scientific reports

OPEN

# New approaches to epidemic modeling on networks

Arturo Gómez[1] & Gonçalo Oliveira[2,3]✉

**In this article, we develop two independent and new approaches to model epidemic spread in a network. Contrary to the most studied models, those developed here allow for contacts with different probabilities of transmitting the disease (transmissibilities). We then examine each of these models using some mean field type approximations. The first model looks at the late-stage effects of an epidemic outbreak and allows for the computation of the probability that a given vertex was infected. This computation is based on a mean field approximation and only depends on the number of contacts and their transmissibilities. This approach shares many similarities with percolation models in networks. The second model we develop is a dynamic model which we analyze using a mean field approximation which highly reduces the dimensionality of the system. In particular, the original system which individually analyses each vertex of the network is reduced to one with as many equations as different transmissibilities. Perhaps the greatest contribution of this article is the observation that, in both these models, the existence and size of an epidemic outbreak are linked to the properties of a matrix which we call the $\mathcal{R}$-matrix. This is a generalization of the basic reproduction number which more precisely characterizes the main routes of infection.**

**Context.** A very natural way to model the spread of a human-to-human transmissible infectious disease is to encode each individual as the vertex of a graph whose edges model the interactions through which the disease can propagate. See[1–9] and references therein for the vast literature of epidemic modeling, including on networks. We also refer the reader to some very interesting related work in[10–14].

However, despite the large body of work, there are substantial difficulties in implementing such methods, the most obvious of which being the difficulty in inferring a realistic network and in analyzing the very high dimensional resulting system of ordinary differential equations. Furthermore, due to such difficulties most models make the additional simplifying assumption that all interactions have the same probability of transmitting the disease.

In fact, extending the theory in order to incorporate interactions with different probability of transmitting the disease, dealing with heterogeneity, developing approximation schemes, and understanding network based interventions are all listed as some of the main challenges facing network epidemic modeling as stated in[7].

These challenges are in fact all linked as, for instance, understanding which interactions are the most responsible for the epidemic spread would allow for better insight on which kind of interventions are the most effective in controlling an outbreak. This is therefore a fundamental research direction which ought to be pursued with more intensity in the future.

In this work we shall use a method that at the same time deals with the two difficulties mentioned above while at the same time incorporating the possibility of different types of interactions. Our results are insightful and our techniques tractable enough so that they can be effectively used in the future in a large amount of situations.

**Summary of results.** We shall now summarize our approach and main results. Consider a large number of individuals $N \in \mathbb{N}$ interacting with each other through $n \in \mathbb{N}$ different types of interactions which have probabilities $T_1, \dots, T_n \in [0, 1]$ of transmitting the disease (typically $1 \leq n \ll N$). To encode the network we use $n$ different graphs $\mathcal{G}_1, \dots, \mathcal{G}_n$ whose edges represent the different interactions and $\mathcal{G} = \mathcal{G}_1 \cup \cdots \cup \mathcal{G}_n$ is the total graph. For example, an edge of the graph $\mathcal{G}_i$ encodes an interaction which has a probability $T_i$ of transmitting the disease.

In reality, we shall require relatively little information on the specific properties of the network encoding the interactions. Namely, we will only need to know the degree distributions for the several types of interactions. Said

[1]Universidade Federal Fluminense, Rua Miguel de Frias, 9, Icaraí, Niterói, RJ 24220-900, Brasil. [2]IST Austria, Am Campus 1, 3400 Klosterneuburg, Austria. [3]Instituto Superior Tecnico, Lisboa, Portugal. ✉email: galato97@gmail.com

in other words, we require knowledge of the probability that a randomly chosen vertex has a certain number of interactions of each type (notice that this is far less than knowing the exact form of the network).

*Summary of methodology and main contribution.* The key technical method which we employ to deal with this is to use multivariate generating functions in order to simplify computations and have a unified approach which only depend on the degree distributions of the graphs encoding the network. Therefore, we start in section "Generating functions" by recalling the definition multivariate generating functions for the excess degree distributions following a random edge of $\mathcal{G}_i$, for $i = 1, \dots, n$. These are denoted by $\tilde{G}_i(x_1, \dots, x_n)$ and we further use them to construct a $n \times n$ matrix with entries

$$\mathcal{R}^{\tilde{G}}_{ij} = T_j \frac{\partial \tilde{G}_i}{\partial x_j}\Big|_{x_1=\cdots=x_n=1},$$

which we shall call the $\mathcal{R}^{\tilde{G}}$-matrix. This matrix has several interesting properties and, as will become later clear, encodes much epidemic information. For example, the sum of all entries in the $i$-th line $R_0(i) = \sum_j \mathcal{R}^{\tilde{G}}_{ij}$ coincides with the basic reproduction number of infections caused by the individuals that originally got infected through an interaction of $\mathcal{G}_i$. Also, the total basic reproduction number can be easily recovered from $\mathcal{R}^{\tilde{G}}_{ij}$ as shown in Remark 8.

Still in section "Generating functions" we define some modified generating functions, which we call $\tilde{H}_i(x_1, \dots, x_n)$, and use to construct a $\mathcal{R}^{\tilde{H}}$-matrix similarly.

Section "Prevalence of infection and percolation (the late stages of an outbreak)" constructs the first, and most basic, of our approximate models. This looks at the late stages of an epidemic outbreak that propagated in $\mathcal{G}$ and assumes that the disease already had enough time to sufficiently spread through in the population and came to some sort of equilibrium with part of the population being removed after infection and transmission. We employ this into a mean field type approximation which, in particular, implies that all individuals with the same number of interactions of each kind, have the same probability of already having been infected. Our analysis can also be considered as a model of percolation with $n$ different types of nodes. This important problem in itself, seems to not have received much attention, see[15] for a honorable exception.

Our findings, stated in Propositions 1, 2, and 3, relate the existence of a phase transition in the fraction of infected vertices and the eigenvalues of the $\mathcal{R}^{\tilde{G}}$-matrix crossing the value 1.

Section "Dynamic modeling in a local mean field approximation" lays out an enormous system of ordinary differential equations (ODEs) modeling a SIR-type epidemic dynamically spreading in a network encoded by $\mathcal{G}$. This model generalizes the more standard version by allowing for different transmissibilities depending on which graph $\mathcal{G}_i$ encodes a specific interaction. In total, this results in a system of $N$ ODEs for $N$ unknown functions.

The analysis of this model is mostly postponed until section "Dynamic modeling in a mean field approximation by degree similarity" , with 4 simply proving that as $t \to +\infty$, the system converges to a disease free equilibrium for which some characterizations are given. For example, Theorem 1 shows that a vertex $v \in \mathcal{G}$ has a probability of at most $1 - \exp(-R(v))$ of ever being infected, where $R(v)$ denotes the number of infections $v$ is expected to cause if infected. Interestingly, this simple bound highlights that individuals expected to infect many others are also more likely to be infected.

Section "Dynamic modeling in a mean field approximation by degree similarity" further analyses the previous system of ODEs by making one extra simplifying assumption. Namely, that vertices with the same joint degree have similar probabilities of being in each state. We then show that such a simplifying assumption reduces the original system of $N$ ODEs to a much smaller one of $n$ ODEs. In this situation, we find that the matrix $\mathcal{R}^{\tilde{H}}$ is of fundamental importance in understanding the dynamics of the epidemic outbreak. For example, we prove that the existence of an eigenvalue greater than one is related with the nearly exponential growth of the outbreak in its early stages.

*A comment on computational methods.* Before embracing in proving the main results we want to make one further comment. There are computational tools which can be used to implement epidemics spreading on networks such as the EoN and SEIRSplus Python packages. While these do exemplify the well known epidemic behaviors that we describe, our main contribution is to rigorously mathematically demonstrate the mentioned results using the very general models we consider (recall that we allow for $n \geq 1$ different transmissibilities). This goes beyond the current state of the art as these results had only been rigorously established in the case $n = 1$.

## Generating functions

This section reviews some basics of generating functions, including multivariate generating functions. For some fascinating early and varied applications of the method of generating functions we recommend[16].

**Graphs and their generating functions.** Let $\mathcal{G}$ be a graph whose vertices $v(\mathcal{G})$ encode individuals and whose edges $e(\mathcal{G})$ encode interactions through which the disease can spread. We consider a family of $n \in \mathbb{N}$ subgraphs $\mathcal{G}_1, \dots, \mathcal{G}_n \subset \mathcal{G}$ with $v(\mathcal{G}_i) = v(\mathcal{G})$, $e(\mathcal{G}_i) \subset e(\mathcal{G})$ for $i = 1, \dots, n$ and $e(\mathcal{G}_1) \cup \cdots \cup e(\mathcal{G}_n) = e(\mathcal{G})$. For each $i = 1, \dots, n$, let $\{P_i(k)\}_{k \in \mathbb{N}_0}$ be the degree distribution of the graph $\mathcal{G}_i$ meaning that a randomly chosen vertex has degree $k$ with probability $P_i(k)$. The corresponding generating functions are given by

$$G_i(x) = \sum_{k=0}^{+\infty} P_i(k) x^k, \quad \text{for } i = 1, \dots, n,$$

and we define the joint multivariate generating function by

$$G(x_1,\dots,x_n)=\sum_{k_1=0}^{+\infty}\dots\sum_{k_n=0}^{+\infty}\prod_{i=1}^{n}P_i(k_i)x_i^{k_i}=\prod_{i=1}^{n}G_i(x_i).$$

***Remark 1*** Notice that $G_i'(x)=\sum_{k=1}^{+\infty}kP_i(k)x^{k-1}$ and so the average degree of $\mathcal{G}_i$, denoted $\mathbf{E}(k_i)$, may be computed to be

$$\mathbf{E}(k_i)=G_i'(1).$$

Similarly, using the fact that $G_i(1)=1$, we find the total average degree to be

$$\mathbf{E}(k)=\sum_{i=1}^{n}G_i'(1)=\frac{d}{dx}G(x,\dots,x)\Big|_{x=1}.$$

***Remark 2*** The idea of having the $n$ subgraphs $\mathcal{G}_1,\dots,\mathcal{G}_n$ is that of modeling different types of contacts between individuals which have unequal probability of transmitting the disease. Thus, we associate a probability of transmission $T_i$ to each $i=1,\dots,n$ and assume with no loss of generality that $T_1>\dots>T_n$.

**Excess degree distributions.** Consider a randomly chosen individual which may have been infected following a randomly chosen transmission. Considering it as a vertex in $\mathcal{G}_i$, we define its excess degree (or ramification) as the number of extra edges emanating from it. The joint probability that such a vertex has ramification $(k_1,\dots,k_n)$ along the graphs $\mathcal{G}_1,\dots,\mathcal{G}_n$ can be computed directly from the degree distributions $P_i(\cdot)$ of the graphs $\mathcal{G}_i$ as in[17]. Indeed, given that fixing the subgraph $\mathcal{G}_i$ there are $k_i+1$ ways of arriving at a vertex with degree $k_i+1$ (ramification $k_i$) we find that

$$\tilde{P}(k_1,\dots,k_n)=\frac{\sum_{l=1}^{n}(k_l+1)P_l(k_l+1)\prod_{i\neq l}P_i(k_i)}{\mathbf{E}(k)}.$$

This will be referred to as the excess degree, or ramification, joint distribution.

***Remark 3*** Here we are working with the excess degree distribution for a randomly chosen individual rather than than the excess degree distribution for a randomly chosen infected individual.

The associated multivariate generating function

$$\tilde{G}(x_1,\dots,x_n):=\sum_{k_1,\dots,k_n}\tilde{P}(k_1,\dots,k_n)x_1^{k_1}\dots x_n^{k_n}, \tag{1}$$

which as shown in[17] can be written in terms of the $G_i(x_i)$ for $i=1,\dots,n$ as follows

$$\tilde{G}(x_1,\dots,x_n)=\frac{\sum_{l=1}^{n}G_l'(x_l)\prod_{i\neq l}G_i(x_i)}{\sum_{l=1}^{n}G_l'(1)}, \tag{2}$$

and recall that $\mathbf{E}(k)=\sum_{l=1}^{n}G_l'(1)$.

***Remark 4*** Let $m_1,\dots,m_n\in\mathbb{N}_0$ with $m_1+\dots+m_n>0$ and consider a randomly chosen individual $v$ and assume that the remaining population is all susceptible. Using the distribution for the excess ramification, the probability that, if infected, $v$ will infect $m_i$ individuals along $\mathcal{G}_i$ for each $i$

$$p(m_1,\dots,m_n)=\sum_{k_1\geq m_1,\dots,k_n\geq m_n}\tilde{P}(k_1,\dots,k_n)\prod_{i=1}^{n}\binom{k_i}{m_i}T_i^{m_i}(1-T_i)^{k_i-m_i}. \tag{3}$$

Then, the generating function for the random variable $M$ given by the number of infections caused by a randomly chosen individual, if infected, is

$$\begin{aligned}G_M(x)&=\sum_{m_1,\dots,m_n}p(m_1,\dots,m_n)x^{m_1+\dots+m_n}\\&=\sum_{k_1,\dots,k_n}\sum_{m_1=0,\dots,m_n=0}^{k_1,\dots,k_n}\tilde{P}(k_1,\dots,k_n)\prod_{i=1}^{n}\binom{k_i}{m_i}(xT_i)^{m_i}(1-T_i)^{k_i-m_i}\\&=\sum_{k_1,\dots,k_n}\tilde{P}(k_1,\dots,k_n)\prod_{i=1}^{n}(xT_i+1-T_i)^{k_i}\\&=\tilde{G}(1+T_1(x-1),\dots,1+T_n(x-1)).\end{aligned} \tag{4}$$

***Remark 5*** Recall that, for a disease starting to propagate in an otherwise completely susceptible population, the basic reproduction number$R_0$ is given by $R_0 = G'_M(1)$, which from the above formula can be computed to be

$$R_0 = \sum_{j=1}^{n} T_j \frac{\partial \tilde{G}}{\partial x_j}\Big|_{x_1=\cdots=x_n=1}.$$

**Specific excess degree distributions.** We now consider the excess degree distribution by following an edge of a specific graph $\mathcal{G}_i$, we define its available ramification as the number of excess edges emanating from it. Then, as before, the joint probability that such a vertex has excess degree $(k_1, \ldots, k_n)$ is

$$\tilde{P}_i(k_1, \ldots, k_n) = \frac{(k_i+1)P_i(k_i+1)\prod_{j\neq i} P_j(k_j)}{\mathbf{E}(k_i)}.$$

Therefore, its multivariate generating function $\tilde{G}_i(x_1, \ldots, x_n) = \sum_{k_1,\ldots,k_n} \tilde{P}_i(k_1, \ldots, k_n)x_1^{k_1} \ldots x_n^{k_n}$ can be computed to be

$$\tilde{G}_i(x_1, \ldots, x_n) = \frac{G'_i(x_i)}{G'_i(1)} \prod_{j\neq i} G_j(x_j). \tag{5}$$

***Remark 6*** Notice, in particular that $\tilde{P}(k_1, \ldots, k_n) = \sum_{i=1}^{n} \frac{\mathbf{E}(k_i)}{\mathbf{E}(k)} \tilde{P}_i(k_1, \ldots, k_n)$. Using either this fact or the previous formulas for the generating functions we find that $\tilde{G}$ can be computed from the $\tilde{G}_i$ as follows

$$\tilde{G}(x_1, \ldots, x_n) = \frac{\sum_{l=1}^{n} G'_l(1)\tilde{G}_l(x_1, \ldots, x_n)}{\sum_{l=1}^{n} G'_l(1)}.$$

**Definition 1** Let $i = 1, \ldots, n$ and $\tilde{G}_i(x_1, \ldots, x_x)$ be as in (5). Define the $\mathcal{R}^{\tilde{G}}$-matrix as the $n \times n$ matrix whose $(i, j)$ entries are

$$\mathcal{R}_{ij}^{\tilde{G}} := T_j \frac{\partial \tilde{G}_i}{\partial x_j}\Big|_{x_1=\cdots=x_n=1}.$$

Furthermore, given $s_{(k_1,\ldots k_n)} \in [0,1]$ for all $(k_1, \ldots k_n) \in \mathbb{N}_0^n$ and $i = 1, \ldots, n$ we shall define the generating function

$$\tilde{H}_i(x_1, \ldots, x_n) = \sum_{k_1,\ldots,k_n} \tilde{P}_i(k_1, \ldots, k_n)s_{(k_1,\ldots,k_i+1,\ldots k_n)}x_1^{k_1} \ldots x_i^{k_i+1} \ldots x_n^{k_n}. \tag{6}$$

It will prove useful to also define a $\mathcal{R}$-matrix associated with these generating functions as follows.

**Definition 2** Let $s = \{s_{(k_1,\ldots k_n)} \in [0,1] \mid (k_1, \ldots k_n) \in \mathbb{N}_0^n\}$ and for $i = 1, \ldots, n$ the function $\tilde{H}_i(x_1, \ldots, x_x)$ be as in (6). Then, we define the $n \times n$ matrix whose $(i, j)$ entries are

$$\mathcal{R}_{ij}^{\tilde{H}} := T_j \frac{\partial \tilde{H}_i}{\partial x_j}\Big|_{x_1=\cdots=x_n=1},$$

which we shall call the $\mathcal{R}^{\tilde{H}}$-matrix.

***Remark 7*** In the case when $s_{(k_1,\ldots,k_n)} = 1$ we have $\tilde{H}_i = x_i\tilde{G}_i$ and so the $\mathcal{R}^{\tilde{H}}$-matrix turns into

$$\mathcal{R}_{ij}^{\tilde{H}} := T_i\delta_{ij} + T_j \frac{\partial \tilde{G}_i}{\partial x_j}\Big|_{x_1=\cdots=x_n=1} = T_i\delta_{ij} + \mathcal{R}_{ij}^{\tilde{G}}.$$

***Remark 8*** We can define a quantity $R_0(i)$ which yields the average number of infections caused by the individuals which were themselves infected from an interaction of $\mathcal{G}_i$. In formulas, such a quantity is given by

$$R_0(i) := \sum_{j=1}^{n} T_j \frac{\partial \tilde{G}_i}{\partial x_j}\Big|_{x_1=\cdots=x_n=1}, \tag{7}$$

or, in terms of the first reproduction matrix, as the sum of all the entries in the i-th line, i.e. $R_0(i) = \sum_{j=1}^{n} \mathcal{R}_{ij}^{\tilde{G}}$. It is easy to check that

$$R_0 = \sum_{i=1}^{n} \frac{\mathbf{E}(k_i)}{\mathbf{E}(k)} R_0(i).$$

## Prevalence of infection and percolation (the late stages of an outbreak)

Recall that for $i = 1, \dots, n$ the transmissibilities $T_i$ denote the probability that an interaction, encoded by an edge of $\mathcal{G}_i$, will transmit the disease if one of its ends is infected. We point out that this approach is in different from that of multiplex networks as in that case it is the different nodes that have different transmissibilities, see for example[18–20]. In our case each node can have several $T_i$ and so our approach can be interpreted as a generalization of the multiplex networks case. We also point out that we shall work in full generality and our results are widely applicable.

At this point we follow a well known trick first introduced in[1] and[21] in the case when there is only one transmissibility. In order to implement this trick in the case when there is more than one transmissibility we introduce $n$ quantities, called $q_1, \dots, q_n$: Each $q_i$ corresponds to the average probability that a vertex is not infected through a specific interaction (edge) of $\mathcal{G}_i$. For this to happen, either:

- the infection is not transmitted (independently of whether the individual in the other end of this interaction is infected or not) which has probability $1 - T_i$, or
- the infection would be transmitted by the interaction, with probability $T_i$, but the other individual was not infected, which happens with probability $\prod_{j=1}^n q_j^{k_j}$ if it has excess degree $(k_1, \dots, k_n)$.

Hence, on average we have

$$q_i = 1 - T_i + T_i \sum_{k_1,\dots,k_n} \tilde{P}_i(k_1, \dots, k_n) \prod_{j=1}^{n} q_j^{k_j} = 1 - T_i + T_i \tilde{G}_i(q_1, \dots, q_n), \tag{8}$$

for $i = 1, \dots, n$. This is a fixed point equation for the function $F(q_1, \dots, q_n) = (F_1, \dots, F_n)$ with $F_i(q_1, \dots, q_n)$ given by the right hand side of Eq. (8).

Then, the average probability that a randomly chosen vertex does not get infected is given by

$$\mathbf{P} = \sum_{k_1,\dots,k_n} P(k_1, \dots, k_n) q_1^{k_1} \dots q_n^{k_n} = G(q_1, \dots, q_n). \tag{9}$$

***Remark 9*** Notice that the probability a vertex with degree $(k_1, \dots, k_n)$ does not get infected is given by

$$\mathbf{P}_{(k_1,\dots,k_n)} = q_1^{k_1} \dots q_n^{k_n},$$

and so $\mathbf{P}$ is simply the average of these probabilities. So, we see that this mean field type approximation also implies the weaker approximation where the probability that a vertex gets or not infected only depends on its degree.

We turn now to the question of finding conditions which guarantee that (8) has a solution other than the obvious one at $(1, \dots, 1)$, which corresponds to the absence of disease. Before we embrace in the general analysis we consider the simple special case when $n = 1$ which already appears in the literature, for example chapter 16 in[2].

***Example 1*** In the case when $n = 1$ the fixed point equation (8) reads

$$q = 1 - T + T\tilde{G}(q). \tag{10}$$

Denoting the right hand side by $F(q)$, we have $F(1) = 1$ and $F(0) = 1 - T + T\tilde{G}(0) \geq 1 - T > 0$ while $F'(q) = T\tilde{G}'(q) > 0$ and $F''(q) = T\tilde{G}''(q) \geq 0$. It then follows from the intermediate value theorem that there is a fixed point $q \in (0, 1)$ of $F$ if and only if $F'(1) > 1$. Such a condition is given by $T\tilde{G}'(q) > 1$ which can equally be written as $R_0 > 1$.

In this simple setting when $n = 1$, we can further try to better understand the transition phenomena at $T = T_c$ such that $T_c\tilde{G}'(1) = 1$. For this we expand $\tilde{G}$ around $q_c = 1$ as a Taylor series

$$\tilde{G}(x) = 1 + \tilde{G}'(1)(1 - q) + \frac{\tilde{G}''(1)}{2}(1 - q)^2 + \cdots,$$

and inserting into Eq. (10) we have

$$\begin{aligned} q &= 1 - T + T\left(1 + \tilde{G}'(1)(q - 1) + \frac{\tilde{G}''(1)}{2}(q - 1)^2\right) + \cdots \\ &= 1 + (q - 1)\frac{T}{T_c} + (q - 1)^2 \frac{T\tilde{G}''(1)}{2} + \cdots, \end{aligned}$$

where we have used $\tilde{G}'(1) = T_c^{-1}$. This can be rewritten as

$$1 = \frac{T}{T_c} - (1 - q)\frac{T\tilde{G}''(1)}{2},$$

from which we find

$$q = 1 - \frac{2}{T\tilde{G}''(1)}\left(\frac{T}{T_c} - 1\right) + \cdots .$$

Then, expanding $\mathbf{P}$ in a Taylor series around $q_c = 1$ we find

$$\begin{aligned}\mathbf{P} &= 1 + (q-1)G'(1) + \cdots \\ &= 1 - \frac{2G'(1)}{\tilde{G}''(1)}\frac{1}{T}\left(\frac{T}{T_c} - 1\right) + \cdots ,\end{aligned} \tag{11}$$

valid for $T \geq T_c$ and which describes the phase transition as a power law with exponent 1.

Continuing to explore the case when $n = 1$ we shall now give two very simple examples which can be solved explicitly.

***Example 2*** (*2 neighbors and* $n = 1$) In this situation each individual contacts with only two other ones, we have $G(x) = x^2$ and $\tilde{G}(x) = x$. Then, the fixed point equation is $q = 1 - T + Tq$ which has the unique solution $q = 1$ independently of $T \in (0,1)$. The only other solution is $q = 0$ which occurs in the case when $T = 1$.

This is to be expected as if there is a probability that the interactions will not transmit the disease, then almost surely there will be someone which does not transmit it and so it does not get passed that individual. In a large population, almost everyone will be left uninfected.

***Example 3*** (*3 neighbors and* $n = 1$) In this example we consider $n = 1$ and $G(x) = x^3$ so $\tilde{G}(x) = x^2$. Then, the fixed point equation turns into $q = 1 - T + Tq^2$. The only solution $T \in (0,1)$ is given by

$$q = \frac{1 - |2T - 1|}{2T} = \begin{cases} 1, & \text{if } T \leq \frac{1}{2}, \\ \frac{1-T}{T}, & \text{if } T > \frac{1}{2}. \end{cases}$$

Hence, we see an interesting explicit phase transition occurring at $T = T_c = 1/2$. In terms of the probability $\mathbf{P}$ that a randomly chosen individual escapes infection we find that

$$\mathbf{P} = \begin{cases} 1, & \text{if } T \leq \frac{1}{2}, \\ \left(\frac{1-T}{T}\right)^3, & \text{if } T > \frac{1}{2}. \end{cases}$$

Notice that this is compatible with Eq. (11). Indeed, expanding $\mathbf{P}$ near $T_c = 1/2$ we find that $\mathbf{P} = 1 - 12(T - T_c) + \cdots$.

When $n \geq 1$ we can equally prove existence of a critical point in $(0,1)^n$ if certain $n$ quantities are greater than one (in the $n = 1$ case there is a single quantity which can be readily identified with $R_0 > 1$). However, in this more general case the proof is slightly less elementary as this is a codimension $n > 1$ problem for which the intermediate value theorem no longer applies.

On possible approach would be to denote by $F : [0,1]^n \to [0,1]^n$ the function defined by the right hand side of (8) and find hypothesis so that there is $\delta > 0$ such that $F([0, 1-\delta]^n) \subset [0, 1-\delta]^n$. Then, the Brower fixed point theorem would guarantee the existence of such a fixed point in $(0,1)^n$. However, we shall instead proceed in a slightly different manner.

**Proposition 1** *Suppose that for all* $i = 1, \ldots, n$, *the quantities* $R_0(i)$ *defined in* (7) *are all greater than* 1, *i.e.*

$$R_0(i) = \sum_{j=1}^{n} T_j \frac{\partial \tilde{G}_i}{\partial q_j}\Big|_{q_1 = \cdots = q_n = 1} > 1.$$

*Then, there is a solution* $(q_1, \ldots, q_n) \in (0,1)^n$ *of Eq.* (8) *and so* $\mathbf{P} \in (0,1)$.

***Proof*** We shall look for solutions of (8) in $(0,1)^n$ and it will prove convenient to write these as $q_i = 1 - T_i\varepsilon_i$. Then, the fixed point equation (8) turns into the equations

$$\varepsilon_i = 1 - \tilde{G}_i(1 - T_1\varepsilon_1, \ldots, 1 - T_n\varepsilon_n), \quad i = 1, \ldots, n,$$

for $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$. Then, we look for fixed points of the function $F$ given by the right hand side of equation above with $\varepsilon \neq 0$. By Brower's fixed point theorem, for such a fixed point to exist, it is enough if $F : [0, T_1^{-1}] \times \cdots [0, T_n^{-1}] \to [0,1]^n$ maps $[\delta, 1]^n$ to itself, for some positive $\delta \ll 1$. First, notice that each entry

$$F_i(\varepsilon_1, \ldots, \varepsilon_n) = 1 - \tilde{G}_i(1 - T_1\varepsilon_1, \ldots, 1 - T_n\varepsilon_n),$$

is nondecreasing in each coordinate and its image lies in $[0,1]^n$. Furthermore, by Taylor's formula

$$F_i(\varepsilon_1, \ldots, \varepsilon_j) = \sum_{j=1}^{n} T_j \frac{\partial \tilde{G}_i}{\partial x_j}\Big|_{x_1=\cdots=x_n=1} \varepsilon_j + o(|\varepsilon|),$$

for $|\varepsilon| \ll 1$. This shows that for sufficiently small $\delta > 0$

$$F_i(\delta, \ldots, \delta) = \left( \sum_{j=1}^{n} T_j \frac{\partial \tilde{G}_i}{\partial x_j}\Big|_{x_1=\cdots=x_n=1} \right) \delta + o(\delta) > \delta,$$

if

$$\sum_{j=1}^{n} T_j \frac{\partial \tilde{G}_i}{\partial x_j}\Big|_{x_1=\cdots=x_n=1} > 1.$$

The quantities in the left hand side can be readily identified with the $R_0(i)$, from which we conclude that under these hypothesis $F([\delta, 1]^n) \subset [\delta, 1]^n$ and a fixed point exists. $\square$

Inspired by this proof and the computation in example 1, also for $n \geq 1$ we shall search for a phase transition (or bifurcation) from $q_c = (1, \ldots, 1)$ due to a variation in the parameters $T = (T_1, \ldots, T_n)$. In order to set up the nomenclature, we shall consider a 1-parameter family $t \in I \subset \mathbb{R} \mapsto T(t)$ of transmissibilities. For all $t \in I$ we have that $q_c = (1, \ldots, 1)$ is a solution to (8). We shall say that a bifurcation from $q_c$ occurs at $T(0)$ if any neighborhood of $(q_c, T(0))$ in $\mathbb{R}^n \times I$ contains solutions of (8) not equal to $q_c$. This will be called a phase transition if such solutions lie in a continuous curve parameterized by $t$. The following result gives a necessary condition for the existence of a phase transition.

**Proposition 2** *Consider a 1-parameter family of parameters $T(t) = (T_1(t), \ldots, T_n(t))$ and suppose that there is a continuous $t \mapsto q(t) = (q_1(t), \ldots, q_n(t))$ solution to (8), for $t \in (-\delta, \delta)$ with $q(t) = (1, \ldots, 1)$ for $t \leq 0$ and $q(t) \in (0, 1)^n$ for $t > 0$. Then, the $\mathcal{R}^{\tilde{G}}$-matrix at $T(0)$ must have 1 as one of its eigenvalues.*

***Proof*** Bifurcations of $q$ from $q_c$ at $T(0)$ are in one to one correspondence with bifurcations of $\varepsilon$ from $\varepsilon_c = 0$ at $T(0)$. Defining the function $\mathcal{G} = (\mathcal{G}_1, \ldots, \mathcal{G}_n)$ whose entries are

$$\mathcal{G}_i(\varepsilon_1, \ldots, \varepsilon_n) = 1 - \tilde{G}_i(1 - T_1\varepsilon_1, \ldots, 1 - T_n\varepsilon_n) - \varepsilon_i,$$

for $i = 1, \ldots, n$. By the implicit function theorem, if for a given $T$ we had $d\mathcal{G}_0 : \mathbb{R}^n \to \mathbb{R}^n$ being an isomorphism, then no bifurcation could occur. This already gives us a necessary condition for a bifurcation to occur.

The $(i, j)$ entry of $d\mathcal{G}_0$ is given by

$$\frac{\partial \mathcal{G}_i}{\partial \varepsilon_j}\Big|_{\varepsilon=0} = T_j \frac{\partial \tilde{G}_i}{\partial x_j}\Big|_{x=1} - \delta_{ij} = \mathcal{R}^{\tilde{G}}_{ij} - \delta_{ij}, \tag{12}$$

and so $d\mathcal{G}_0$ is non-invertible if and only if the $\mathcal{R}^{\tilde{G}}$-matrix has 1 as an eigenvalue. We have thus concluded that for a bifurcation to occur at a given $T$, the $\mathcal{R}^{\tilde{G}}$-matrix must have a unit eigenvalue. $\square$

***Remark 10*** We also mention in passing that in biology, such phase transitions and bifurcation phenomena are sometimes referred to as branching processes. We direct the reader to the Refs.[22,23] for more on such branching processes in biology.

Suppose now that the $\mathcal{R}^{\tilde{G}}$-matrix associated with $T(t)$, which we shall denote by $\mathcal{R}^{\tilde{G}}(t)$ has exactly one eigenvalue $\lambda(t)$ such that $\lambda(0) = 1$ (this means that the algebraic multiplicity of $\lambda(t)$ is one). If we further assume that $\dot{\lambda}(0) \neq 0$, i.e $\lambda(t)$ crosses 1 transversely at $t = 0$. Then, the computation (12) shows that at $t = 0$ the map $d\mathcal{G}_0$ has a 1-dimensional kernel and cokernel and that $(\partial_t|_{t=0} d\mathcal{G}_0)(\ker(d\mathcal{G}_0)) \notin \mathrm{im}(d\mathcal{G}_0)$. Then, the Crandall–Rabinowitz theorem (Theorem 1.7. in[24]) shows that a phase transition must occur at $t = 0$. We shall state this separately as follows.

**Proposition 3** *Consider a 1-parameter family of parameters $T(t) = (T_1(t), \ldots, T_n(t))$ whose associated $\mathcal{R}^{\tilde{G}}$-matrix has a unique eigenvalue $\lambda(t)$ satisfying $\lambda(0) = 1$. If $\dot{\lambda}(0) \neq 0$, then a phase transition must occur at $t = 0$.*

***Example 4*** $(n = 2)$ We have $\tilde{G}_i(x_1, x_2) = \frac{G_i'(x_i)}{G_i'(1)} G_j(x_j)$ for $i = 1, 2$ and $j \neq i$. In particular,

$$\frac{\partial \tilde{G}_1}{\partial x_1} = \frac{G_1''(x_1)}{G_1'(1)} G_2(x_2), \qquad \frac{\partial \tilde{G}_1}{\partial x_2} = \frac{G_1'(x_1) G_2'(x_2)}{G_1'(1)},$$

and similarly for $\tilde{G}_2$. Hence, the $\mathcal{R}^1$-matrix can be written as

$$\mathcal{R}^{\tilde{G}} = \begin{pmatrix} T_1 \frac{G_1''(1)}{G_1'(1)} & T_1 G_1'(1) \\ T_2 G_2'(1) & T_2 \frac{G_2''(1)}{G_2'(1)} \end{pmatrix} = \begin{pmatrix} T_1 \frac{\mathbf{E}(k_1^2) - \mathbf{E}(k_1)}{\mathbf{E}(k_1)} & T_1 \mathbf{E}(k_1) \\ T_2 \mathbf{E}(k_2) & T_2 \frac{\mathbf{E}(k_2^2) - \mathbf{E}(k_2)}{\mathbf{E}(k_2)} \end{pmatrix}.$$

***Example 5*** ($n = 2$ *with 2 neighbors each*) Consider $T_1 < T_2$ and $G_1(x) = x^2$, $G_2(x) = x^2$. Then, we have $\tilde{G}_i(x_1, x_2) = x_i x_j^2$ for $i = 1, 2$ and $j \neq i$. Then, the $\mathcal{R}^{\tilde{G}}$-matrix is

$$\mathcal{R}^{\tilde{G}} = \begin{pmatrix} T_1 & 2T_1 \\ 2T_2 & T_2 \end{pmatrix},$$

whose eigenvalues are

$$\lambda_{\pm} = \frac{T_1 + T_2}{2} + \frac{1}{2}\sqrt{T_1^2 + T_2^2 + 14T_1T_2}.$$

Clearly, $\lambda_- < 0$ and so a phase transition must occurs when $\lambda_+$ crosses 1, i.e. when

$$T_1 + T_2 + \sqrt{T_1^2 + T_2^2 + 14T_1T_2} = 2.$$

Furthermore, in this case the equations for $(q_1, q_2)$ read

$$q_1 = 1 - T_1 + T_1 q_1 q_2^2$$
$$q_2 = 1 - T_2 + T_2 q_1^2 q_2.$$

In particular, using the first equation to write $q_1$ in terms of $q_2$ and inserting in the second we find that solutions are given by $q_2 = 1$ and solutions of the quartic equation

$$T_1^2 q_2^4 + T_1^2 T_2 q_2^3 + (T_2T_1 - 2)T_1 q_2^2 + T_1T_2(T_1 - 2)q_2 - T_2 + 1 = 0.$$

***Example 6*** ($n = 2$ *with 2 exponentially distributed graphs*) Again, we consider $T_1 < T_2$ and $G_1(x) = x^{-N_1(1-x)}$, $G_2(x) = x^{-N_2(1-x)}$ for $N_1 > N_2$. In this situation we have

$$\tilde{G}_1(x_1, x_2) = e^{-N_1(1-x_1) - N_2(1-x_2)} = \tilde{G}_2(x_1, x_2).$$

Then, the $\mathcal{R}^1$-matrix is

$$\mathcal{R}^{\tilde{G}} = \begin{pmatrix} T_1N_1 & T_1N_1 \\ T_2N_2 & T_2N_2 \end{pmatrix},$$

whose eigenvalues are 0 and

$$\lambda = T_1N_1 + T_2N_2.$$

which in this case coincides with $R_0$ and we therefore find that a phase transition occurs when $R_0$ crosses 1. In fact, it is tempting to regard $T_1N_1$ and $T_2N_2$ as the respective contributions to $R_0$ by the networks $\mathcal{G}_1$ and $\mathcal{G}_2$. Indeed, $R_0(1) = 2T_1N_1$ and $R_0(2) = 2T_2N_2$ so that

$$R_0 = \frac{1}{2}(R_0(1) + R_0(2)),$$

as shown in Remark 8.

These interpretations of $R_0(1)$ and $R_0(2)$ may, however, not be appropriate to interpret some non-intuitive phenomena. For example, one may be lead to think that both $q_1$ and $q_2$ are non-increasing with respect to $R_0(1)$ and $R_0(2)$. However, this need not be true as we shall illustrate in an example. In this case the equations for $(q_1, q_2)$ read

$$q_1 = 1 - T_1 + T_1 e^{-N_1(1-q_1) - N_2(1-q_2)}$$
$$q_2 = 1 - T_2 + T_2 e^{-N_1(1-q_1) - N_2(1-q_2)},$$

and we shall now iterate the right hand side to approximate two solutions. Say, in the case when $N_1 = 50$, $T_1 = 0.1$ and $N_2 = 2$, $T_2 = 0.6$, which corresponds to $R_0(1) = 5$ and $R_0(2) = 1.2$ we have $(q_1, q_2) \sim (0.9, 0.4)$. On the other hand, if instead $N_1 = 5$, $T_1 = 0.2$, which corresponds to $R_0(1) = 1$, while the rest remains as before, we have $(q_1, q_2) \sim (0.83, 0.49)$ and so $q_1$ did decrease but $q_2$ increased.

However, this situation is not totally counter-intuitive as overall, the value of $\mathbf{P}$, the probability that a random vertex escapes infection, does increase in the second example where $\mathbf{P} \sim 0.17$ in comparison with a $\mathbf{P} \sim 0.13$ in the first example.

## Dynamic modeling in a local mean field approximation

In section "Prevalence of infection and percolation (the late stages of an outbreak)" we studied a mathematical framework, related to percolation models, and used the properties of the network in order to compute the probability that a given node will eventually be infected. However, this framework does not look at the specific way the infection propagates in the network through time. That will be the topic of the current section. Here, we investigate an extended SIR type system modeling the spread of an epidemic on a network with different types of contacts, meaning that the transmissibilities are not all the same and can be gauged to approximate different kinds of contacts. As before, we consider a set of $n$ graphs $\mathcal{G}_1, \ldots, \mathcal{G}_n$ with the same vertices but different edges. These encode the interactions and each graph $\mathcal{G}_i$ is weighted by a transmissibility per unit time $\beta_i$ encoding the probability of transmitting the disease through that interaction (per time unit). Our model is a simple alternative to[25] which more directly deals with contact duration.

### The model.

For each vertex $v$ of $\mathcal{G}$ we shall denote by $\mathcal{G}_i(v)$ all its neighbors through the graph $\mathcal{G}_i$, in other words $\mathcal{G}_i(v)$ is the set of all vertices which are connected to $v$ through a an edge of $\mathcal{G}_i$. Then, we respectively denote by $s_v$, $x_v$ and $r_v$ the probabilities that $v$ is either susceptible to the disease, infected, or removed. The dynamics of this network SIR model is then approximately governed by

$$\begin{aligned} \dot{s_v} &= -\sum_{i=1}^{n} \beta_i \sum_{w \in \mathcal{G}_i(v)} s_v x_w \\ \dot{x_v} &= -\gamma x_v + \sum_{i=1}^{n} \beta_i \sum_{w \in \mathcal{G}_i(v)} s_v x_w \\ \dot{r_v} &= \gamma x_v, \end{aligned} \tag{13}$$

where $\gamma > 0$ is the rate of recovery.

***Remark 11*** Alternatively, we can let $A^i_{vw}$ be the entries of the adjunction matrix of the graph $\mathcal{G}_i$ and write the sum $\sum_{w \in \mathcal{G}(v_i)} s_v x_w$ as $\sum_w A^i_{vw} s_v x_w$.

***Remark 12*** The system (13) is an approximation because the average probability that $v$ is susceptible and $w$ infected is only approximately given by $s_v x_w$. In order to work with a non-approximate model one would have to write an infinite array of equations modeling the dynamics of the average probabilities of all such nonlinear quantities. See also[3] for such an analysis carried out in the situation where there is only one type of interaction.

### Classifying the disease free equilibrium.

It is clear from the equation $\dot{r_v} = \gamma x_v$ that any equilibrium solution of the system (13) we must have $x_v = 0$ for all $v$. Hence, any equilibrium solution is disease free. In fact, setting $x_v = 0$ and $s_v + r_v = 1$ gives a 1-parameter family of equilibrium solutions of the system and any equilibrium solution must be one of these.

An important question is then to understand if a non-constant solution converges to one of these equilibria and to which? The fact that any solution $\{(s_v(t), x_v(t), r_v(t)) | \; t \in [0, +\infty)\}_v$ converges to an equilibrium is immediate from the fact that $r_v(t)$ is nondecreasing and bounded, hence the limit

$$r_v(\infty) := \lim_{t \to +\infty} r_v(t) \in [0, 1]$$

exists. Furthermore $\lim_{t \to +\infty} \dot{r_v} = 0$ which implies $x_v(\infty) = 0$ as we wanted to show. The main question then becomes:

**Question 1** What is the disease free equilibrium $(s_v(\infty), 0, r_v(\infty))$ to which a solution starting at $(s_v(0), x_v(0), r_v(0))$ converges? Can we understand how this depends on the properties of $\mathcal{G}_1, \ldots, \mathcal{G}_n$ and $\beta_1, \ldots, \beta_n$?

From the system of equations (13) we find that for each vertex $v$ of $\mathcal{G}$, there are two conserved quantities

$$1 = s_v + x_v + r_v,$$

and

$$H(v) := s_v \exp\left( \frac{1}{\gamma} \sum_{i=1}^{n} \beta_i \sum_{w \in \mathcal{G}_i(v)} r_w \right).$$

Let $N$ be the total number of vertices. These conserved quantities reduces the system of $3N$ equations for $3N$ functions to a system involving only $N$ functions. In fact introducing $s_v = 1 - r_v - x_v$ and the last equation of system (13) into the previous conserved quantity we find that

$$H(v) = \left(1 - r_v - \frac{\dot{r_v}}{\gamma}\right) \exp\left( \frac{1}{\gamma} \sum_{i=1}^{n} \beta_i \sum_{w \in \mathcal{G}_i(v)} r_w \right).$$

Suppose $v$ starts susceptible, then $\dot{r_v}(0) = 0 = r_v(0)$ and so $H(v) = 1$, from which we have

$$\left(1 - r_v - \frac{\dot{r}_v}{\gamma}\right) \exp\left(\frac{1}{\gamma}\sum_{i=1}^{n}\beta_i \sum_{w\in\mathcal{G}_i(v)} r_w\right) = 1$$

for all time. Given that the solution converges to an equilibrium, we must have $\lim_{t\to+\infty}\dot{r}_v = 0$ and so

$$1 = (1 - r_v)\exp\left(\frac{1}{\gamma}\sum_{i=1}^{n}\beta_i \sum_{w\in\mathcal{G}_i(v)} r_w\right), \tag{14}$$

where in this equation we have written $r_v(\infty)$ as $r_v$ to simplify notation. It will prove convenient to have the following notion at hand.

**Definition 3** Let $v$ be a vertex of $\mathcal{G}$ and $k_i(v)$ its degree as a vertex of $\mathcal{G}_i$ we shall denote by

$$R(v) := \sum_{i=1}^{n}\frac{\beta_i}{\gamma}k_i(v)$$

the expected number of infections $v$ will cause if infected.

***Remark 13*** Notice that the quantity $\beta_i/\gamma$ represents the probability that an interaction of $\mathcal{G}_i$, between an infected and a susceptible individual, results in an infection. Hence, the quantity $\sum_{i=1}^{n}\frac{\beta_i k_i(v)}{\gamma}$ can be regarded as the average number of infections the individual represented by $v$ is expected to cause if it is infected.

***Remark 14*** For example, let us assume we have a sufficiently simple situation so that $r_w \approx r_v$ for all $w \in \mathcal{G}_i(v)$. Then, inserting this into (14) we find that

$$1 = (1 - r_v)\exp\left(r_v\sum_{i=1}^{n}\frac{\beta_i}{\gamma}k_i(v)\right) = (1 - r_v)\exp\left(r_v R(v)\right).$$

In this situation, the right hand side equals 1 when $r_v = 0$ and vanishes when $r_v = 1$. Hence, by the mean value theorem, a solution with $r_v \in (0, 1)$ exists if and only if the derivative of the right hand at $r_v = 0$, is positive. Such a derivative can be computed to be $R(v) - 1$ which is positive if and only if $R(v) > 1$.

In order to investigate the existence of solutions to Eq. (14) with $r_v \neq 0$ it is convenient to rewrite this equation as

$$r_v = 1 - \exp\left(-\sum_{i=1}^{n}\frac{\beta_i}{\gamma}\sum_{w\in\mathcal{G}_i(v)} r_w\right) = 1 - \exp\left(-\sum_{i=1}^{n}\frac{\beta_i}{\gamma}\sum_{w} A^i_{vw} r_w\right). \tag{15}$$

Hence, our problem is now recast as the problem of looking for fixed points of the function $F : [0, 1]^N \to [0, 1]^N$ given by $F(r_1, \ldots, r_N) = (F_1(r_1, \ldots, r_N), \ldots, F_N(r_1, \ldots, r_N))$ where

$$F_v(r_1, \ldots, r_N) = 1 - \exp\left(-\sum_{i=1}^{n}\frac{\beta_i}{\gamma}\sum_{w} A^i_{vw} r_w\right).$$

The first obvious fixed point is that occurring at the origin which corresponds to all individuals still being susceptible, i.e. the disease never spread. In the next result we give a criteria for the existence of another fixed point.

**Theorem 1** *Suppose that for all vertices $v$ of $\mathcal{G}$ we have $R(v) > 1$. Then, the solution to the system* (13) *starting with $r_v(0) = 0$ converges to a disease free equilibrium with $r_v \neq 0$ for all $v$. This satisfies Eq.* (15) *and the bound*

$$r_v \leq 1 - \exp\left(-R(v)\right).$$

*In particular, we find that the upper bound is increasing with $R(v)$* (*in agreement with basic intuition*).

***Proof*** The proof that the solution converges to a disease free equilibrium is given in the beginning of this subsection. This must satisfy (15) and we shall now prove that, under the conditions stated, $r_v \in (0, 1)$, i.e. $r_v \neq 0$. We proceed as in the proof of Proposition 1 by applying Bower's fixed point theorem. Let $\varepsilon > 0$ to be fixed later, then the Taylor expansion of $F_v$ around the origin reads

$$F_v(\varepsilon, \ldots, \varepsilon) = \sum_p \frac{\partial F_v}{\partial r_p}\varepsilon + O(\varepsilon^2) = \left(\sum_{i=1}^{n}\frac{\beta_i}{\gamma}\sum_p A^i_{vp}\right)\varepsilon + O(\varepsilon^2) = \sum_{i=1}^{n}\frac{\beta_i k_i(v)}{\gamma}\varepsilon + O(\varepsilon^2) = R(v)\varepsilon + O(\varepsilon^2).$$

Furthermore, $F_v$ is non-decreasing with respect to each entry. Hence, if $R(v) > 1$ for all $v$ we find that for sufficiently small $\varepsilon > 0$ the function $F|_{[\varepsilon,1]^N}$ maps $[\varepsilon, 1]^N$ to itself and by the Brower fixed point theorem must have a fixed point in $[\varepsilon, 1]^N$.

We turn now to the proof of the upper bound for $r_v$ given in the statement. Recall that $k_i(v)$ denotes the degree of $v$ in $\mathcal{G}_i$. From Eq. (14) we immediately find that the argument of the exponential in the right hand side satisfies

$$\frac{1}{\gamma}\sum_{i=1}^{n}\beta_i \sum_{w\in\mathcal{G}_i(v)} r_w \leq \sum_{i=1}^{n}\frac{\beta_i}{\gamma}\sum_{w\in\mathcal{G}(v_i)} 1 = \sum_{i=1}^{n}\frac{\beta_i}{\gamma}k_i(v).$$

Hence, we discover that $r_v = r_v(\infty)$ satisfies

$$r_v \leq 1 - \exp\left(-\sum_{i=1}^{n}\frac{\beta_i}{\gamma}k_i(v)\right) \tag{16}$$

Inserting the Definition 3, of $R(v)$, in this bound gives $r_v \leq 1 - \exp(-R(v))$, as claimed in the statement. □

***Remark 15*** Inserting the equation $\dot{r_w} = \gamma x_w$ for $w \in \mathcal{G}_i(v)$, in the first equation of (13) we find that

$$\dot{s_v} = -\left(\sum_{i=1}^{n}\frac{\beta_i}{\gamma}\sum_{w\in\mathcal{G}_i(v)}\dot{r_w}\right)s_v.$$

this can be integrated using the initial condition $r_w(0) = 0$ to obtain

$$s_v(t) = s_v(0)\exp\left(-\sum_{i=1}^{n}\frac{\beta_i}{\gamma}\sum_{w\in\mathcal{G}_i(v)} r_w(t)\right).$$

This is a very nice and beautiful formula, but it is not of much use if we know nothing about $r_w(t)$. However, we do can take the limit as $t \to +\infty$ and assume we are converging to a disease free equilibrium $(s_v, x_v, r_v) = (s_v, 0, r_v)$ whose existence is assured, for example, under the assumptions of Proposition 1. In such a situation we have $\lim_{t\to+\infty} s_v(t) = 1 - \lim_{t\to+\infty} r_v(t)$. Hence, in this limit the previous equation turns into

$$1 - r_v(\infty) = s_v(0)\exp\left(-\sum_{i=1}^{n}\frac{\beta_i}{\gamma}\sum_{w\in\mathcal{G}_i(v)} r_w\right).$$

By setting $s_v(0) = 1$ this is the same equation which we have previously derived. This situation is slightly more general than the one we have analyzed in Theorem 1 and a similar results hold yielding the bound

$$r_v(\infty) \leq 1 - s_v(0)\exp(-R(v)).$$

## Dynamic modeling in a mean field approximation by degree similarity

The system of the previous section, though very general is extremely large and difficult to investigate. For this reason, it is convenient to find simpler systems which we can more easily analyze. This is the content of this section where we will consider a system for the average probability that vertices of a given degree are in specific states. This is an oversimplified assumption which nevertheless allows one to gain a lot of insight on the dynamics of an epidemic in a network. The approach we take here is mostly inspired from that of[2], where the authors first learned it for the case $n = 1$. There are also interesting individual based approaches which however require knowing the whole network structure and they also only have $n = 1$, see for example[26]. See also[27] for some computational results using a model on a weighted network.

**The model.** For each $k = (k_1, \ldots, k_n)$ we shall denote by $s_k$, $x_k$ and $r_k$ the average probabilities that a vertex with degree $k$ is susceptible, infected and removed respectively. Such a network SIR model is governed by the following system

$$\begin{aligned}
\dot{s_k} &= -\sum_{i=1}^{n}\beta_i k_i v_i s_k \\
\dot{x_k} &= -\gamma x_k + \sum_{i=1}^{n}\beta_i k_i v_i s_k \\
\dot{r_k} &= \gamma x_k,
\end{aligned} \tag{17}$$

where for $i = 1, \ldots, n$

$$v_i := \sum_{k_1,\ldots,k_n}\tilde{P}_i(k_1,\ldots,k_n)x_{(k_1,\ldots,k_i+1,\ldots,k_n)},$$

is the averaged probability of an individual being infected after following a random edge of the graph $\mathcal{G}_i$.

***Remark 16*** Again, in writing the system (17) we have used one large simplification. Namely, we have approximated the average value of $x_{k'}s_k$ by the product of the average values of $x_{k'}$ and $s_k$.

**An equivalent (much reduced) system of equations.** Notice that (17) can potentially be an extremely large system, as there are as many groups of 3-equations as degree combinations. It is remarkable that this system can be extremely reduced to one that only involves $n$ equations for $n$ functions. For this, it is convenient to introduce quantities $w_i$ measuring the average number of removed individuals following an edge of $\mathcal{G}_i$. These are given by

$$w_i := \sum_{k_1,\dots,k_n} \tilde{P}_i(k_1,\dots,k_n) r_{(k_1,\dots,k_i+1,\dots,k_n)},$$

for $i = 1,\dots,n$. Then, we have $\dot{w}_i = \gamma v_i$ which inserting into the equation for $\dot{s_k}$ yields $\dot{s_k} = -\frac{1}{\gamma}\sum_{i=1}^n \beta_i k_i \dot{w}_i s_k$. Assuming that the epidemic outbreak starts with no-one removed from previous infections, we have $w_i(0) = 0$ (which follows from $r_k(0) = 0$ for all $k \in \mathbb{N}^n$) and so

$$s_k(t) = s_k(0) \exp\left(-\sum_{i=1}^n \frac{\beta_i}{\gamma} k_i w_i\right),$$

which we can also rewrite as

$$s_k(t) = s_k(0) u_1^{k_1} \dots u_n^{k_n},$$

for $u_i(t) = \exp\left(-\frac{\beta_i}{\gamma} w_i\right)$. Now, recall that $s_k + x_k + r_k = 1$ and so

$$\begin{aligned} v_i &= \sum_{k_1,\dots,k_n} \tilde{P}_i(k_1,\dots,k_n)(1 - r_{(k_1,\dots,k_i+1,\dots,k_n)} - s_{(k_1,\dots,k_i+1,\dots,k_n)}) \\ &= 1 - \sum_{k_1,\dots,k_n} \tilde{P}_i(k_1,\dots,k_n) r_{(k_1,\dots,k_i+1,\dots,k_n)} \\ &\quad - \sum_{k_1,\dots,k_n} \tilde{P}_i(k_1,\dots,k_n) s_{(k_1,\dots,k_i+1,\dots,k_n)}(0) u_1^{k_1} \dots u_i^{k_i+1} \dots u_n^{k_n} \\ &= 1 - w_i - \tilde{H}_i(u_1,\dots,u_n), \end{aligned} \tag{18}$$

where in the last equality we have used the generating function (6) whose definition we recall to be

$$\tilde{H}_i(x_1,\dots,x_n) = \sum_{k_1,\dots,k_n} \tilde{P}_i(k_1,\dots,k_n) s_{(k_1,\dots,k_i+1,\dots,k_n)}(0) x_1^{k_1} \dots x_i^{k_i+1} \dots x_n^{k_n},$$

which depends on the initial conditions $s_{(k_1,\dots,k_n)}(0)$.

***Remark 17*** Consider the case where one is willing to make the simplifying assumption that $s_k(0) = s(0)$ for all $k \in \mathbb{N}^n$, i.e. the initial proportion of susceptible individuals is independent of their degree distribution (this may be a reasonable assumption for a new disease which begins to spread in an unknown part of the population). Then, $\tilde{H}_i(x_1,\dots,x_n) = s(0)\tilde{G}_i(x_1,\dots,x_n)$ for all $i = 1,\dots,n$. Notice a few properties of the function $\tilde{H}$, namely it is nondecreasing with respect to each coordinate, and

$$\tilde{H}_i(1) = \sum_{k_1,\dots,k_n} \tilde{P}_i(k_1,\dots,k_n) s_{(k_1,\dots,k_i+1,\dots,k_n)}(0) = \mathbf{E}[s(0)|\mathcal{G}_i],$$

i.e. the average number of initially susceptible individuals in $\mathcal{G}_i$, meaning those which have edges in $\mathcal{G}_i$.

Using $v_i = \frac{1}{\gamma}\dot{w}_i$ and $w_i = -\frac{\gamma}{\beta_i}\log u_i$ and rearranging, we find that (17) can be written as a system for $w = (w_1,\dots,w_n)$ given by

$$\gamma^{-1}\dot{w}_i = 1 - w_i - \tilde{H}_i\left(e^{-\frac{\beta_1}{\gamma}w_1},\dots,e^{-\frac{\beta_n}{\gamma}w_n}\right). \tag{19}$$

and further using and $w_i = -\frac{\gamma}{\beta_i}\log u_i$ this can be rewritten as a system for $u = (u_1,\dots,u_n)$

$$\dot{u}_i = -\beta_i u_i \left(1 + \frac{\gamma}{\beta_i}\log u_i - \tilde{H}_i(u_1,\dots,u_n)\right), \tag{20}$$

which is a substantially smaller than the initial one in (17).

**Existence of an equilibrium.** At an equilibrium point we either have the rights hand side of (19) vanishing, i.e.

$$1 - w_i - \tilde{H}_i\left(e^{-\frac{\beta_1}{\gamma}w_1}, \ldots, e^{-\frac{\beta_n}{\gamma}w_n}\right) = 0,$$

or, written as a fixed point equation, as

$$w_i = 1 - \tilde{H}_i\left(e^{-\frac{\beta_1 w_1}{\gamma}}, \ldots, e^{-\frac{\beta_n w_n}{\gamma}}\right), \tag{21}$$

for $i = 1, \ldots, n$. In particular, notice from Eq. (18) that this implies that $v_i = 0$ for all $i = 1, \ldots, n$ and so this equation encodes a disease free equilibrium to which the solution of the system is expected to converge.

As in section "Prevalence of infection and percolation (the late stages of an outbreak)", we shall start by analyzing the case when $n = 1$ in the following example. It will serve as a good exercise for the $n > 1$ case. See also[2] for this analysis in the case $n = 1$.

***Example 7*** ($n = 1$) In this situation, there is only one $\tilde{H}_i(x)$ which is given by

$$\tilde{H}(x) = \sum_{k=0}^{+\infty} \tilde{P}(k) s_{k+1}(0) x^{k+1} = \frac{1}{\mathbf{E}(k)} \sum_{k=1}^{+\infty} k P(k) s_k(0) x^k.$$

and the Eq. (21) for an equilibrium is

$$w = 1 - \tilde{H}(e^{-\frac{\beta w}{\gamma}}). \tag{22}$$

Hence, an equilibrium is determined by fixed points of the function $g(w)$ given by the right hand side of (22). Notice that $g(0) = 1 - \mathbf{E}[s(0)] > 0$ (even though it may be very small) while $g(1) = 1 - \tilde{H}(e^{-\frac{\beta}{\gamma}}) < 1$, and so a fixed point $w^{eq}$ always exist by the intermediate value theorem. Moreover, we find from

$$g'(w) = \frac{\beta}{\gamma}\tilde{H}'(e^{-\frac{\beta w}{\gamma}}) > 0,$$

that $g(w)$ is increasing while from

$$g''(w) = -\left(\frac{\beta}{\gamma}\right)^2 \tilde{H}''(e^{-\frac{\beta w}{\gamma}}) < 0,$$

we find that its concavity always faces down. Hence, the equilibrium point $w^{eq}$ must be unique. We further deduce the equilibrium value of $s_k$ to be

$$s_k^{eq} = s(0) e^{-\frac{k\beta}{\gamma} w^{eq}}.$$

However, when $\mathbf{E}[s(0)] = 1$, then there is fixed point of $g$ at $w = 0$ which corresponds to never having a disease spreading. For another fixed point $w^{eq} \in (0, 1)$ to exist we must have $g'(0) > 1$ which can be written as

$$\frac{\beta}{\gamma}\tilde{H}'(1) > 1,$$

and can be identified with the initial basic reproduction number. This situation with $\mathbf{E}[s(0)] = 1$ is relevant, for example, when considering an approximately infinite number of individuals with only a finite number of infected individuals.

***Example 8*** (*$n = 1$ and every vertex with k neighbors*) In this case $\tilde{H}(x) = x^k$ and so the equilibrium is attained at a $w^{eq} = w$ such that $w = 1 - s(0)e^{-\frac{\beta}{\gamma}kw}$.

Also in this case, the insights given by the previous example can be extended to higher dimensions to prove the existence of an equilibrium point of the system starting at a given configuration of initially susceptible individuals.

**Proposition 4** *Suppose that the average number of initially susceptible individuals $\mathbb{E}[s(0)] < 1$. Then, there is an equilibrium point $w^{eq} = (w_1^{eq}, \ldots, w_n^{eq}) \in (0, 1)^n$ such that*

$$s_k^{eq} = s_k(0) \exp\left(-\sum_{i=1}^{n} \frac{\beta_i}{\gamma} k_i w_i^{eq}\right).$$

*In particular,*

$$r_k^{eq} = 1 - s_k(0) \exp\left(-\sum_{i=1}^{n} \frac{\beta_i}{\gamma} k_i w_i^{eq}\right) \leq 1 - s_k(0) \exp\left(-\sum_{i=1}^{n} \frac{\beta_i}{\gamma} k_i\right)$$

***Proof*** We are looking for fixed points of the (continuous) map $g : [0, 1]^n \to [0, 1]^n$ is given by

$$g(w_1,\ldots,w_n)=\left(1-\tilde{H}_1\left(e^{-\frac{\beta_1 w_1}{\gamma}},\ldots,e^{-\frac{\beta_n w_n}{\gamma}}\right),\ldots,1-\tilde{H}_n\left(e^{-\frac{\beta_1 w_1}{\gamma}},\ldots,e^{-\frac{\beta_n w_n}{\gamma}}\right)\right),$$

and again at least one such equilibrium $w^{eq}=(w_1^{eq},\ldots,w_n^{eq})$ exists from Brower's fixed point theorem. Using this we find

$$s_k^{eq}=s_k(0)\exp\left(-\sum_{i=1}^{n}\frac{\beta_i}{\gamma}k_i w_i^{eq}\right),$$

which again we can see to be exponentially decreasing with $k_i$ and $T_i$ (notice however that the $w_i^{eq}$ themselves are also functions of the $\beta_i/\gamma$ so this statement is somewhat imprecise).

Given that at the equilibrium point all $x_k=0$ (recall that $\gamma v_i=\dot{w}_i$) we have $s_k^{eq}+r_k^{eq}=1$ from which the last equality and inequality in the statement follow. ☐

***Remark 18*** Of course, when $\mathbb{E}[s(0)]=1$ there is an equilibrium point with $w^{eq}=0$ and $s_k^{eq}=1$.

**Convergence to the equilibrium.** We consider the cases $n=1$ and $n=2$ for which we have the flow equation (20) which we rewrite as $\dot{u}_i=f_i(u_1,\ldots,u_n)$, for

$$f_i(u_1,\ldots,u_n)=-\beta_i u_i\left(1+\frac{\gamma}{\beta_i}\log u_i-s(0)\tilde{G}_i(u_1,\ldots,u_n)\right). \tag{23}$$

In the case $n=1$ we have

$$\begin{aligned} f'(u)&=-\beta\left(1+\frac{\gamma}{\beta}\log u-s(0)\tilde{G}(u)\right)-\beta u\left(\frac{\gamma}{\beta}\frac{1}{u}-s(0)\tilde{G}'(u)\right)\\ &=-\beta-\gamma\log u+s(0)T\tilde{G}(u)-\gamma+s(0)u\beta\tilde{G}'(u). \end{aligned}$$

In particular, $\lim_{u\to 0^+}f(u)=0$ with $\lim_{u\to 0^+}f'(u)=+\infty$ and so there is $\varepsilon>0$ such that $f(u)>0$ in $(0,\varepsilon)$. On the other hand while $\lim_{u\to 1^-}f(u)=-\beta(1-s(0))<0$. Hence, the solution to (20) with $n=1$ stays bounded inside $(0,1)$ and therefore converges to an equilibrium point $u^{eq}\in(0,1)$ with $f(u^{eq})=0$. Hence,

$$f'(u^{eq})=-\beta u\left(\frac{\gamma}{\beta}\frac{1}{u}-s(0)\tilde{G}'(u)\right)=-\big(\gamma-\beta s(0)\tilde{G}'(u)u\big).$$

We now turn to the case when $n=2$. In this situation we have $\lim_{u_i\to 0^+}f_i(u_1,u_2)=0$ and

$$\begin{aligned} \frac{\partial f_1}{\partial u_1}&=-\beta_1\left(1+\frac{\gamma}{\beta_1}\log u_1-s(0)\tilde{G}_1(u_1,u_2)\right)-\beta_1 u_1\left(\frac{\gamma}{\beta_1}\frac{1}{u_1}-s(0)\frac{\partial\tilde{G}_1}{\partial x_1}(u_1,u_2)\right)\\ &=-\beta_1-\gamma\log u_1-s(0)\beta_1\tilde{G}_1(u_1,u_2)-\gamma-s(0)u_1\beta_1\frac{\partial\tilde{G}_1}{\partial x_1}(u_1,u_2), \end{aligned}$$

and similarly for $\frac{\partial f_2}{\partial u_2}$. In particular, we find that $\lim_{u_i\to 0^+}f_i(u_1,u_2)=+\infty$. Hence, we have that there is $\varepsilon>0$ such that $f_i(u_1,u_2)>0$ for $u_i\in(0,\varepsilon)$. Furthermore,

$$\begin{aligned} \lim_{u_1\to 1^-}f_1(u_1,u_2)&=-\beta_1\big(1-s(0)\tilde{G}_1(1,u_2)\big)\\ &=-\beta_1(1-s(0)G_2(u_2))\\ &\le-\beta_1(1-s(0))<0. \end{aligned}$$

Hence, the solutions stay inside the square $[0,1]^2$ and by the Poincaré–Bendixon theorem, it must therefore converge to an equilibrium point $(u_1^{eq},u_2^{eq})$ with $f_i(u_1^{eq},u_2^{eq})=0$ for $i=1,2$. Notice that here we have implicitly used the fact that there are no non-constant periodic solutions. That can easily be proven by looking at the system (17) from which we find that the $r_k$ must be non-decreasing and can only be constant at a disease free equilibrium.

**Short time behavior.** Recall from Eq. (19) that the system is governed by the set of ordinary differential equations

$$\gamma^{-1}\dot{w}_i=1-w_i-\tilde{H}_i\left(e^{-\frac{\beta_1}{\gamma}w_1},\ldots,e^{-\frac{\beta_n}{\gamma}w_n}\right), \tag{24}$$

for $i=1,\ldots,n$. Given that initially we have $w_i(0)=0$ we may Taylor expand right hand side above using $\tilde{H}_i(1)=\mathbf{E}[s(0)]$ as alluded to in remark 17. This gives

$$\begin{aligned}\gamma^{-1}\dot{w}_i &= 1 - w_i - \left(\tilde{H}_i(1) - \sum_{j=1}^{n} \frac{\partial \tilde{H}_i}{\partial x_j}\Big|_{x=1} \frac{\beta_j}{\gamma} w_j\right) + \cdots \\ &= 1 - \mathbf{E}[s(0)] + \sum_{j=1}^{n} \left(\frac{\beta_j}{\gamma} \frac{\partial \tilde{H}_i}{\partial x_j}\Big|_{x=1} - \delta_{ij}\right) w_j + \cdots \\ &= 1 - \mathbf{E}[s(0)] + \sum_{j=1}^{n} \left(\mathcal{R}^{\tilde{H}}_{ij} - \delta_{ij}\right) w_j + \cdots\end{aligned}$$

where $\mathcal{R}^{\tilde{H}}_{ij}$ is the $(i,j)$ entry of the $\mathcal{R}^{\tilde{H}}$-matrix associated with the transmissibilities $T_k = \beta_k/\gamma$ as in definition 2, and the $\cdots$ denote terms of order $O(|w|^2)$. In this way, the above equation may be written as

$$\gamma^{-1}\dot{w} = 1 - \mathbf{E}[s(0)] + (\mathcal{R}^{\tilde{H}} - \mathbb{I}_{n,n})\, w + \cdots \tag{25}$$

Suppose that $\mathcal{R}^{\tilde{H}}$ can be diagonalised and let with eigenbasis $v_1, \ldots, v_n$ and $\lambda_1 \leq \cdots \leq \lambda_n$ the corresponding eigenvalues. Then, we write $w(t) = \sum_{l=1}^{n} w^l(t) v_l$ we find that each $w^l$ must solve

$$\dot{w^l} = 1 - \mathbf{E}[s(0)] + (\lambda_l - 1) w^l + \cdots,$$

which we can integrate to obtain

$$w(t) = (1 - \mathbf{E}[s(0)]) \sum_{l=1}^{n} \frac{1}{\lambda_l - 1}\left(e^{(\lambda_l - 1)t} - 1\right).$$

We therefore find that the initial exponential growth observed at the beginning of an outbreak is codified in the existence of an eigenvector of $\mathcal{R}^{\tilde{H}}$ greater than 1, as alluded to in the introduction.

## Major limitations

As with any model, those considered in this article have a scope and are therefore heavily limited. Obviously, there are limitations associated with the several assumptions and approximations made, but there are also several others. For instance, the fact that one does not need to consider the full form of the network, which can be interpreted as a strength of the model, may also be a severe limitation. Indeed, there are many different multigraphs $\mathcal{G} = \mathcal{G}_1 \cup \cdots \cup \mathcal{G}_n$ having the same degree distributions and the spread of an epidemic outbreak may have different features in such in-equivalent networks. This is not incorporated by our models which simply use the information on the degree distributions. However, as previously mentioned, this is limitation can also be seen positively. Indeed, while the exact form of the network is practically impossible to obtain in practice, estimating the degree distributions is a feasible endeavor. Nevertheless, one must bear such limitations into consideration anytime these models are used.

## Data availibility

All data generated or analysed during this study are included in this published article.

Received: 30 April 2022; Accepted: 5 September 2022
Published online: 10 January 2023

## References

1. Mollison, D. Spatial contact models for ecological and epidemic spread. *J. R. Stat. Soc. Ser. B (Methodol.)* **39**(3), 283–313 (1977).
2. Newman, M. *Networks: An Introduction* (Oxford University Press, 2018).
3. Kiss, I. Z., Miller, J. C. & Simon, P. L. *Mathematics of Epidemics on Networks* (Springer, 2017).
4. Keeling, M. J. & Ken, T. D. E. Networks and epidemic models. *J. R. Soc. Interface* **2**(4), 295–307 (2005).
5. Brauer, F., Castillo-Chavez, C. & Feng, Z. *Mathematical Models in Epidemiology* Vol. 32 (Springer, 2019).
6. Brauer, F. *An Introduction to Networks in Epidemic Modeling. Mathematical Epidemiology* 133–146 (Springer, 2008).
7. Pellis, L. *et al.* Eight challenges for network epidemic models. *Epidemics* **10**, 58–62 (2015).
8. Capasso, V. *Mathematical Structures of Epidemic Systems* (Springer Science and Business Media, 2008).
9. Azizi, A. *et al.* Epidemics on networks: Reducing disease transmission using health emergency declarations and peer communication. *Infect. Dis. Model.* **5**, 12–22 (2020).
10. Liu, J.-B., Zhao, J. & Cai, Z.-Q. On the generalized adjacency, Laplacian and signless Laplacian spectra of the weighted edge corona networks. *Physica A Stat. Mech. Appl.* **540**, 123073 (2020).
11. Liu, J.-B. *et al.* Network coherence analysis on a family of nested weighted n-polygon networks. *Fractals* **29**(08), 1–15 (2021).
12. Zhang, Y. Q. *et al.* Human interactive patterns in temporal networks. *IEEE Trans. Syst. Man Cybern. Syst.* **45**(2), 214–222 (2015).
13. Zhang, Y. Q. *et al.* Spectral analysis of epidemic thresholds of temporal networks. *IEEE Trans. Cybern.* **50**(5), 1965–1977 (2020).
14. Callaway, D. S. *et al.* Network robustness and fragility: Percolation on random graphs. *Phys. Rev. Lett.* **85**(25), 5468 (2000).
15. Osat, S., Faqeeh, A. & Radicchi, F. Optimal percolation on multiplex networks. *Nat. Commun.* **8**(1), 1–7 (2017).
16. Harris, T. E. *The Theory of Branching Processes* (Springer, 1963).
17. Oliveira, G. Early epidemic spread, percolation and Covid-19. *J. Math. Biol.* **81**(4), 1143–1168 (2020).
18. Barnard, R. C. *et al.* Edge-based compartmental modelling of an SIR epidemic on a dual-layer static-dynamic multiplex network with tunable clustering. *Bull. Math. Biol.* **80**(10), 2698–2733 (2018).
19. Buono, C. *et al.* Epidemics in partially overlapped multiplex networks. *PLoS One* **9**(3), e92200 (2014).
20. Zhang, X. *et al.* Multiplex network reconstruction for the coupled spatial diffusion of infodemic and pandemic of COVID-19. *Int. J. Digit. Earth* **14**(4), 401–423 (2021).

21. Grassberger, P. On the critical behavior of the general epidemic process and dynamical percolation. *Math. Biosci.* **63**(2), 157–172 (1983).
22. Jagers, P. *Branching Processes with Biological Applications* (Wiley, 1975).
23. Kimmel, M. & Axelrod, D. E. *Branching Processes in Biology* (Springer, 2002).
24. Crandall, M. G. & Rabinowitz, P. H. Bifurcation from simple eigenvalues. *J. Funct. Anal.* **8**(2), 321–340 (1971).
25. Miller, J. C., Slim, A. C. & Volz, E. M. Edge-based compartmental modelling for infectious disease. *J. R. Soc. Interface* **9**, 890–906 (2012).
26. Youssef, M. & Scoglio, C. An individual-based approach to SIR epidemics in contact networks. *J. Theor. Biol.* **283**(1), 136–144 (2011).
27. Kamp, C., Moslonka-Lefebvre, M. & Alizon, S. Epidemic spread on weighted networks. *PLoS Comput. Biol.* **9**(12), e1003352 (2013).

## Acknowledgements

Gonçalo Oliveira is supported by the NOMIS Foundation, Fundação Serrapilheira 1812-27395, by CNPq grants 428959/2018-0 and 307475/2018-2, and by FAPERJ through the grant Jovem Cientista do Nosso Estado E-26/202.793/2019.

## Author contributions

Both authors contributed to all aspects of the production of this manuscript. This manuscript consists of material that will be included in A.G. Ph.D. thesis carried out under the supervision of G.O.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to G.O.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.