1-(1-\eta_{out})\frac{k^{in(3)}}{\langle k^{out}\rangle}\right)e^{-\delta t}$$

$$+(1-\eta_{out})\overline{N}\frac{(1-\eta_{out})\delta-(1-\eta_{in})p}{(1-\eta_{out})\delta+\eta_{in}p}\frac{k^{in(3)}}{\langle k^{out}\rangle}e^{-(\delta+\eta_{in}p/(1-\eta_{out}))t}$$

$$+(1-\eta_{out})\overline{N}\frac{p}{(1-\eta_{out})\delta+\eta_{in}p}\frac{k^{in(3)}}{\langle k^{out}\rangle}, \tag{63}$$

(2) *Transit nodes*: The solution of the rate equation of the average population size in a transit node with joint-degree $\overrightarrow{k}^{(2)}$ in Eq. (47) can be similarly solved by

$$N_{\overrightarrow{k}^{(2)}}(t)=e^{-pt}\left(C_{\overrightarrow{k}^{(2)}}+p\frac{k^{in(2)}}{\langle k^{out}\rangle}\int e^{pt}\left(\overline{N}-\psi_{in}(t)\right)dt\right),$$

$$N_{\overrightarrow{k}^{(2)}}(t)=C_{\overrightarrow{k}^{(2)}}e^{-pt}+(1-\eta_{out})\overline{N}\frac{\delta}{(1-\eta_{out})\delta+\eta_{in}p}\frac{k^{in(2)}}{\langle k^{out}\rangle}$$
$$+\eta_{in}\overline{N}\frac{(1-\eta_{out})p\left[(1-\eta_{out})\delta-(1-\eta_{in})p\right]}{\left[(1-\eta_{out})\delta+\eta_{in}p\right]\left[(1-\eta_{out})\delta-(1-\eta_{out}-\eta_{in})p\right]}$$
$$\times\frac{k^{in(2)}}{\langle k^{out}\rangle}e^{-(\delta+\eta_{in}p/(1-\eta_{out}))t}, \tag{64}$$

where $C_{\overrightarrow{k}^{(2)}}$ is a constant fixed by the initial condition $N_{\overrightarrow{k}^{(2)}}(0)=\overline{N}$. Inserting the initial condition leads to

$$N_{\overrightarrow{k}^{(2)}}(t)=\overline{N}\left[1-\frac{(1-\eta_{out})}{(1-\eta_{out})\delta+\eta_{in}p}\right.$$
$$\left.\times\left(\delta+\eta_{in}p\frac{(1-\eta_{out})\delta-(1-\eta_{in})p}{(1-\eta_{out})\delta-(1-\eta_{out}-\eta_{in})p}\right)\frac{k^{in(2)}}{\langle k^{out}\rangle}\right]e^{-pt}$$
$$+\eta_{in}\overline{N}\frac{(1-\eta_{out})p\left[(1-\eta_{out})\delta-(1-\eta_{in})p\right]}{\left[(1-\eta_{out})\delta+\eta_{in}p\right]\left[(1-\eta_{out})\delta-(1-\eta_{out}-\eta_{in})p\right]}$$
$$\times\frac{k^{in(2)}}{\langle k^{out}\rangle}e^{-(\delta+\eta_{in}p/(1-\eta_{out}))t}$$
$$+(1-\eta_{out})\overline{N}\frac{\delta}{(1-\eta_{out})\delta+\eta_{in}p}\frac{k^{in(2)}}{\langle k^{out}\rangle}, \tag{65}$$

(3) *Source nodes*: Recall the rate equation in Eq. (50) that determines the change of the total population size in the source nodes. This first order differential equation can easily be solved by

$$\psi_{out}(t)=e^{-pt}\left(C_{out}+\delta\int e^{pt}\psi_{in}(t)\,dt\right),$$

$$\psi_{out}(t)=C_{out}e^{-pt}+\delta e^{-pt}$$
$$\times\int e^{pt}\eta_{in}\overline{N}\frac{(1-\eta_{out})\delta-(1-\eta_{in})p}{(1-\eta_{out})\delta+\eta_{in}p}e^{-(\delta+\eta_{in}p/(1-\eta_{out}))t}\,dt$$
$$+\delta e^{-pt}\int e^{pt}\eta_{in}\overline{N}\frac{p}{(1-\eta_{out})\delta+\eta_{in}p}dt$$

$$\psi_{out}(t)=C_{out}e^{-pt}-\eta_{in}\overline{N}$$
$$\times\frac{(1-\eta_{out})\delta\left[(1-\eta_{out})\delta-(1-\eta_{in})p\right]}{\left[(1-\eta_{out})\delta-(1-\eta_{out}-\eta_{in})p\right]\left[(1-\eta_{out})\delta+\eta_{in}p\right]}e^{-(\delta+\eta_{in}p/(1-\eta_{out}))t}$$
$$+\eta_{in}\overline{N}\frac{\delta}{(1-\eta_{out})\delta+\eta_{in}p}, \tag{66}$$

where the constant $C_{out}$ is given by the initial condition $\psi_{out}(0)=\eta_{out}\overline{N}$, yielding

$$\psi_{out}(t)=\eta_{out}\overline{N}\frac{(1-\eta_{out}-\eta_{in})(\delta-p)}{(1-\eta_{out})\delta-(1-\eta_{out}-\eta_{in})p}e^{-pt}$$
$$-\eta_{in}\overline{N}\frac{(1-\eta_{out})\delta\left[(1-\eta_{out})\delta-(1-\eta_{in})p\right]}{\left[(1-\eta_{out})\delta-(1-\eta_{out}-\eta_{in})p\right]\left[(1-\eta_{out})\delta+\eta_{in}p\right]}e^{-(\delta+\eta_{in}p/(1-\eta_{out}))t}$$
$$+\eta_{in}\overline{N}\frac{\delta}{(1-\eta_{out})\delta+\eta_{in}p}. \tag{67}$$

Now we focus on the degree-block variable $N_{\overrightarrow{k}^{(1)}}(t)$. In the following, we consider two cases of $p_\beta$:

- $p_\beta=1$: In this case, the rate equation leads to the solution $N_{\overrightarrow{k}^{(1)}}(t)=\eta_{out}^{-1}\psi_{out}(t)$.
- $p_\beta=0$: In this case, the rate equation is given by Eq. (56) and the solution follows as

$$N_{\overrightarrow{k}^{(1)}}(t)=e^{-pt}\left(C_{\overrightarrow{k}^{(1)}}+\delta\frac{(1-\eta_{in})k^{out(1)}}{\eta_{out}\langle k^{out}\rangle}\int e^{pt}\psi_{in}(t)\,dt\right), \tag{68}$$

$$N_{\overrightarrow{k}^{(1)}}(t)=C_{\overrightarrow{k}^{(1)}}e^{-pt}+\delta\frac{(1-\eta_{in})k^{out(1)}}{\eta_{out}\langle k^{out}\rangle}e^{-pt}$$
$$\times\int e^{pt}\eta_{in}\overline{N}\frac{(1-\eta_{out})\delta-(1-\eta_{in})p}{(1-\eta_{out})\delta+\eta_{in}p}e^{-(\delta+\eta_{in}p/(1-\eta_{out}))t}\,dt$$
$$+\delta\frac{(1-\eta_{in})k^{out(1)}}{\eta_{out}\langle k^{out}\rangle}e^{-pt}\int e^{pt}\eta_{in}\overline{N}\frac{p}{(1-\eta_{out})\delta+\eta_{in}p}dt$$

$$N_{\overrightarrow{k}^{(1)}}(t)=C_{\overrightarrow{k}^{(1)}}e^{-pt}$$
$$-\eta_{in}\overline{N}\frac{(1-\eta_{in})(1-\eta_{out})\delta\left[(1-\eta_{out})\delta-(1-\eta_{in})p\right]}{\eta_{out}\left[(1-\eta_{out})\delta-(1-\eta_{out}-\eta_{in})p\right]\left[(1-\eta_{out})\delta+\eta_{in}p\right]}$$
$$\times\frac{k^{out(1)}}{\langle k^{out}\rangle}e^{-(\delta+\eta_{in}p/(1-\eta_{out}))t}+\eta_{in}\overline{N}\frac{(1-\eta_{in})\delta}{\eta_{out}\left[(1-\eta_{out})\delta+\eta_{in}p\right]}\frac{k^{out(1)}}{\langle k^{out}\rangle}, \tag{69}$$

where $C_{\overrightarrow{k}^{(1)}}$ is a constant set by the initial condition $N_{\overrightarrow{k}^{(1)}}(0)=\overline{N}$. Using the initial value, we obtain

$$N_{\overrightarrow{k}^{(1)}}(t)=\left(1-\overline{N}\frac{\eta_{in}(1-\eta_{in})\delta}{(1-\eta_{out})\delta-(1-\eta_{out}-\eta_{in})p}\frac{k^{out(1)}}{\langle k^{out}\rangle}\right)e^{-pt}$$
$$-\eta_{in}\overline{N}\frac{(1-\eta_{in})(1-\eta_{out})\delta\left[(1-\eta_{out})\delta-(1-\eta_{in})p\right]}{\eta_{out}\left[(1-\eta_{out})\delta-(1-\eta_{out}-\eta_{in})p\right]\left[(1-\eta_{out})\delta+\eta_{in}p\right]}$$
$$\times\frac{k^{out(1)}}{\langle k^{out}\rangle}e^{-(\delta+\eta_{in}p/(1-\eta_{out}))t}+\eta_{in}\overline{N}\frac{(1-\eta_{in})\delta}{\eta_{out}\left[(1-\eta_{out})\delta+\eta_{in}p\right]}\frac{k^{out(1)}}{\langle k^{out}\rangle}. \tag{70}$$

## Appendix B. Derivation of global invasion thresholds

In this appendix, we present the derivations of the global epidemic invasion thresholds, their respective critical movement rates, and the relationship between the two thresholds as seen through a comparison of the critical movement rates.

### B.1. Global invasion thresholds

We consider the $n$th generation number of diseased subpopulations with degree $\overrightarrow{k}^{(x)}$ from node class $x$, $D^n_{\overrightarrow{k}^{(x)}}$, as a function of the three sets $\{D^{n-1}_{\overrightarrow{k}^{(1)}}\}$, $\{D^{n-1}_{\overrightarrow{k}^{(2)}}\}$, and $\{D^{n-1}_{\overrightarrow{k}^{(3)}}\}$ of the $(n-1)$th generation. The "infection" of a node occurs when infected cattle move from one node into another node containing only susceptible individuals. The expression for the branching process as expressed by Colizza and Vespignani (2008) and Balcan and

Vespignani (2012) follows as

$$D^n_{\vec{k}} = \sum_{\vec{j}} D^{n-1}_{\vec{j}} P_a\left(\vec{j}, \vec{k} \mid \vec{j}\right) j^{out} p(\vec{j}, \vec{k}) \prod_{m=0}^{n-1} \left(1 - \frac{D^m_{\vec{k}}}{V_{\vec{k}}}\right) \tag{71}$$

This branching equation models the $n$th generation of newly infected nodes by considering that each subpopulation in a node with a (out-) degree of $j$ has the potential to infect $j$ other nodes by the intersection of three events. These events are that a node with degree $\vec{k}$ exists in the out neighborhood of a node with degree $\vec{j}$, $P_a\left(\vec{j}, \vec{k} \mid \vec{j}\right)$, that the neighbor with degree $k$ has not been infected in previous generations, $\prod_{m=0}^{n-1}\left(1 - D^m_{\vec{k}}/V_{\vec{k}}\right)$, and that the disease will spread from the node with degree $j$ to the node with degree $k$, $p(\vec{j}, \vec{k})$. For the SIR model (Bailey, 1975), $p(\vec{j}, \vec{k})$ is given by

$$p(\vec{j}, \vec{k}) = 1 - R_0^{-\lambda_{\vec{j}\vec{k}}}, \tag{72}$$

where $\lambda_{\vec{j}\vec{k}}$ is the number of infected individuals moving from the infected subpopulation to the fully susceptible subpopulation. The SIR model solves for

$$\lambda_{\vec{j}\vec{k}} = \frac{\alpha}{\mu} N_{\vec{j}} d_{\vec{j}\vec{k}}, \tag{73}$$

where $\alpha$ is the final size of an SIR epidemic. Furthermore, the probability $p(\vec{j}, \vec{k})$ can be approximated as $p(\vec{j}, \vec{k}) \simeq \alpha N_{\vec{j}} d_{\vec{j}\vec{k}} (R_0 - 1)/\mu \simeq 2\left(1 - R_0^{-1}\right)^2 N_{\vec{j}} d_{\vec{j}\vec{k}}/\mu$ if we assume $R_0 \simeq 1$. The assumption that we are considering as an epidemic in its early stages allows the approximation $\prod_{m=0}^{n-1}\left(1 - D^m_{\vec{k}}/V_{\vec{k}}\right) \simeq 1$ and yields the approximated branching model:

$$D^n = \frac{2}{\mu}\left(1 - R_0^{-1}\right)^2 \sum_{\vec{j}} D^{n-1}_{\vec{j}} P_a\left(\vec{j}, \vec{k} \mid \vec{j}\right) j^{out} N_{\vec{j}} d_{\vec{j}\vec{k}}. \tag{74}$$

Note that this approximation of the process ignores bidirectional arcs and also has an assumption that the next generation is only infected by the immediately previous generation and none prior. This second assumption will remove the impact of the initial condition of the sink and sources nodes. Let us expand this expression to correctly consider the 3 classes of nodes. The lack of in-degrees for the source nodes and out-degrees for the sink nodes yields three expressions for the source and sink branching model on directed networks:

$$D^n_{\vec{k}^{(1)}} = \frac{2}{\mu}\left(1 - R_0^{-1}\right)^2 \left[ \sum_{\vec{j}^{(1)}} D^{n-1}_{\vec{j}^{(1)}} 0 j^{out(1)} N_{\vec{j}^{(1)}} d_{\vec{j}^{(1)}\vec{k}^{(1)}} + \sum_{\vec{j}^{(2)}} D^{n-1}_{\vec{j}^{(2)}} 0 j^{out(2)} N_{\vec{j}^{(2)}} d_{\vec{j}^{(2)}\vec{k}^{(1)}} \right] = 0, \quad n > 0, \tag{75}$$

$$D^n_{\vec{k}^{(2)}} = \frac{2}{\mu}\left(1 - R_0^{-1}\right)^2 \left[ \sum_{\vec{j}^{(1)}} D^{n-1}_{\vec{j}^{(1)}} P_a\left(\vec{j}^{(1)}, \vec{k}^{(2)} \mid \vec{j}^{(1)}\right) j^{out(1)} N_{\vec{j}^{(1)}} d_{\vec{j}^{(1)}\vec{k}^{(2)}} + \sum_{\vec{j}^{(2)}} D^{n-1}_{\vec{j}^{(2)}} P_a\left(\vec{j}^{(2)}, \vec{k}^{(2)} \mid \vec{j}^{(2)}\right) j^{out(2)} N_{\vec{j}^{(2)}} d_{\vec{j}^{(2)}\vec{k}^{(2)}} \right], \quad n > 0, \tag{76}$$

$$D^n_{\vec{k}^{(3)}} = \frac{2}{\mu}\left(1 - R_0^{-1}\right)^2 \left[ \sum_{\vec{j}^{(1)}} D^{n-1}_{\vec{j}^{(1)}} P_a\left(\vec{j}^{(1)}, \vec{k}^{(3)} \mid \vec{j}^{(1)}\right) j^{out(1)} N_{\vec{j}^{(1)}} d_{\vec{j}^{(1)}\vec{k}^{(3)}} + \sum_{\vec{j}^{(2)}} D^{n-1}_{\vec{j}^{(2)}} P_a\left(\vec{j}^{(2)}, \vec{k}^{(3)} \mid \vec{j}^{(2)}\right) j^{out(2)} N_{\vec{j}^{(2)}} d_{\vec{j}^{(2)}\vec{k}^{(3)}} \right], \quad n > 0. \tag{77}$$

From Eq. (75), when $n > 0$ we have that $D^n_{\vec{k}_1} = 0$, and Eqs. (76) and (77) simplify to expressions that only consider infections arriving from the transit class:

$$D^n_{\vec{k}^{(2)}} = \frac{2}{\mu}\left(1 - R_0^{-1}\right)^2 \sum_{\vec{j}^{(2)}} D^{n-1}_{\vec{j}^{(2)}} P_a\left(\vec{j}^{(2)}, \vec{k}^{(2)} \mid \vec{j}^{(2)}\right) j^{out(2)} N_{\vec{j}^{(2)}} d_{\vec{j}^{(2)}\vec{k}^{(2)}}, \quad n > 1 \tag{78}$$

$$D^n_{\vec{k}^{(3)}} = \frac{2}{\mu}\left(1 - R_0^{-1}\right)^2 \sum_{\vec{j}^{(2)}} D^{n-1}_{\vec{j}^{(2)}} P_a\left(\vec{j}^{(2)}, \vec{k}^{(3)} \mid \vec{j}^{(2)}\right) j^{out(2)} N_{\vec{j}^{(2)}} d_{\vec{j}^{(2)}\vec{k}^{(3)}}, \quad n > 1 \tag{79}$$

With the traffic functions as mentioned above, these simplify to

$$D^n_{\vec{k}^{(2)}} = \frac{2p}{\mu}\left(1 - R_0^{-1}\right)^2 \sum_{\vec{j}^{(2)}} D^{n-1}_{\vec{j}^{(2)}} P_a\left(\vec{j}^{(2)}, \vec{k}^{(2)} \mid \vec{j}^{(2)}\right) N_{\vec{j}^{(2)}}, \quad n > 1, \tag{80}$$

$$D^n_{\vec{k}^{(3)}} = \frac{2p}{\mu}\left(1 - R_0^{-1}\right)^2 \sum_{\vec{j}^{(2)}} D^{n-1}_{\vec{j}^{(2)}} P_a\left(\vec{j}^{(2)}, \vec{k}^{(3)} \mid \vec{j}^{(2)}\right) N_{\vec{j}^{(2)}}, \quad n > 1. \tag{81}$$

Inserting the quasi-equilibrium populations, $N_{\vec{j}^{(2)}}$, as derived in Appendix A yields

$$D^n_{\vec{k}^{(2)}} = \frac{2p(1-\eta_{out})\delta N}{\mu\left[(1-\eta_{out})\delta + \eta_{in} p\right]\left\langle k^{in}\right\rangle}\left(1 - \frac{1}{R_0}\right)^2 \sum_{\vec{j}^{(2)}} D^{n-1}_{\vec{j}^{(2)}} P_a\left(\vec{j}^{(2)}, \vec{k}^{(2)} \mid \vec{j}^{(2)}\right) j^{in(2)}, \quad n > 1, \tag{82}$$

$$D^n_{\vec{k}^{(3)}} = \frac{2p(1-\eta_{out})\delta N}{\mu\left[(1-\eta_{out})\delta + \eta_{in} p\right]\left\langle k^{in}\right\rangle}\left(1 - \frac{1}{R_0}\right)^2 \sum_{\vec{j}^{(2)}} D^{n-1}_{\vec{j}^{(2)}} P_a\left(\vec{j}^{(2)}, \vec{k}^{(3)} \mid \vec{j}^{(2)}\right) j^{in(2)}, \quad n > 1. \tag{83}$$

As we are considering uncorrelated networks here,

$$P_a\left(\vec{j}^{(2)}, \vec{k}^{(2)} \mid \vec{j}^{(2)}\right) = P^{in}_a\left(\vec{k}^{(2)}\right) = \frac{k^{in(2)} P_v\left(\vec{k}^{(2)}\right)}{\left\langle k^{in}\right\rangle}, \tag{84}$$

$$P_a\left(\vec{j}^{(2)}, \vec{k}^{(3)} \mid \vec{j}^{(2)}\right) = P^{in}_a\left(\vec{k}^{(3)}\right) = \frac{k^{in(3)} P_v\left(\vec{k}^{(3)}\right)}{\left\langle k^{in}\right\rangle}. \tag{85}$$

Inserting Eqs. (84) and (85) simplifies the branching processes to

$$D^n_{\vec{k}^{(2)}} = \frac{2p(1-\eta_{out})\delta \overline{N}}{\mu\left[(1-\eta_{out})\delta + \eta_{in} p\right]\left\langle k^{in}\right\rangle^2}\left(1 - \frac{1}{R_0}\right)^2 P_v\left(\vec{k}^{(2)}\right) k^{in(2)}$$

$$\times \sum_{\vec{j}^{(2)}} D^{n-1}_{\vec{j}^{(2)}} j^{in(2)}, \quad n>1, \tag{86}$$

$$D^{n}_{\vec{k}^{(3)}} = \frac{2p(1-\eta_{out})\delta\overline{N}}{\mu\left[(1-\eta_{out})\delta+\eta_{in}p\right]\left\langle k^{in}\right\rangle^2}\left(1-\frac{1}{R_0}\right)^2 P_v\left(\vec{k}^{(3)}\right)k^{in(3)} \times \sum_{\vec{j}^{(2)}} D^{n-1}_{\vec{j}^{(2)}} j^{in(2)}, \quad n>1. \tag{87}$$

Let us define, similar to the undirected analysis of Balcan and Vespignani (2012),

$$\Theta_2^n = \sum_{\vec{j}^{(2)}} D^{n}_{\vec{j}^{(2)}} j^{in(2)} \quad \text{and} \quad \Theta_3^n = \sum_{\vec{j}^{(3)}} D^{n}_{\vec{j}^{(3)}} j^{in(3)}. \tag{88}$$

Combining these definition with equations (86) and (87), we have

$$\Theta_2^n = \sum_{\vec{j}^{(2)}} D^{n}_{\vec{j}^{(2)}} j^{in(2)} = \frac{2p(1-\eta_{out})\delta\overline{N}}{\mu\left[(1-\eta_{out})\delta+\eta_{in}p\right]\left\langle k^{in}\right\rangle^2}\left(1-\frac{1}{R_0}\right)^2 \times \sum_{\vec{k}^{(2)}} P_v\left(\vec{k}^{(2)}\right)\left(k^{in(2)}\right)^2\Theta_2^{n-1}, \quad n>1 \tag{89}$$

$$\Theta_3^n = \sum_{\vec{j}^{(3)}} D^{n}_{\vec{j}^{(3)}} j^{in(3)} = \frac{2p(1-\eta_{out})\delta\overline{N}}{\mu\left[(1-\eta_{out})\delta+\eta_{in}p\right]\left\langle k^{in}\right\rangle^2}\left(1-\frac{1}{R_0}\right)^2 \times \sum_{\vec{k}^{(3)}} P_v\left(\vec{k}^{(3)}\right)\left(k^{in(3)}\right)^2\Theta_2^{n-1}, \quad n>1. \tag{90}$$

Noting that the node-type degree distributions can be expressed as $P_v\left(\vec{k}^{(2)}\right) = (1-\eta_{out}-\eta_{in})P_{v^{(2)}}\left(\vec{k}^{(2)}\right)$ and $P_v\left(\vec{k}^{(3)}\right) = \eta_{in}P_{v^{(3)}}\left(\vec{k}^{(3)}\right)$, we derive from the above equations the global epidemic invasion threshold, $R_*$, and the transit-to-sink invasion threshold, $R_*^{TS}$, respectively as

$$R_* = \frac{\Theta_2^n}{\Theta_2^{n-1}} = \frac{2p(1-\eta_{out})(1-\eta_{out}-\eta_{in})\delta\overline{N}}{\mu\left[(1-\eta_{out})\delta+\eta_{in}p\right]}\left(1-\frac{1}{R_0}\right)^2\frac{\left\langle\left(k^{in(2)}\right)^2\right\rangle}{\left\langle k^{in}\right\rangle^2}, \tag{91}$$

$$R_*^{TS} = \frac{\Theta_3^n}{\Theta_2^{n-1}} = \frac{2p(1-\eta_{out})\eta_{in}\delta\overline{N}}{\mu\left[(1-\eta_{out})\delta+\eta_{in}p\right]}\left(1-\frac{1}{R_0}\right)^2\frac{\left\langle\left(k^{in(3)}\right)^2\right\rangle}{\left\langle k^{in}\right\rangle^2}. \tag{92}$$

As seen in Eq. (92), the disease process in the sink nodes is driven by the disease process in the transit nodes. This is only the uncorrelated case, with uniform movement rates, and further work could extend this analysis to other types of networks and mobility patterns. We proceed next to consider the critical movement rates $p_c$, $p_c^{TS}$ that define the tipping points for these invasion thresholds.

### B.2. Critical movement rates

Here we extract the critical movement rates that are defined by $R_*(p_c)=1$ and $R_*^{TS}(p_c^{TS})=1$. These critical rates express the smallest movement rates necessary for the disease to spread among the transit nodes and from the transit nodes to the sink nodes, respectively, as

$$p_c = \frac{\mu\delta(1-\eta_{out})\left\langle k^{in}\right\rangle^2}{2\delta\overline{N}(1-\eta_{out})(1-\eta_{out}-\eta_{in})\left(1-\frac{1}{R_0}\right)^2\left\langle\left(k^{in(2)}\right)^2\right\rangle-\eta_{in}\mu\left\langle k^{in}\right\rangle^2}, \tag{93}$$

$$p_c^{TS} = \frac{\mu\delta(1-\eta_{out})\left\langle k^{in}\right\rangle^2}{2\delta\overline{N}(1-\eta_{out})\eta_{in}\left(1-\frac{1}{R_0}\right)^2\left\langle\left(k^{in(3)}\right)^2\right\rangle-\eta_{in}\mu\left\langle k^{in}\right\rangle^2}. \tag{94}$$

Of significant interest is the relationship between these two critical movement probabilities. If the system and disease outbreak are such that $p_c > p_c^{TS}$, then the disease process will move through both transit and sink nodes easily when the individuals' movement rate exceeds $p_c$. However, the reverse is true, three regions of movement rates will be defined wherein a second outbreak situation arises where the disease may persist within the transit nodes, but not reach the sink nodes. The relationship is then derived as follows:

$$\begin{aligned}
&p_c^{TS} > p_c \\
&\Downarrow \\
&\frac{\mu\delta(1-\eta_{out})\left\langle k^{in}\right\rangle^2}{2\delta\overline{N}(1-\eta_{out})\eta_{in}\left(1-\frac{1}{R_0}\right)^2\left\langle\left(k^{in(3)}\right)^2\right\rangle-\eta_{in}\mu\left\langle k^{in}\right\rangle^2} \\
&> \frac{\mu\delta(1-\eta_{out})\left\langle k^{in}\right\rangle^2}{2\delta\overline{N}(1-\eta_{out})(1-\eta_{out}-\eta_{in})\left(1-\frac{1}{R_0}\right)^2\left\langle\left(k^{in(2)}\right)^2\right\rangle-\eta_{in}\mu\left\langle k^{in}\right\rangle^2} \\
&\Downarrow \\
&2\delta\overline{N}(1-\eta_{out})(1-\eta_{out}-\eta_{in})\left(1-\frac{1}{R_0}\right)^2\left\langle\left(k^{in(2)}\right)^2\right\rangle-\eta_{in}\mu\left\langle k^{in}\right\rangle^2 \\
&\quad > 2\delta\overline{N}(1-\eta_{out})\eta_{in}\left(1-\frac{1}{R_0}\right)^2\left\langle\left(k^{in(3)}\right)^2\right\rangle-\eta_{in}\mu\left\langle k^{in}\right\rangle^2 \\
&\Downarrow \\
&2\delta\overline{N}(1-\eta_{out})(1-\eta_{out}-\eta_{in})\left(1-\frac{1}{R_0}\right)^2\left\langle\left(k^{in(2)}\right)^2\right\rangle \\
&\quad > 2\delta\overline{N}(1-\eta_{out})\eta_{in}\left(1-\frac{1}{R_0}\right)^2\left\langle\left(k^{in(3)}\right)^2\right\rangle \\
&\Downarrow \\
&(1-\eta_{out}-\eta_{in})\left\langle\left(k^{in(2)}\right)^2\right\rangle > \eta_{in}\left\langle\left(k^{in(3)}\right)^2\right\rangle \\
&\Downarrow \\
&\frac{\left\langle\left(k^{in(2)}\right)^2\right\rangle}{\left\langle\left(k^{in(3)}\right)^2\right\rangle} > \frac{\eta_{in}}{(1-\eta_{out}-\eta_{in})} \iff p_c^{TS} > p_c.
\end{aligned} \tag{95}$$

Note that, in this derivation, from the second line to the third, it is assumed that the denominators are positive. If the denominators were not strictly positive, the case would be that the critical rates would be either undefined or negative values, and the respective thresholds would always be greater than unity. We limited our derivation to consider only the case of positive denominators in this step. The comparison concludes that $p_c^{TS} > p_c$ if and only if the ratio of the second moments of the node-type in-degree distributions (transit nodes over sink nodes) is greater than the ratio of the sink nodes to the transit nodes.

## Appendix C. Stochastic simulation processes

Here we present the details of the numerical implementation of the dynamical processes described in the main text. We denote the number of Susceptible, Infected, and Recovered individuals in node $i$ at time $t$ respectively by $S_i(t)$, $I_i(t)$, and $R_i(t)$. The number of individuals in each of these populations is varied by discrete and stochastic dynamics describing the demographic, diffusion, and disease processes. By definition, the number of individuals in the population of node $i$ at time $t$ is given by $N_i(t) = S_i(t) + I_i(t) + R_i(t)$ .

### C.1. Disease dynamics

Susceptible individuals in the population of node $i$ at time $t$ receive infections from any infected individuals also present in node $i$ during the time interval $\Delta t$ with a probability $p_{S_i \to I_i}$ defined as

$$p_{S_i \to I_i} = \lambda_i(t)\Delta t, \tag{96}$$

where $\lambda_i(t)$ is the per capita force of infection $\lambda_i(t) = \beta I_i(t)/N_i(t)$. The number of individuals $\mathcal{Q}_i(S,I)$ transferring from the susceptible state $S_i(t)$ to the infected state $I_i(t)$ at time $t$ in node $i$ is then extracted from the binomial distribution as

$$\mathcal{Q}_i(S,I) = \Pr^{\text{Binom}}\left(S_i(t), p_{S_i \to I_i}\right). \tag{97}$$

The number of individuals $\mathcal{Q}_i(I,R)$ transferring from the infected state $I_i(t)$ to the recovered state $R_i(t)$ at time $t$ in node $i$ is similarly extracted from the binomial distribution

$$\mathcal{Q}_i(I,R) = \Pr^{\text{Binom}}\left(I_i(t), p_{I_i \to R_i}\right), \tag{98}$$

where $p_{I_i \to R_i}$ is the per capita recovery probability, $p_{I_i \to R_i} = \mu \Delta t$, within the time interval $\Delta t$. The classical SIR model contains only these two transitions. After extracting $\mathcal{Q}_i(S,R)$ and $\mathcal{Q}_i(I,R)$ for all nodes, we update the subpopulations of each node $i$ as

$$\begin{aligned} &S_i \to S_i - \mathcal{Q}_i(S,R), \\ &I_i \to I_i + \mathcal{Q}_i(S,R) - \mathcal{Q}_i(I,R), \\ &R_i \to R_i + \mathcal{Q}_i(I,R). \end{aligned} \tag{99}$$

After each iteration of the disease process, we similarly update the population dynamics.

### C.2. Population dynamics

The number of individuals $N_i(t)$ in subpopulation $i$ at time $t$ is subject to changes in the following time interval $\Delta t$ due to death, birth, importation, and migration. These events will be assumed to occur in the following order:

(1) *Death*: Each individual in node $i$ dies at rate $\delta_i$. The number of individuals who die $\mathcal{D}_i = \mathcal{D}_i^S + \mathcal{D}_i^I + \mathcal{D}_i^R$ in subpopulation $i$ within this time interval is a random integer number extracted from three binomial distributions:

$$\mathcal{D}_i = \Pr^{\text{Binom}}(S_i(t), \delta_i \Delta t) + \Pr^{\text{Binom}}(I_i(t), \delta_i \Delta t) + \Pr^{\text{Binom}}(R_i(t), \delta_i \Delta t). \tag{100}$$

After the extraction of the random numbers $\{\mathcal{D}_i\}$, we update the population sizes by

$$S_i \to S_i - \mathcal{D}_i^S; \quad I_i \to I_i - \mathcal{D}_i^I; \quad R_i \to R_i - \mathcal{D}_i^R. \tag{101}$$

The total number of individuals who are dropped out of the system due to death is then given by

$$\mathcal{D} = \sum_i \mathcal{D}_i. \tag{102}$$

Note that the death rate $\delta_i$ is subpopulation dependent and will be assumed to be equal to $\delta$ for all subpopulations of sink nodes and 0 otherwise.

(2) *Birth and importation*: We reintroduce the total number of lost individuals $\mathcal{D}$ into the system in order to keep the total population invariant over time. We consider that a fraction $p_\beta$ of these deaths are introduced back into the system as births and the rest as importations. The total number of births $\mathcal{B}$ is an integer random number extracted from the binomial distribution:

$$\mathcal{B} = \Pr^{\text{Binom}}\left(\mathcal{D}, p_\beta\right). \tag{103}$$

After extracting $\mathcal{B}$, we calculate the number of imported individuals $\mathcal{I}$ by $\mathcal{I} = \mathcal{D} - \mathcal{B}$. Once we compute $\mathcal{B}$ and $\mathcal{I}$, we distribute these individuals in the system as follows:

- *Births*: We assume that node $i$ gets a newly born individual with probability $p_{\beta_i}$ proportional to its current population:

$$p_{\beta_i} = \frac{r_{\beta_i} N_i(t)}{\sum_h r_{\beta_h} N_h(t)}, \tag{104}$$

where $r_{\beta_i}$ is going to take values of 0 or 1 and is introduced for normalization purposes. For now, we will assume that $r_{\beta_i} = 1$ for all the source nodes and 0 for the rest of the nodes. The number of newly born individuals $\mathcal{B}_i$ in node $i$ is extracted from the multinomial distribution:

$$\Pr^{\text{Multinom}}\left(\mathcal{B}, \{p_{\beta_i}\}\right). \tag{105}$$

- *Importation*: Each node $i$ receives an imported case with probability $p_{\mathcal{I}_i}$ proportional to its out-degree:

$$p_{\mathcal{I}_i} = \frac{r_{\mathcal{I}_i} k_i^{out}}{\sum_h r_{\mathcal{I}_h} l_h^{out}}, \tag{106}$$

where $r_{\mathcal{I}_i}$ is introduced for the purpose of normalization, similar to $r_{\beta_i}$, and will take values of 1 for all the source nodes and 0 otherwise. Then the total number of importations for each (source) subpopulation is given by an integer random number extracted from the multinomial distribution:

$$\Pr^{\text{Multinom}}\left(\mathcal{I}, \{p_{\mathcal{I}_i}\}\right). \tag{107}$$

After the extraction of the sets $\{\mathcal{B}_i\}$ and $\{\mathcal{I}_i\}$, the source node populations are updated with the addition of susceptible cattle by

$$S_i \to S_i + \mathcal{B}_i + \mathcal{I}_i. \tag{108}$$

In this implementation, we have assumed that imported individuals are not entering with the system with any infection. This permits the study of a single source of the outbreak, but in general this may not be an accurate representation of a livestock system.

(3) *Migration*: Outward migration occurs from both transit and source node subpopulations as they have non-zero out-degrees. Individuals in node $i$ migrate to subpopulation $j$ in the out-neighborhood $v_i^{out}$ at per capita rate $d_{ij}$:

$$d_{ij} = \frac{p}{k_i^{out}}, \quad k^{out} \neq 0 \tag{109}$$

yielding the probability of migration *pij* within the time interval $\Delta t$, $p_{ij} = d_{ij}\Delta t$. The number of individuals $\mathcal{M}_{ij}^X$ leaving node $i$ and arriving to node $j$ among each disease state $X$ is an integer random number extracted from the

multinomial distributions:

$$\Pr^{\text{Multinom}}\left(X_i, \{p_{ij} | j \in v_i^{out}\}\right). \tag{110}$$

After the extraction of all the integer numbers $\{\mathcal{M}_{ij}^X\}$, the population sizes of each subpopulation is updated according to

$$\begin{aligned} S_i &\to S_i + \sum_j \left[\mathcal{M}_{ij}^S - \mathcal{M}_{ji}^S\right], \\ I_i &\to I_i + \sum_j \left[\mathcal{M}_{ij}^I - \mathcal{M}_{ji}^I\right], \\ R_i &\to R_i + \sum_j \left[\mathcal{M}_{ij}^R - \mathcal{M}_{ji}^R\right]. \end{aligned} \tag{111}$$

## References


Anderson, R.M., May, R.M., 1984. Spatial, temporal, and genetic heterogeneity in host populations and the design of immunization programmes. IMA J. Math. Appl. Med. Biol. 1 (3), 233–266.

Bailey, N.T.J., 1975. The Mathematical Theory of Infectious Diseases and Its Applications. Charles Griffin, London, High Wycombe.

Bajardi, P., Barrat, A., Natale, F., Savini, L., Colizza, V., 2011. Dynamical patterns of cattle trade movements. PLoS One 6 (May (5)), e19869.

Bajardi, P., Barrat, A., Savini, L., Colizza, V., 2012. Optimizing surveillance for livestock disease spreading through animal movements. J. R. Soc. Interface 9 (November (76)), 2814–2825.

Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J.J., Vespignani, A., 2009. Multiscale mobility networks and the spatial spreading of infectious diseases. Proc. Natl. Acad. Sci. 106 (December), 21484–21489.

Balcan, D., Vespignani, A., 2011. Phase transitions in contagion processes mediated by recurrent mobility patterns. Nat. Phys. 7 (July (7)), 581–586.

Balcan, D., Vespignani, A., 2012. Invasion threshold in structured populations with recurrent mobility patterns. J. Theor. Biol. 293 (January), 87–100.

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

Barrat, A., Barthélemy, M., Pastor-Satorras, R., Vespignani, A., 2004. The architecture of complex weighted networks. Proc. Natl. Acad. Sci. USA 101 (March (11)), 3747–3752.

Barrat, A., Barthélemy, M., Vespignani, A., 2008. Dynamical Processes on Complex Networks. Cambridge University Press, Cambridge, UK.

Bascompte, J., Solé, R.V., 1998. Modeling Spatiotemporal Dynamics in Ecology. Springer, New York.

Battiston, S., Puliga, M., Kaushik, R., Tasca, P., Caldarelli, G., 2012. DebtRank: too central to fail? financial networks, the FED and systemic risk. Sci. Rep. 2 (August), 541.

Bigras-Poulin, M., Barfod, K., Mortensen, S., Greiner, M., 2007. Relationship of trade patterns of the danish swine industry animal movements network to potential disease spread. Prev. Vet. Med. 80 (July (2–3)), 143–165.

Bigras-Poulin, M., Thompson, R.A., Chriel, M., Mortensen, S., Greiner, M., 2006. Network analysis of Danish cattle industry trade patterns as an evaluation of risk potential for disease spread. Prev. Vet. Med. 76 (September (1–2)), 11–39.

Bolker, B., Grenfell, B., 1995. Space persistence and dynamics of measles epidemics. Philos. Trans.: Biol. Sci. 348 (May (1325)), 309–320.

Bolker, B.M., Grenfell, B.T., 1993. Chaos and biological complexity in measles dynamics. Proc. R. Soc. Lond. Biol. Sci. 251 (Janurary (1330)), 75–81.

Brockmann, D., Hufnagel, L., Geisel, T., 2006. The scaling laws of human travel. Nature 439 (January (7075)), 462–465.

Buhnerkempe, M.G., Grear, D.A., Portacci, K., Miller, R.S., Lombard, J.E., Webb, C.T., 2013. A national-scale picture of U.S. cattle movements obtained from interstate certificate of veterinary inspection data. Prev. Vet. Med., Nov 1; 112 (3–4), 318–329. http://dx.doi.org/10.1016/j.prevetmed.2013.08.002 (Epub 2013 Aug 16).

Castellano, C., Fortunato, S., Loreto, V., 2009. Statistical physics of social dynamics. Rev. Mod. Phys. 81 (April–June (2)), 591–646.

Chowell, G., Hyman, J.M., Eubank, S., Castillo-Chavez, C., 2003. Scaling laws for the movement of people between locations in a large city. Phys. Rev. E 68 (December (6)), 066102.

Colizza, V., Barrat, A., Barthelemy, M., Valleron, A.-J., Vespignani, A., 2007. Modeling the worldwide spread of pandemic influenza: vaseline case and containment interventions. PLoS Med. 4 (Janurary (1)), e13.

Colizza, V., Barrat, A., Barthélemy, M., Vespignani, A., 2006. The role of the airline transportation network in the prediction and predictability of global epidemics. Proc. Natl. Acad. Sci. USA 103 (February (7)), 2015–2020.

Colizza, V., Pastor-Satorras, R., Vespignani, A., 2007. Reaction–diffusion processes and metapopulation models in heterogeneous networks. Nat. Phys. 3 (April (4)), 276–282.

Colizza, V., Vespignani, A., 2007. Invasion threshold in heterogeneous metapopulation networks. Phys. Rev. Lett. 99 (October (14)), 148701.

Colizza, V., Vespignani, A., 2008. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. J. Theor. Biol. 251 (April (3)), 450–467.

Cooper, B.S., Pitman, R.J., Edmunds, W.J., Gay, N.J., 2006. Delaying the international spread of pandemic influenza. PLoS Med. 3 (May (6)), e212.

Cross, P.C., Johnson, P.L., Lloyd-Smith, J.O., Getz, W.M., 2007. Utility of $r_0$ as a predictor of disease invasion in structured populations. J. R. Soc. Interface 4 (April (13)), 315–324.

Cross, P.C., Lloyd-Smith, J.O., Johnson, P.L.F., Getz, W.M., 2005. Duelling timescales of host movement and disease recovery determine invasion of disease in structured populations. Ecol. Lett. 8 (6), 587–595.

Earn, D.J., Rohani, P., Grenfell, B.T., 1998. Persistence, chaos and synchrony in ecology and epidemiology. Proc. R. Soc. Biol. Sci. 265 (1390), 7–10.

Danon, L., Ford, A.P., House, T., Jewell, C.P., Keeling, M.J., Roberts, G.O., Ross, J.V., Vernon, M.C., 2011. Networks and the epidemiology of infectious disease. Interdiscip. Perspect. Infect. Dis. 2011 (March (284909)).

Ferguson, N.M., Donnelly, C.A., Anderson, R.M., 2001. The foot-and-mouth epidemic in great britain: Pattern of spread and impact of interventions. Science 292 (May (5519)), 1155–1160.

Ferguson, N.M., Keeling, M.J., John Edmunds, W., Gani, R., Grenfell, B.T., Anderson, R. M., Leach, S., 2003. Planning for smallpox outbreaks. Nature 425 (October (6959)), 681–685.

Baptista, Filipa M., Nunes, Telmo, Almeida, Virgilio, Louza, Armando, 2008. Cattle movements in Portugal—an insight into the potential use of network analysis. Rev. Port. Cienc. Vet. 107 (565–566), 35–40.

Flahault, A., Valleron, A., 1992. A method for assessing the global spread of HIV-1 infection based on air travel. Math. Popul. Stud. 3 (3), 161–171.

Gonçalves, B., Balcan, D., Vespignani, A., 2013. Human mobility and the worldwide impact of intentional localized highly pathogenic virus release. Sci. Rep. 3 (July).

González, M.C., Hidalgo, C.A., Barabási, A.-L., 2008. Understanding individual human mobility patterns. Nature 453 (June (7196)), 779–782.

Grais, R.F., Ellis, J.H., Glass, G.E., 2003. Assessing the impact of airline travel on the geographic spread of pandemic influenza. Eur. J. Epidemiol. 18 (11), 1065–1072.

Grais, R.F., Ellis, J.H., Kress, A., Glass, G.E., 2004. Modeling the spread of annual influenza epidemics in the U.S.: the potential role of air travel. Health Care Manag. Sci. 7 (May (2)), 127–134.

Green, D.M., Kiss, I.Z., Kao, R.R., 2006. Modelling the initial spread of foot-and-mouth disease through animal movements. Proc. R. Soc. B: Biol. Sci. 273 (November (1602)), 2729–2735.

Grenfell, Bolker, 1998. Cities and villages: infection hierarchies in a measles metapopulation. Ecol. Lett. 1 (1), 63–70.

Grenfell, B., Harwood, J., 1997. (meta)population dynamics of infectious diseases. Trends Ecol. Evol. 12 (October (10)), 395–399.

Guimerà, R., Mossa, S., Turtschi, A., Amaral, L.A.N., 2005. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. Proc. Natl. Acad. Sci. 102 (May (22)), 7794–7799.

Hanski, I., Gaggiotti, O.E., 2004. Ecology, Genetics, and Evolution of Metapopulations. Elsevier, Academic Press, Amsterdam.

Hanski, I.A., Gilpin, M.E., 1997. Metapopulation Biology: Ecology, Genetics, and Evolution. Academic Press, San Diego.

Harris, T.E., 1989. The Theory of Branching Processes. Courier Dover Publications, Mineola, New York.

Hethcote, H.W., 1978. An immunization model for a heterogeneous population. Theor. Popul. Biol. 14 (December (3)), 338–349.

Hollingsworth, T.D., Ferguson, N.M., Anderson, R.M., 2006. Will travel restrictions control the international spread of pandemic influenza? Nat. Med. 12 (May (5)), 497–499.

Hsu, W.H., Roy Chowdhury, S., Scoglio, C., 2011. Mitigation strategies for foot and mouth disease: A learning-based approach. Int. J. Artif. Life Res. 2 (April (2)), 42–76.

Hufnagel, L., Brockmann, D., Geisel, T., 2004. Forecast and control of epidemics in a globalized world. Proc. Natl. Acad. Sci. USA 101 (Octobe (42)), 15124–15129.

Kao, R., Danon, L., Green, D., Kiss, I., 2006. Demographic structure and pathogen dynamics on the network of livestock movements in great britain. Proc. R. Soc. B: Biol. Sci. 273 (August (1597)), 1999–2007.

Kao, R.R., Green, D.M., Johnson, J., Kiss, I.Z., 2007. Disease dynamics over very different time-scales: foot-and-mouth disease and scrapie on the network of livestock movements in the UK. J. R. Soc. Interface 4 (October (16)), 907–916.

Keeling, M.J., 2000. Metapopulation moments: coupling, stochasticity and persistence. J. Anim. Ecol. 69 (September (5)), 725–736.

Keeling, M.J., Danon, L., Vernon, M.C., House, T.A., 2010. Individual identity and movement networks for disease metapopulations. Proc. Natl. Acad. Sci. 107 (May (19)), 8866–8870.

Keeling, M.J., Rohani, P., 2002. Estimating spatial coupling in epidemiological systems: a mechanistic approach. Ecology Letters 5 (1), 20–29.

Keeling, M.J., Rohani, P., 2011. Modeling Infectious Diseases in Humans and Animals. Princeton University Press, Princeton September.

Keeling, M.J., Woolhouse, M.E.J., Shaw, D.J., Matthews, L., Chase-Topping, M., Haydon, D.T., Cornell, S.J., Kappey, J., Wilesmith, J., Grenfell, B.T., 2001. Dynamics of the 2001 UK foot and mouth epidemic: Stochastic dispersal in a heterogeneous landscape. Science 294 (October (5543)), 813–817.

Lentz, H.H.K., Selhorst, T., Sokolov, I.M., 2012. Spread of infectious diseases in directed and modular metapopulation networks. Phys. Rev. E 85 (June), 066111.

Liu, S., Perra, N., Karsai, M., Vespignani, A., 2014. Controlling contagion processes in activity driven networks. Phys. Rev. Lett. 112 (March (11)), 118702.

Liu, S.-Y., Baronchelli, A., Perra, N., 2013. Contagion dynamics in time-varying metapopulation networks. Phys. Rev. E 87 (March (3)), 032805.
Lloyd, A.L., May, R.M., 1996. Spatial heterogeneity in epidemic models. J. Theor. Biol. 179 (March (1)), 1–11.
Lloyd, A.L., May, R.M., 2001. How viruses spread among computers and people. Science 292 (May (5520)), 1316–1317.
Longini Jr., I.M., 1988. A mathematical model for predicting the geographic spread of new infectious agents. Math. Biosci. 90 (July (1–2)), 367–383.
May, R.M., Anderson, R.M., 1979. Population biology of infectious diseases: Part II. Nature 280 (August (5722)), 455–461.
May, R.M., Anderson, R.M., 1984. Spatial heterogeneity and the design of immunization programs. Math. Biosci. 72 (November (1)), 83–111.
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 (December (4)), 341–350.
Parikh Nidhi, Youssef Mina, Swarup Samarth, Eubank Stephen, 2013. Modeling the effect of transient populations on epidemics in washington DC. Submitted, http://dx.doi.org/10.1038/srep03152.
Park, A.W., Gubbins, S., Gilligan, C.A., 2002. Extinction times for closed epidemics: the effects of host spatial structure. Ecol. Lett. 5 (6), 747–755.
Pastor-Satorras, R., Vespignani, A., 2001. Epidemic spreading in scale-free networks. Phys. Rev. Lett. 86 (April (14)), 3200–3203.
Patuelli, R., Reggiani, A., Gorman, S.P., Nijkamp, P., Bade, F.-J., 2007. Network analysis of commuting flows: a comparative static approach to german data. Netw. Spat. Econ. 7 (December (4)), 315–331.
Poletto, C., Meloni, S., Colizza, V., Moreno, Y., Vespignani, A., 2013. Host mobility drives pathogen competition in spatially structured populations. PLoS Comput. Biol. 9 (August (8)), e1003169.
Poletto, C., Tizzoni, M., Colizza, V., 2012. Heterogeneous length of stay of hosts' movements and spatial epidemic spread. Sci. Rep. 2 (June (476)).
Rautureau, S., Dufour, B., Durand, B., 2011. Vulnerability of animal trade networks to the spread of infectious diseases: a methodological approach applied to evaluation and emergency control strategies in cattle, France, 2005. Transbound. Emer. Dis. 58 (April (2)), 110–120.
Rvachev, L.A., Longini Jr., I.M., 1985. A mathematical model for the global spread of influenza. Math. Biosci. 75 (July (1)), 3–22.
Riley, S., 2007. Large-scale spatial-transmission models of infectious disease. Science 316 (June (5829)), 1298–1301.
Robinson, S.E., Christley, R.M., 2007. Exploring the role of auction markets in cattle movements within great britain. Prev. Vet. Med. 81 (September (1–3)), 21–37.
Rohani, P., Earn, D.J.D., Grenfell, B.T., 1999. Opposite patterns of synchrony in sympatric disease metapopulations. Science 286 (October (5441)), 968–971.
Ruan, S., Wang, W., Levin, S.A., 2006. The effect of global travel on the spread of sars. Math. Biosci. Eng. 3 (January (1)), 205–218.
Sattenspiel, L., Dietz, K., 1995. A structured epidemic model incorporating geographic mobility among regions. Math. Biosci. 128 (August (1–2)), 71–91.
Schweitzer, F., Fagiolo, G., Sornette, D., Vega-Redondo, F., Vespignani, A., White, D. R., 2009. Economic networks: the new challenges. Science 325 (July (5939)), 422–425.
Taylor, C., Marathe, A., Beckman, R., 2010. Same influenza vaccination strategies but different outcome across US cities? Int. J. Infect. Dis.: Official Publication of the International Society for Infectious Diseases 14 (September (9)), e792–e795.
Tilman, D., Kareiva, P.M., 1997. Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions. Princeton University Press, Princeton.
Vazquez, A., 2006. Polynomial growth in branching processes with diverging reproductive number. Phys. Rev. Lett. 96 (January (3)), 038702.
Vazquez, A., 2007. Epidemic outbreaks on structured populations. J. Theor. Biol. 245 (March (1)), 125–129.
Vespignani, A., 2009. Predicting the behavior of techno-social systems. Science 325 (July (5939)), 425–428.
Vespignani, A., 2012. Modelling dynamical processes in complex socio-technical systems. Nat. Phys. 8 (Janurary (1)), 32–39.
Xue, L., Cohnstaedt, L.W., Scott, H.M., Scoglio, C., 2013. A hierarchical network approach for modeling rift valley fever epidemics with applications in North America. PLoS One 8 (May (5)), e62049.
Xue, L., Scott, H.M., Cohnstaedt, L.W., Scoglio, C., 2012. A network-based metapopulation approach to model rift valley fever epidemics. J. Theor. Biol. 306 (August), 129–144.
Zhou, J., Xiao, G., Wong, L., Fu, X., Ma, S., Cheng, T.H., 2010. Generation of arbitrary two-point correlated directed networks with given modularity. Phys. Lett. A 374 (July (31–32)), 3129–3135.