PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Connectivity, reproduction number, and mobility interact to determine communities' epidemiological superspreader potential in a metapopulation network

Brandon Lieberthal*, Allison M. Gardner

University of Maine, Orono, Maine, United States of America

* brandon.lieberthal@maine.edu

Check for updates

OPEN ACCESS

Citation: Lieberthal B, Gardner AM (2021) Connectivity, reproduction number, and mobility interact to determine communities' epidemiological superspreader potential in a metapopulation network. PLoS Comput Biol 17(3): e1008674. https://doi.org/10.1371/journal.pcbi.1008674

Editor: Benjamin Muir Althouse, Institute for Disease Modeling, UNITED STATES

Received: July 13, 2020

Accepted: January 5, 2021

Published: March 18, 2021

Copyright: © 2021 Lieberthal, Gardner. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and at the following link: http://dataverse.acg.maine.edu/dvn/dv/superspreader.

Funding: BL and AG were funded by National Science Foundation Coupled Natural-Human Systems award #1824961 (https://www.nsf.gov/pubs/2018/nsf18503/nsf18503.htm). AG was funded by USDA National Institute of Food and Agriculture, Hatch Project Number ME021826 through the Maine Agricultural and Forest

Abstract

Disease epidemic outbreaks on human metapopulation networks are often driven by a small number of superspreader nodes, which are primarily responsible for spreading the disease throughout the network. Superspreader nodes typically are characterized either by their locations within the network, by their degree of connectivity and centrality, or by their habitat suitability for the disease, described by their reproduction number (R). Here we introduce a model that considers simultaneously the effects of network properties and R on superspreaders, as opposed to previous research which considered each factor separately. This type of model is applicable to diseases for which habitat suitability varies by climate or land cover, and for direct transmitted diseases for which population density and mitigation practices influences R. We present analytical models that quantify the superspreader capacity of a population node by two measures: probability-dependent superspreader capacity, the expected number of neighboring nodes to which the node in consideration will randomly spread the disease per epidemic generation, and time-dependent superspreader capacity, the rate at which the node spreads the disease to each of its neighbors. We validate our analytical models with a Monte Carlo analysis of repeated stochastic Susceptible-Infected-Recovered (SIR) simulations on randomly generated human population networks, and we use a random forest statistical model to relate superspreader risk to connectivity, R, centrality, clustering, and diffusion. We demonstrate that either degree of connectivity or R above a certain threshold are sufficient conditions for a node to have a moderate superspreader risk factor, but both are necessary for a node to have a high-risk factor. The statistical model presented in this article can be used to predict the location of superspreader events in future epidemics, and to predict the effectiveness of mitigation strategies that seek to reduce the value of R, alter host movements, or both.

Author summary

Infectious disease outbreaks on human mobility networks often are driven by a small number of superspreader individuals or communities, which are primarily responsible for

Experiment Station. (https://nifa.usda.gov/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

propagating the disease throughout the network. In this paper, we introduce a model that considers how the properties of the network and spatial variance in disease transmission intensity (i.e., the reproduction number) due to social and ecological conditions interact to influence the occurrence of superspreaders. This type of model is applicable to diseases for which habitat suitability is influenced by climate or land cover, such as vector-borne diseases, and to directly transmitted diseases for which population density and practices to mitigate transmission may vary spatially. We present mathematical models that quantify the superspreader capacity of a population node, based on the extent area of the disease spread attributable to that node and the rate at which the disease spreads. We validate our models with a simulation of epidemic spread across randomly generated networks. The statistical model presented here can be used to predict the location of superspreader events in future epidemics and to predict the effectiveness of mitigation strategies that seek to reduce the disease reproduction rate, alter host movements, or both.

Introduction

Network spreading phenomena, including epidemic disease spread and information diffusion on social media, tend to be fueled by a small number of individuals in the network. Known as the 20/80 rule or the Pareto principle [1], this pattern is apparent in a variety of infectious disease systems, including the 2003 SARS outbreak in Hong Kong [2], the 2015 MERS outbreak in South Korea [3], and most recently, the COVID-19 pandemic [4]. The concept of superspreaders can be expanded from individuals to include entire communities. In a metapopulation network in which each node represents a city, for example, a few nodes containing highly trafficked airports or other transportation hubs are typically responsible for propagation of the outbreak throughout the network [5]. The identification of these superspreader nodes is an important topic of research in network science and spatial epidemiology, to reduce the area or velocity of disease outbreak spread.

A node's potential as a superspreader is often estimated based on a variety of network characteristics [6]. Early studies of superspreading dynamics were based on stochastic models, in which each node has a given probability of transmitting the disease to a neighboring node. Therefore, nodes with more neighbors, i.e., higher connectivity, would be expected to spread the disease to a greater portion of the network [7]. More complex models consider not only a node's degree of connectivity but also the connectivity of its neighbors [8]. Superspreader metrics that consider the structure of the entire network include centrality (i.e., the inverse of the node's average distance to all other nodes) and k-core values (i.e., the node's location in the core area of the network) [9], both of which pertain to the potential for a pathogen to disseminate from a given node to the rest of the network. Several studies develop novel definitions of centrality that are particularly well suited to identifying superspreaders because they consider the paths that a pathogen might take to spread through the network [10–13].

None of these network-based approaches to predicting superspreader status consider the common situation that the severity of an epidemic, characterized by either its infection rate or its reproduction number R_0, may vary in space. For instance, for vector-borne diseases, where R_0 is directly related to the habitat suitability for production of the disease vector [14], species distribution models are developed to estimate the probability of presence of vectors based on environmental, climate, and socioeconomic variables [15]. Directly transmitted diseases, such as influenza, are also dependent on spatial factors as their transmission rates may depend on environmental variables such as local temperature and relative humidity [16]. Human factors

such as population density and efforts to mitigate disease spread, such as social distancing, quarantines, and sanitation, also affect transmission rates of directly transmitted diseases based on location [17]. Superimposing a metapopulation network, derived from census data and travel routes, on a spatial map of transmission rates to produce a joint metapopulation network/spatial R_0 model is a useful and underexplored approach to visualize and analyze multiple, interacting potential drivers of pathogen spread simultaneously.

Empirical methods to determine superspreader potential typically involve simulating an outbreak originating from a particular node and measuring the extent of disease spread, in terms of the total number of infected nodes, a process which can be computationally intensive for large networks [18]. This study introduces two definitions of superspreader capacity, based on the number of nodes that become infected and the rate of the disease spread originating from a single node, and provides analytical models to predict these based on a node's connectivity, R_0, and diffusion, along with the properties of its neighbors. We validate these analytical models using a Monte Carlo simulation, involving hundreds of randomized networks superimposed on random R_0 spatial fields. In each simulation, the superspreader capacity of each node is measured, along with several key metrics including degree of connectivity, clustering, centrality, R_0, and diffusion. These data are then used to construct a random forest regression model which predicts the superspreader capacity of any node in a metapopulation network based on its properties [19]. In theory, any real-world metapopulation network, along with a spatial map of R_0 values, can be input to this model to produce a superspreader risk map for a potential future epidemic.

Methods

Metapopulation SIR model

The classic SIR (Susceptible-Infected-Recovered) epidemiological model is a mathematical framework to characterize the transmission dynamics of an infectious disease. Individuals from an at-risk population of size N are classified among three states (i.e., susceptible; infected; or recovered). Individuals transition from the susceptible to the infected state based upon the transmission rate β, and from the infected to the recovered state based upon the recovery rate μ. The stochastic SIR model, used in this article, proceeds as follows:

1. The simulation time is incremented by a chosen value dt.
2. If at least one individual in the population is infected, a random selection of Susceptible individuals become Infected, as a binomial distribution with probability $1 - \left(1 - \frac{\beta * \mathrm{d}t}{N}\right)^I$, where I is the current number of infected individuals. This reflects the assumption that each individual has an independently distributed probability of becoming infected.
3. Simultaneously, a random selection of Infected individuals become Recovered, as a binomial distribution with probability $\mu * \mathrm{d}t$.
4. The simulation repeats, over each time step, until no infected individuals remain.

A metapopulation SIR model is an elaboration of the basic model that considers a network of population centers or nodes, each with its own set of SIR equations, and the rates of population migration between nodes [20]. Individuals migrate from node i to node j based on a mobility matrix M_{ij}, which is typically calibrated by a diffusion parameter p [21]. Here, the processes of transmission, recovery, and migration are simulated stochastically, as described in the Monte Carlo Method section. A node is designated as "infected" when it has at least one infected individual. For the purposes of this study we quantify the spread of the epidemic by

the number of infected nodes over time, as we are concerned with measuring the growing area of the epidemic spread.

Network generation

For our Monte Carlo simulation, we generate a series of random population networks that exhibit properties of scale-free, small-world, and triangulation models [22–24]. This allows us to construct and test models that have a wide range of metrics for connectivity, clustering, and centrality. The algorithm is as follows. A scale-free network with n nodes, each representing a homogeneous community, is constructed from m_0 seed nodes to produce a network with degree distribution $P(k) \propto k^{-3}$ [22]. A forcing algorithm is used to plot this network geographically, scaled so that the average length of any edge is 1. From this, a fraction b of all edges are rewired randomly to generate small-world properties (i.e. a short nodal distance between any pair of nodes). Finally, to simulate node clustering, a Delaunay triangulation method is used to connect any nodes within a certain distance threshold d. The typical uniformly distributed random ranges of values for each parameter are as follows:

- n: 500
- m_0: 1-5
- b: 0-0.1
- d: 0-1

This algorithm creates networks with an average connectivity degree of 3. The majority of nodes have connectivity less than 10, with hub nodes ranging from 20 to 140. About one third of networks have an average clustering coefficient of nearly 0, the rest are distributed between 0 and 0.5, with most networks between 0.25 and 0.4. The average centrality among networks ranges from 0.14 to 0.26, with a peak at 0.18. This algorithm allows for the rapid generation of human mobility networks that bear a resemblance to real-world case studies [25].

Monte Carlo method

After a network has been constructed, a random value of diffusion p is assigned, ranging from 0.1 to 1, and movement heterogeneity θ is assigned as 0.5. A population of $500 * n$ individuals is assigned to the network, such that the population of each node N is proportional to $k^{1+\theta}$. All of these individuals are initially classified as Susceptible. The mobility matrix is then defined based on a traffic-dependent model, $M_{ij} = p\frac{(k_i k_j)^{\theta}}{Ak_i^{1+\theta}}$, where A is a calibration factor. The recovery rate μ is set to 1, and β is assigned over a continuous spatial field with an exponential distribution with mean 1.5, to mimic spatial variability in the reproduction rate of seasonal influenza [26]. R_0 fields generated with this method tend to feature a few hot spots with multiple incidences of spatial clustering. Based on this spatial field, an infection rate value β_i is assigned to each node, and a reproduction rate R_i is computed as β_i/μ.

We introduce 10 infected individuals to one node chosen at random, build a tree data structure with this node at the root, and run a stochastic metapopulation SIR simulation on the network. At each time interval of duration dt, the following steps are taken:

1. In each node i with at least one Infected individual, a random selection of Susceptible individuals become Infected, as a binomial distribution with probability $1 - \left(1 - \frac{\beta_i \mathrm{d}t}{N_i}\right)^{I_i}$. A random selection of Infected individuals become Recovered, as a binomial distribution with probability $\mu * \mathrm{d}t$.

2. A random selection of individuals moves from node i to each neighboring node j, as a binomial distribution with probability $p * dt * \frac{k_j^{1+\theta}}{\sum k_j^{1+\theta}}$ [27]. Each individual who moves between nodes may be Susceptible, Infected, or Recovered.
3. If neighboring node j receives its first infected individual, the current time is recorded, and it is added to the data tree as a subnode under node i.

An example of the data tree is shown in Fig 1. We run the simulation until there are no new infected individuals, typically up to $t = 750$. For each node, we record the epidemic's time of

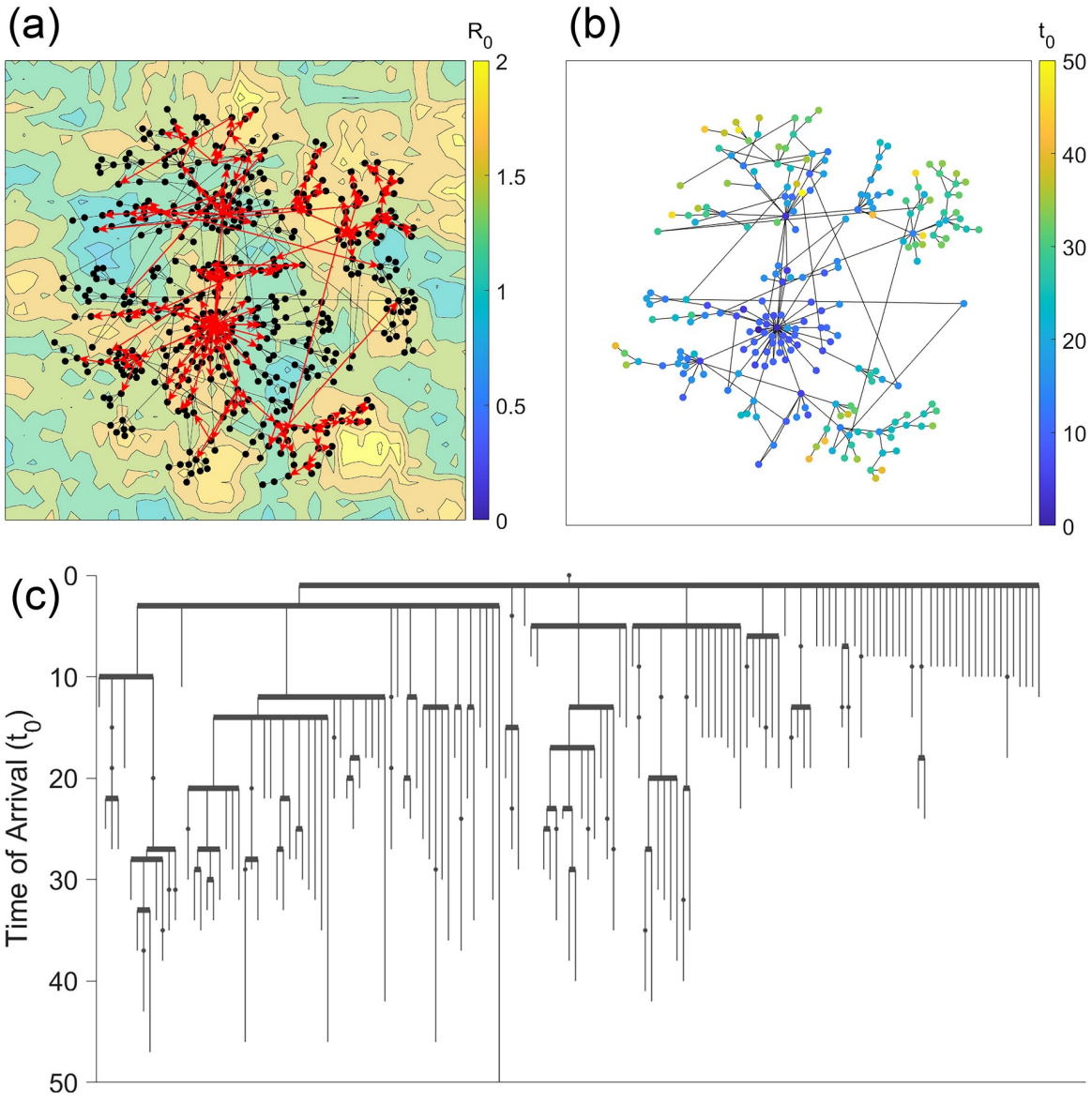

Fig 1. (a) An example of a human network model with 1000 nodes and 1,000,000 individuals. A stochastic metapopulation SIR model was processed on this network, and the red arrows represent the spread of the outbreak. (b) A data tree showing the order in which the outbreak spread throughout the network. Nodes are sorted by the order in which they first spread the infection to a neighboring node, and the y-axis represents the time of arrival of the epidemic.

https://doi.org/10.1371/journal.pcbi.1008674.g001

arrival, peak prevalence, and the infection tree of the network. We repeat this simulation 20 times, each with a different initially infected node, and average our results over these iterations to ensure that the simulation is not biased by the location of the initial outbreak. About 500 networks are processed this way, for a total of about 10,000 metapopulation SIR simulations.

Superspreader capacity and risk factor

For each simulation, we define two metrics of superspreader capacity for each node:

1. **Probability-dependent superspreader capacity**: the total number of children, including subchildren, of each node on the tree graph. This is analogous to the definition of nodal scope introduced by [28].
2. **Time-dependent superspreader capacity**: the average rate, in nodes per unit time, at which the disease spreads from each node to each of its children.

Each metric is averaged for each node over the 20 simulations run, so that the starting point of the epidemic does not bias our numerical results more than about 20%. Each node is assigned a probability-dependent and time-dependent risk index, ranging from 0 to 1, based on its superspreader capacity among all nodes in its network. We hypothesize that risk indices for probability-dependent and time-dependent superspreader capacity will be strongly correlated, but that time-dependent superspreader risk will be less influenced by the starting point of the epidemic.

We run a random forest model using the randomForest package in R [29] to process a model with risk index as the response (ranging from 0 to 1), and with R_i, degree of connectivity, centrality, clustering, and diffusion as predictor variables. The output is a statistical model which takes as input a metapopulation network with an R_i value for each node and outputs a superspreader risk factor index for each node in the network.

Table of variables

Table 1 gives a list of key variables in our metapopulation model and their definition.

Results

This section is divided into three parts. First, we derive analytically a formula for the probability-dependent superspreader capacity, the expected number of nodes to which a certain node

Table 1. A list of key variables used in this article and their definitions.

Symbol	Meaning
n	number of nodes in the network
N	total population of a certain node
$I(t)$	number of infected individuals of a certain node
β	infection rate of the disease in a certain node
μ	recovery rate of the disease
R, R_0	reproduction rate of the disease in a certain node
diffusion p	the fraction of a nodal population that migrates per unit time
θ	heterogeneity of movement of the network
κ	the probability that an individual migrates from one specific node to another per unit time
degree of connectivity k	the number of adjacent nodes connected to a certain node
centrality	the reciprocal of a node's average distance to all other nodes in the network
clustering	the fraction of a node's neighbors that are also connected to each other

https://doi.org/10.1371/journal.pcbi.1008674.t001

is predicted to spread an epidemic. We apply this formula to various randomly generated networks and discuss the correlation between degree of connectivity, reproduction rate R, and superspreader capacity. Next, we derive the time-dependent superspreader capacity, defined as the velocity at which a given node spreads the epidemic to its neighbors, and compare our two definitions of superspreader capacity. Finally, we use a Monte Carlo simulation to develop a Random Forest model relating several network parameters and reproduction rate to a node's risk of becoming a superspreader site.

Probability-Dependent superspreader capacity in uncorrelated graphs

To derive a formula for probability-dependent superspreader capacity, consider an uncorrelated network with nodes of varying degree and vector habitat suitability. Within this network, consider a single node i with degree k_i and reproduction number R_i. This node is adjacent to one node from which the epidemic originated, along with $k - 1$ other uninfected nodes. These $k - 1$ nodes can be decomposed as $k - 1 = k_1 + k_2 + k_3$, where

- k_1 = the number of nodes with $R \gg 1$
- k_2 = the number of nodes with $R \simeq 1$ (typically between 0.8 and 1.2, say)
- k_3 = the number of nodes with $R < 1$

The node will almost certainly spread the outbreak to its neighbors with $R \gg 1$, and although infected individuals may travel to neighboring nodes with $R < 1$, it is very unlikely that an outbreak will be able to take root there. Only nodes with $R \simeq 1$ require special consideration.

The probability that node i of degree k_i and reproduction number R_i will cause an outbreak in an adjacent node j of degree k_j and reproduction number R_j is given by

$$P(\text{outbreak}) = 1 - (R_j)^{-\lambda_{ij}^{R_i}} \quad \text{with} \quad \lambda_{ij}^{R_i} = d_{ij}\frac{\alpha N_i}{\mu} \tag{1}$$

The matrix d_{ij} represents the mobility matrix from node i to node j, generally a function only of their respective degrees, N_i is the population of node i, and μ is the recovery rate [21]. The fraction of the nodal population that becomes infected over the course of the outbreak, designated as α, is a function of R, which can be derived from [30] as

$$\alpha(R) = \begin{cases} 0 & R < 1 \\ 1 + \dfrac{W(-Re^{-R})}{R} & R \geq 1 \end{cases} \tag{2}$$

where $W()$ is the product log function. Note that this function is equal to 0 when $R = 1$ and asymptotically increases to 1 as R increases to infinity.

We use a traffic-dependent mobility model, which assigns a mobility rate $d_{ij} = p\frac{(k_i k_j)^\theta}{T(k_i)}$, where $T(k) = \frac{k^{1+\theta}\langle k^{1+\theta}\rangle}{\langle k\rangle}$, and nodal populations are assigned as $N(k) = \frac{k^{1+\theta}}{\langle k^{1+\theta}\rangle}\bar{N}$, where p is the diffusion rate, θ is the heterogeneity of movement (typically 0.5 or 1), and $\bar{N}$ is the average population among all nodes. This model assumes that a node's population is proportional to its connectivity, and the rate of mobility between two nodes is proportional to the product of their degrees. The diffusion rate, a fixed value between 0 and 1, represents the average rate of human movement and serves as the constant of proportionality.

If we further assume that $R_j \simeq 1$, then the outbreak probability can be approximated to first order [21].

$$1 - (R_j)^{-\lambda_{ij}^{R_i}} \simeq \lambda_{ij}^{R_i}(R_j - 1) \tag{3}$$

Substituting the value for λ_{ij} from Eq 1 and the value of d_{ij} from the traffic-dependent mobility model, we derive the outbreak probability:

$$\begin{aligned} P(\text{outbreak}) &= \left(p\frac{(k_i k_j)^{\theta}\langle k_i\rangle}{k_i^{1+\theta}\langle k_i^{1+\theta}\rangle}\right)\frac{k_i^{1+\theta}}{\langle k_i^{1+\theta}\rangle}\frac{\bar{N}}{\mu}\alpha(R_i)(R_j-1) \\ &= p\frac{\langle k_i\rangle}{\langle k_i^{1+\theta}\rangle^2}\frac{\bar{N}}{\mu}(k_i k_j)^{\theta}\alpha(R_i)(R_j-1) \end{aligned} \tag{4}$$

The superspreader capacity (SSC) of a node is defined as the expected number of neighbors it will infect in the first generation.

$$\text{Superspreader capacity (SSC)} = k * P(\text{outbreak}) \tag{5}$$

Assuming the node's neighbors can be divided into k_1 nodes that will definitely be infected ($P(outbreak) = 1$), k_2 nodes that have the probability given in Eq (4), and k_3 nodes that cannot support an outbreak ($P(outbreak) = 0$), and assuming each outbreak event is independent, the superspreader capacity as defined in Eq 5 can be decomposed as:

$$\begin{aligned} SSC_1 &= k_1 * 1 + k_2\left\langle \sum_{k_j, R_j \simeq 1} p\frac{\langle k\rangle}{\langle k^{1+\theta}\rangle^2}(k_i k_j)^{\theta}\frac{\bar{N}}{\mu}\alpha(R_i)(R_j-1)\right\rangle + k_3 * 0 \\ SSC_1 &= k_1 + k_2\left\langle \sum_{k_j, R_j \simeq 1} p\frac{\langle k\rangle}{\langle k^{1+\theta}\rangle^2}(k_i k_j)^{\theta}\frac{\bar{N}}{\mu}\alpha(R_i)(R_j-1)\right\rangle \\ SSC_1 &= k_1 + k_2\left(p\frac{\bar{N}}{\mu}\frac{\langle k\rangle}{\langle k^{1+\theta}\rangle^2}\right)\left(k_i^{\theta}\alpha(R_i)\right)\left\langle \sum_{k_j, R_j \simeq 1}(k_j)^{\theta}(R_j-1)\right\rangle \\ SSC_1 &= k_1 + k_2\left(p\frac{\bar{N}}{\mu}\frac{\langle k\rangle\langle k^{\theta}\rangle}{\langle k^{1+\theta}\rangle^2}\right)\left(k_i^{\theta}\alpha(R_i)\right)\langle R_j - 1\rangle \end{aligned} \tag{6}$$

where the term $\langle R_j - 1\rangle$ is the mean only over nodes with $R_j \simeq 1$. A good practice is to consider only nodes for which the quantity $\left[\left(p\frac{\bar{N}}{\mu}\frac{\langle k\rangle\langle k^{\theta}\rangle}{\langle k^{1+\theta}\rangle^2}\right)\left(k_i^{\theta}\alpha(R_i)\right)(R_j-1)\right]$ is less than 1. This threshold can vary considerably depending on the values of p, μ, and $\bar{N}$.

Assuming that $k_2 \simeq k_i$, this implies that the superspreader capacity of the node is proportional to $k_i^{1+\theta} * \alpha(R_i)$. A contour plot example of this is shown in Fig 2, with values $p = 0.5, \bar{N} = 1000, \mu = 1, P(k) \propto k^{-3}$, and $\langle R_j - 1\rangle = 0.001$. The isolines in Fig 2 show nodes with varying values of k and R that should have the same superspreader capacity. For example, a node with $k = 14$ and $R = 1.2$ has an equivalent superspreader capacity to a node with $k = 6$ and $R = 2$. The effects of R on superspreader capacity starts to diminish for $R > 3$, but superspreader capacity increases with k indefinitely.

Superspreader capacity in classic network models. The formula given in Eq 4 can be applied directly to a particular metapopulation network to estimate the superspreader risk of each node. To that end, a more precise but less elegant equation for first generation superspreader capacity is found by discarding the assumption that $R \simeq 1$. We substitute Eq 1 and

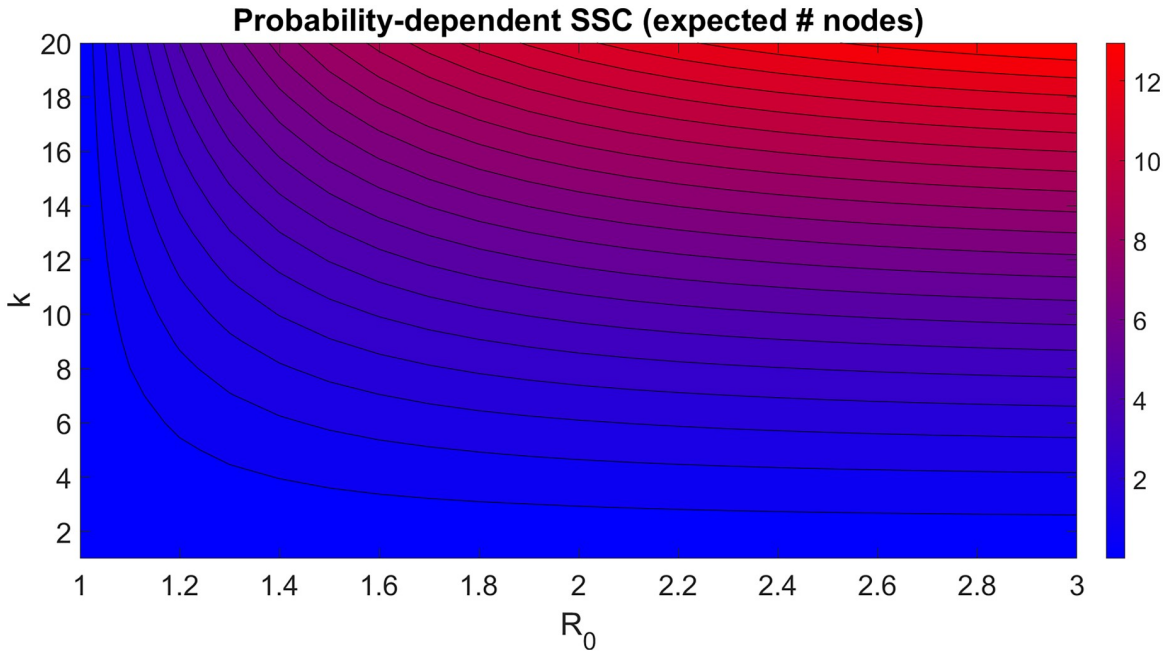

Fig 2. Estimation of probability-dependent superspreader capacity for a node in an uncorrelated network graph, assuming all its neighbors have $R \simeq 1$. In this figure Superspreader capacity is defined as the expected number of neighbors to which the node will spread the outbreak within one generation.

https://doi.org/10.1371/journal.pcbi.1008674.g002

the traffic-dependent mobility model into this definition:

$$\begin{aligned} SSC_1 &= \sum P(\text{outbreak}) \\ SSC_1 &= \sum_{k_j,R_j>1} [1-(R_j)^{-\lambda_{k_ik_j}^{R_i}}] \\ SSC_1 &= \sum_{k_j,R_j>1} \left[1-(R_j)^{-p\frac{\langle k\rangle}{\langle k^{1+\theta}\rangle}\frac{\bar{N}}{2}\mu(k_ik_j)^{\theta}\alpha(R_i)}\right] \end{aligned} \tag{7}$$

where the summation is computed over all neighboring nodes with $R_j \geq 1$. This equation can be used to directly compute the superspreader capacity for each node in a given network. For example, Fig 3A shows a plot of superspreader capacity for a 10,000 node Erdős–Rényi (random) network, where β is randomly, independently assigned as a Weibull distribution with shape factor 1.2 and scale factor 2, and μ is equal to 1. Other parameters are $\bar{N} = 1000, \theta = 0.5, p = 0.1$, and $\mu = 1$. These data closely resemble the analytical prediction in Fig 2. A high degree of connectivity or a high R is sufficient for a superspreader node, although this correlation is diminished with high R. As expected, superspreader capacity is highly correlated with the quantity $k^{1+\theta} * \alpha(R)$, with a linear model fit of $R^2 = 0.72$.

In a Barabási–Albert (scale-free) model, on the other hand, only the hub nodes near the center of the network have the potential to be superspreaders. For individual nodes, a high degree of connectivity and any value of R greater than 1 are necessary conditions for being a superspreader. The correlation between superspreader capacity and the quantity $k^{1+\theta} * \alpha(R)$ is $R^2 = 0.88$, but this is mostly because nodes with high degree also tend to have a high superspreader capacity. Among high degree nodes, there is a strong positive correlation between R and superspreader capacity, but this correlation is not apparent among lower degree nodes.

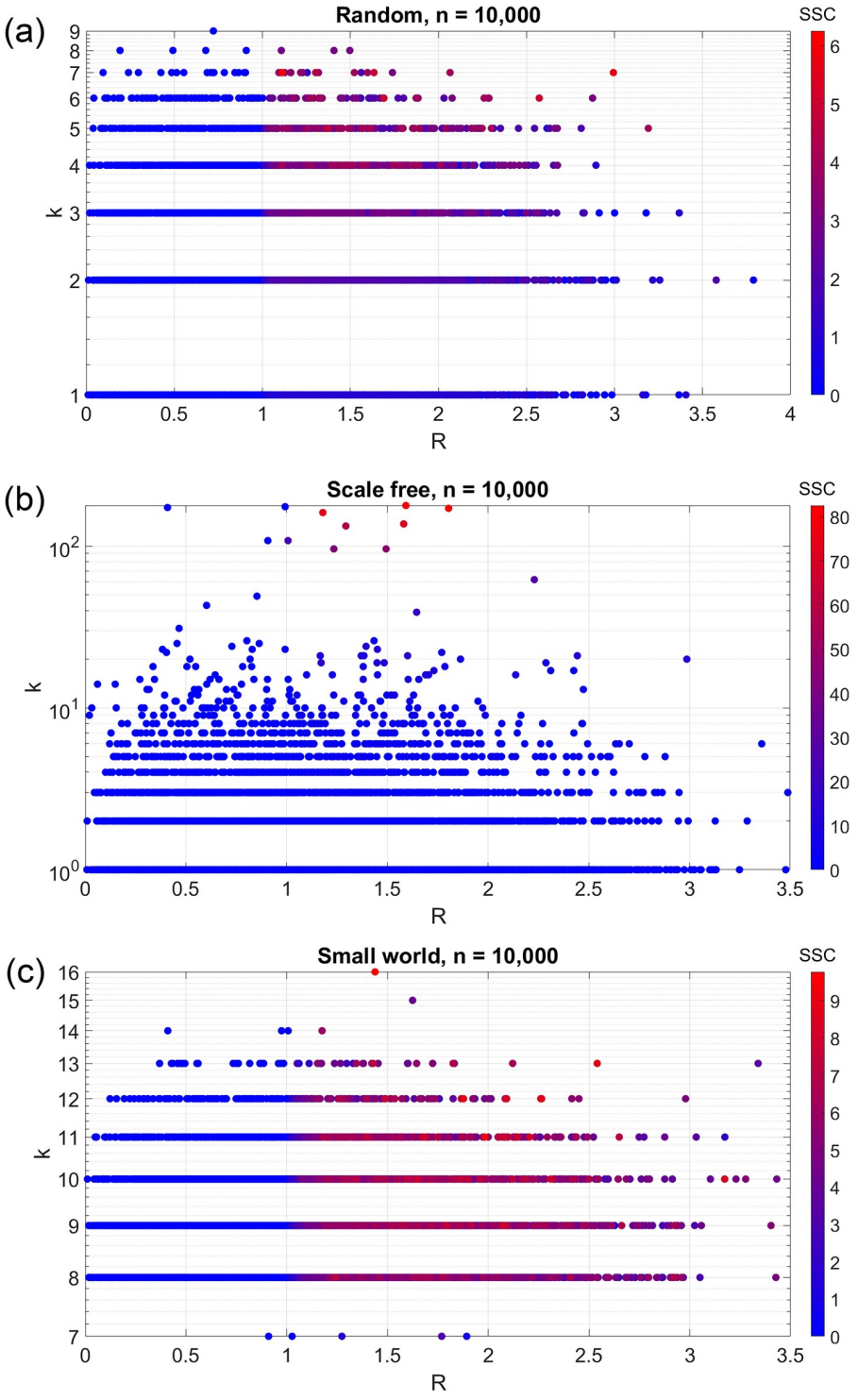

Fig 3. Estimated probability-dependent superspreader capacity in a human population network with 10,000 nodes, with (a) random uncorrelated network, (b) scale free network, and (c) small world network.

https://doi.org/10.1371/journal.pcbi.1008674.g003

In the Watts-Strogatz (small-world) model, every node with $R > 1$ is equally likely to be a superspreader, regardless of its degree of connectivity. This network has the opposite properties of the scale-free network, in that there is a strong positive correlation between R and superspreader capacity, in terms of $\alpha(R)$ ($R^2 = 0.84$), but a weak correlation between degree of connectivity and superspreader capacity ($R^2 = 0.11$). This is due to a small-world model having about four times as many edges as a scale-free model, so it is much more likely that any given node is connected to other nodes with high R.

Time-dependent superspreader capacity

To determine a formula for time-dependent superspreader capacity, we focus on a single node and its neighbors, rather than the network as a whole. Consider a city that is connected to k other cities by road. At time $t = 0$, a number of people I_0 in the hub city have been infected with a contagious illness. Assume the infection spreads exponentially, a valid assumption in the early stages of the outbreak when nearly the entire population is susceptible. The number of infected people over time is therefore given by

$$I(t) = I_0 e^{(\beta-\mu)t} \tag{8}$$

where β and μ are the infection and recovery rates, respectively [31].

The probability that a given individual leaves the hub city for one of the neighboring cities between times t and $t + \mathrm{d}t$ is given by $p * \mathrm{d}t$, where p is the diffusion rate. The probability that individual moves to a particular neighboring city is given by κ_1, κ_2, etc. Note that $\kappa_1 + \kappa_2 + \cdots + \kappa_k = 1$, and that they are constant among all individuals and over time. In a homogeneous system, $\kappa_i = 1/k$ for all i, and in a traffic-dependent system, $\kappa_i = \frac{k_i^\theta}{\sum k_i^\theta}$. Assume each individual's movement is independent from each other.

Therefore, the probability that an individual moves from the hub city to neighboring city i between times t and $t + \mathrm{d}t$ is given by

$$p\kappa_i \mathrm{d}t \tag{9}$$

and the probability that at least one infected individual moves from the hub city to city i is given by

$$1 - [1 - p\kappa_i \mathrm{d}t]^{I(t)} \simeq I(t)p\kappa_i \mathrm{d}t \tag{10}$$

Therefore, $f_i(t) = I(t)p\kappa_i$ is the probability density function for the event that at least one infected individual moves from the hub city to city i at time t.

The probability that at least one infected individual travels between the hub city and city i between time 0 and time T is given by

$$F_i(T) = \int_0^T I(t)p\kappa_i \,\mathrm{d}t = \int_0^T I_0 e^{(\beta-\mu)t} p\kappa_i \,\mathrm{d}t = p\kappa_i \frac{I_0}{\beta-\mu}\left(e^{(\beta-\mu)T} - 1\right) \tag{11}$$

This probability is equal to 0 at time 0 and increases to 1 at time $T_1 = \frac{1}{\beta-\mu} \log\left[1 + \frac{\beta-\mu}{pI_0\kappa_i}\right]$.

The expected value of the time until the first infected individual travels from the hub city to city i is given by

$$\begin{aligned} E_i &= \int_0^\infty [1 - F_i(T)]\,\mathrm{d}T \\ &= \int_0^{T_1} \left[1 - p\kappa_i \frac{I_0}{\beta - \mu}\left(e^{(\beta-\mu)T} - 1\right)\right] \mathrm{d}T \\ &= -\frac{1}{\beta - \mu} + \left[\frac{p\kappa_i I_0 + \beta - \mu}{(\beta - \mu)^2}\right] \log\left[1 + \frac{\beta - \mu}{pI_0\kappa_i}\right] \end{aligned} \tag{12}$$

The probability that at least one infected individual moves from the hub city to *at least* one neighboring city between time 0 and time T is given by

$$\begin{aligned} F_{any}(T) &= 1 - \prod_{i=1}^{k} [1 - F_i(T)] \\ &= 1 - \prod_{i=1}^{k} \left[1 - p\kappa_i \frac{I_0}{\beta - \mu}\left(e^{(\beta-\mu)T} - 1\right)\right] \end{aligned} \tag{13}$$

We evaluate this product with a binomial expansion and ignore higher order terms:

$$\begin{aligned} F_{any}(T) &\simeq 1 - \left[1 - \left(\sum_{i=1}^{k} \kappa_i\right) p \frac{I_0}{\beta - \mu}\left(e^{(\beta-\mu)T} - 1\right)\right] \\ &= p \frac{I_0}{\beta - \mu}\left(e^{(\beta-\mu)T} - 1\right) \end{aligned} \tag{14}$$

This probability approaches 1 at time $T_2 = \frac{1}{\beta-\mu} \log\left[1 + \frac{\beta-\mu}{pI_0}\right]$.

Then the expected value of the time until the first infected individual leaves the hub city to any neighboring city is given by

$$\begin{aligned} E_{any} &= \int_0^\infty [1 - F_{any}(T)]\,\mathrm{d}T \\ &= \int_0^{T_2} \left[1 - p \frac{I_0}{\beta - \mu}\left(e^{(\beta-\mu)T} - 1\right)\right] \mathrm{d}T \\ &= -\frac{1}{\beta - \mu} + \left[\frac{pI_0 + \beta - \mu}{(\beta - \mu)^2}\right] \log\left[1 + \frac{\beta - \mu}{pI_0}\right] \end{aligned} \tag{15}$$

which is incidentally the same value as E_i but with κ_i replaced with 1.

The probability that at least one infected individual leaves the hub city to each of the neighboring cities is given by

$$\begin{aligned} F_{all}(T) &= \prod_{i=1}^{k} F_i(T) \\ &= \prod_{i=1}^{k} p\kappa_i \frac{I_0}{\beta-\mu}\left(e^{(\beta-\mu)T}-1\right) \\ &= \left[p\frac{I_0}{\beta-\mu}\left(e^{(\beta-\mu)T}-1\right)\right]^k * \prod_{i=1}^{k}\kappa_i \end{aligned} \tag{16}$$

Unfortunately, we cannot assume that T is sufficiently small to approximate the exponential term as a polynomial. This probability approaches 1 at $T_3 = \frac{1}{\beta-\mu}\log\left[1+\frac{\beta-\mu}{pI_0}\left(\prod_{i=1}^{k}\kappa_i\right)^{-1/k}\right]$.

Finally, the expected value of the time it takes for an infected individual to travel from the hub city to all neighboring cities is given by

$$\begin{aligned} E_{all} &= \int_0^\infty [1-F_{all}(T)]\,\mathrm{d}T \\ &= \int_0^{T_3}\left[1-\left[p\frac{I_0}{\beta-\mu}\left(e^{(\beta-\mu)T}-1\right)\right]^k * \prod_{i=1}^{k}\kappa_i\right]\mathrm{d}T \end{aligned} \tag{17}$$

This function is not analytically solvable, but it can be computed numerically for given values of p, I_0, β, μ, k, and κ_i. The value of β will always be greater than μ, as a requirement for an outbreak scenario to occur, and we can consider the two extreme cases for β. In the case that $\beta \simeq \mu$, E_{all} can be approximated as:

$$\begin{aligned} E_{all} &\simeq \int_0^{T_3}\left[1-(pI_0T)^k * \prod_{i=1}^{k}\kappa_i\right]\mathrm{d}T \\ &= \frac{1}{pI_0}\frac{k}{k+1}\left(\prod_{i=1}^{k}\kappa_i\right)^{-1/k} \end{aligned} \tag{18}$$

In the case that $\beta \gg \mu$, E_{all} can be approximated as:

$$\begin{aligned} E_{all} &\simeq \int_0^{T_3}\left[1-\left[p\frac{I_0}{\beta-\mu}e^{(\beta-\mu)T}\right]^k * \prod_{i=1}^{k}\kappa_i\right]\mathrm{d}T \\ &= \int_0^{T_3}\left[1-\left[p\frac{I_0}{\beta-\mu}\right]^k e^{k(\beta-\mu)T} * \prod_{i=1}^{k}\kappa_i\right]\mathrm{d}T \\ &\simeq \frac{1}{k(\beta-\mu)}\left[k\log\left[\frac{\beta-\mu}{pI_0}\left(\prod_{i=1}^{k}\kappa_i\right)^{-1/k}\right]-1\right] \end{aligned} \tag{19}$$

As another note, if human movement is homogeneous and $\kappa_1 = \kappa_2 = \cdots = \kappa_k$, then $\kappa_i = 1/k$ for all i and $\left(\prod_{i=1}^{k}\kappa_i\right)^{-1/k}$ is equal to k.

We can therefore define a time-dependent superspreader capacity as the velocity at which a node spreads the disease to at least one neighboring node:

$$V_{any} = \frac{1}{E_{any}} = \left\{ -\frac{1}{\beta - \mu} + \left[\frac{pI_0 + \beta - \mu}{(\beta - \mu)^2} \right] \log \left[1 + \frac{\beta - \mu}{pI_0} \right] \right\}^{-1} \tag{20}$$

or the velocity at which the node spreads the disease to each of its neighbors:

$$V_{all} = \frac{k}{E_{all}} \simeq \frac{k^2(\beta - \mu)}{k \log \left[\frac{\beta - \mu}{pI_0} \left(\prod_{i=1}^{k} \kappa_i \right)^{-1/k} \right] - 1} \tag{21}$$

A contour plot of V_{all} is shown in Fig 4, with $\mu = 1$, $R = \beta/\mu = \beta$, $p = 0.5$, $I_0 = 1$, and assuming homogenous movement. Similar to the computed results of probability-dependent superspreader capacity, time-dependent superspreader capacity increases both with R and k. However, time-dependent superspreader capacity increases indefinitely with R, without diminishing returns, and its value is not dependent on the R value of the node's neighbors. Because of the $(\prod_{i=1}^{k} \kappa_i)^{-1/k}$ term in the denominator of V_{all}, we expect the superspreader capacity of networks with more equitable distributions of nodal degree, such as small-world networks, to be higher than in networks characterized by hub nodes such as scale-free networks. Unlike probability-dependent superspreader capacity, we do not expect time-dependent superspreader capacity to vary significantly across generations.

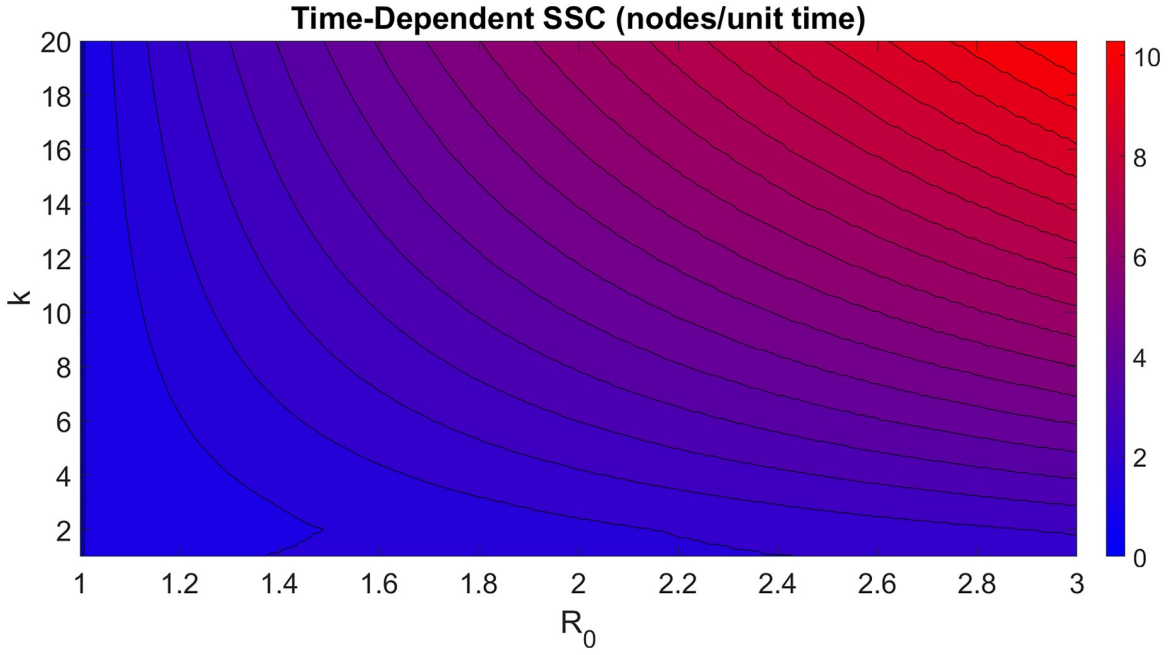

Fig 4. Estimation of time-dependent superspreader capacity for a node in an uncorrelated network graph. Superspreader capacity is defined as the expected velocity at which a node will spread the outbreak to all its neighbors with $R \geq 1$. We do not expect the structure of the network to affect this relationship significantly.

https://doi.org/10.1371/journal.pcbi.1008674.g004

Numerical simulations of superspreader capacity

An analytical model is insufficient to describe the relationship between more abstract network properties, such as centrality and clustering, to superspreader capacity. Therefore, we supplement our model with a Monte Carlo numerical simulation. As described previously, to test the hypothesis that degree of connectivity and R each increase superspreader risk, we randomly generated 500 metapopulation networks, each with about 500 nodes, and each network was simulated 20 times with the initial disease cases occurring in a different node. For each network, the mean values of probability-dependent and time-dependent superspreader capacity were linearly correlated with a correlation coefficient of 97.8%. Therefore, the two metrics of superspreader capacity have very similar means, but different variances. Depending on the point of origin, probability-dependent superspreader capacity varied by about 28%, and time-dependent superspreader capacity varied by about 16%. Running more than 20 simulations did not significantly reduce these variances. When modeling risk indices for future epidemics, time-dependent superspreader capacity may be a more reliable metric if the origin point of the epidemic is unknown.

The output of the Monte Carlo simulation was a list of about 250,000 nodes, about 36% of which had a non-zero time-dependent superspreader capacity, meaning they spread the epidemic to at least one neighboring node. Each node was assigned a risk index, ranging from 0 to 1, based on its superspreader capacity relative to the other nodes within the same network.

We ran a random forest model on 2000 randomly selected training nodes with degrees of connectivity equal to five or greater. After five hundred iterations, the mean square residual error among training nodes was 6.1%. The model revealed that the most important network and epidemic properties for determining the risk index of an individual node, in descending level of importance, are degree of connectivity, R, clustering, centrality, and diffusion. Clustering coefficient is negatively correlated with risk index, the other parameters are all positively correlated. Fig 5 shows the normalized response curve of superspreader risk in terms of each of these factors. Note that for the sake of clarity, each curve has been rescaled to range from "Lowest risk" to "Highest risk." Fig 5 also shows a scatter plot of observed versus predicted risk factor for a randomly selected set of 2000 testing nodes. The root mean squared error in prediction accuracy was 16.5%. This statistical model can predict the presence of nodes in the 90% percentile of risk index with about a 75% success rate, with a Type I error rate of 23% and a Type II error rate of about 25%. In contrast, a model that considers only degree of connectivity or only R has about a 45% success rate (not shown).

Fig 6 shows the two-dimensional response curves of superspreader risk in terms of R and degree of connectivity on the left, and clustering and centrality on the right. Note that the contours are colored by decile, that is, each shaded region represents 10% of tested nodes. For example, the highest risk region on the R-degree of connectivity chart takes up the largest area on the chart, but because degree of connectivity is power-law distributed and R is exponentially distributed, a node only has a 10% chance of lying within that area. These graphs show that a high degree of connectivity is a sufficient condition to be a moderate-risk node, but both high connectivity and high R are necessary for a high-risk node, although connectivity plays a greater role. Similarly, low clustering is a sufficient condition to be a moderate-risk node, but low clustering and high centrality are necessary for a high-risk node.

Discussion

A critical task in epidemic preparedness is to conduct place-based health interventions for population centers that face the highest risk of becoming superspreaders. While previous research has considered the structure of the metapopulation network or spatial heterogeneity

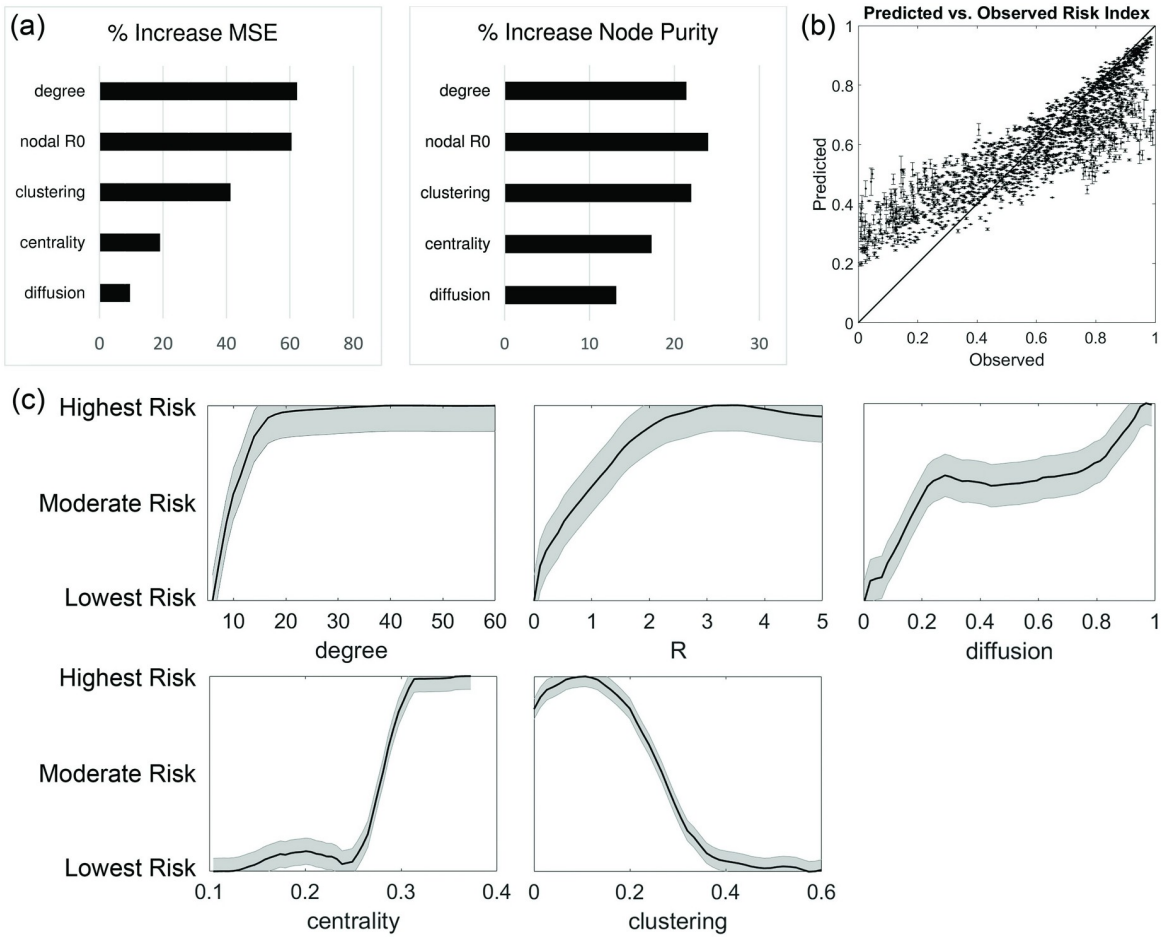

Fig 5. One-variable response curves of superspreader risk in terms of degree, *R*, diffusion, centrality, and clustering. Note that these curves have been rescaled for clarity. A scatter plot showing observed versus predicted risk indices for individual nodes is also shown, with a line indicating a 1:1 relationship.

https://doi.org/10.1371/journal.pcbi.1008674.g005

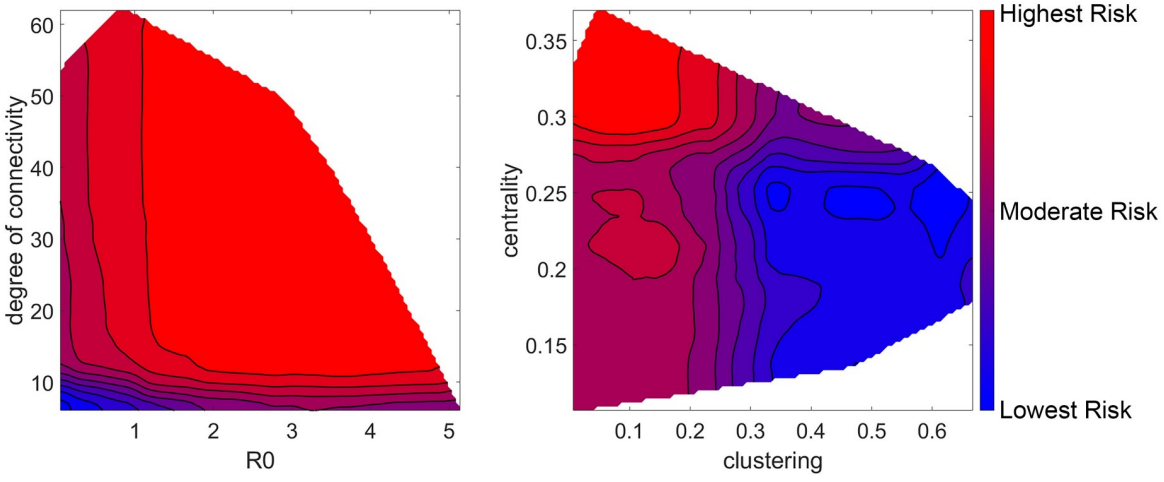

Fig 6. Two-dimensional response curves of superspreader risk index. Each contour line represents one decile of risk, and each shaded area contains roughly 10% of all tested nodes.

https://doi.org/10.1371/journal.pcbi.1008674.g006

in infection rates as predictors of superspreader potential, limited research has considered both factors simultaneously. This manuscript improves on previous knowledge by considering analytically and numerically how both of these factors contribute to nodal risk of superspreader events. Our analyses show that nodal degree of connectivity is indeed the most important factor in determining superspreader capacity, consistent with the mechanistic results demonstrated in [31]. However, the value of R, which may be spatially heterogeneous, is the second most important, followed by other network characteristics such as clustering, centrality, and diffusion. This research extends previous research that demonstrated the superspreader capacity of nodes with high connectivity and centrality [9, 12, 32] and the importance of infection hot-spots to the spread of disease outbreaks [33–35]. This article also complements research that discuss heterogeneous infection rates among subpopulations within a group [36].

The numerical model of risk conforms well with the analytical models presented in the Results section. Degree of connectivity is strongly correlated with superspreader risk only up to about $k = 15$, after which the risk index increases only slightly with degree. However, note the difference between the likelihood and severity of a superspreader event. The probability of a particular node becoming a superspreader does not increase indefinitely with degree of connectivity, but the velocity at which that node could potentially spread the disease does increase indefinitely. This implies that a different calculus must be made for identifying superspreader nodes and estimating the potential harm caused by those nodes. For a scale-free network in particular, the node at the center of the network is almost always the highest risk superspreader, unless its R value is less than 1. This analysis also implies that any mitigation strategy that reduces a node's degree of connectivity, such as travel restrictions, is almost always worthwhile to reduce the rate at which individuals become infected, but extreme mitigation might be necessary to reduce the total number of individuals that become infected over the course of the epidemic.

Although less important than connectivity, the relationship between node centrality and superspreader risk also is positively correlated. Most of the increase in risk occurs when centrality is greater than 0.25, which encompasses about 10% of all nodes in our simulations. In other words, a given node is likely to have a low risk index if within the bottom 90% of nodes, in terms of centrality, but a moderate to high-risk index if within the top 10% of nodes. As a point of comparison, there are 3,143 incorporated counties in the United States, and based on highway and air routes, approximately 300 counties would be high-risk superspreaders based on this metric [37, 38]. The inverse relationship between clustering and superspreader risk is less intuitive. Essentially, nodes with high clustering coefficients have several neighboring nodes that are also adjacent to each other, so the spreading disease is not forced to spread through the node under consideration. Although networks with high rates of diffusion are more prone to spreading the epidemic over shorter time scales [39], diffusion does not significantly affect the prevalence of individual superspreader nodes. This analysis does not consider the presence of community structures, such as a group of nodes that are densely clustered with each other but are loosely connected to neighboring nodes [40, 41]. The effect of nodal communities on superspreader risk merits further research. In addition, future analysis should consider the case in which movement between adjacent nodes is not symmetric, which would require a directed network graph structure for the metapopulation network, as well as networks with multiple layers of mobility [42].

The relationship between R and risk strongly resembles the $\alpha(R)$ function in Eq (2), in that risk increases with R but converges at about $R = 3$. For the small number of nodes with R much greater than 3, there was no correlation between their R values and the area or velocity of the epidemic spread originating from those nodes. This is consistent with case studies in vector-borne and directly transmitted diseases, as an R value of 3 is typically considered the

threshold for a superspreader event [43–45]. Spatial heterogeneity in R can take several forms, each of which has different implications for evaluating superspreader risk. The most common form is continuously varying R, which typically applies when infection risk is correlated with climate or environmental factors [46–48]. In these cases, R does not tend to differ significantly between adjacent nodes, and superspreader nodes are likely to be concentrated in one section of the network. The second form consists of a few hotspot nodes of high R, surrounded by a network with low to moderate R. This is relevant to diseases that are prevalent in high-population centers [49], and these locales tend to be particularly effective superspreaders relative to their neighbors. The third and fourth forms, which are not addressed in this paper, are cases in which R can vary over time, usually as a response to changes in policy during an ongoing epidemic [50, 51], and cases in which R can vary within a node, such as diseases that have varying R values among different species [52]. Both these cases require alternative methods of measuring superspreader capacity than what is considered in this article.

An unanticipated result of the random forest model for superspreader risk is that there is some small degree of risk in nodes with high connectivity and $R < 1$, whereas the analytical models assume that these nodes have no superspreader risk at all. A moderately high value of R or degree of connectivity is a sufficient condition for a node to pose a superspreader risk, and if the node has a degree higher than 10, it is not necessary for R to be greater than 1 for the node to be a superspreader by this method of measurement. The literature does not typically consider such nodes to be outbreak sites, let alone candidates to become superspreaders, and what is likely occurring is that a high number of infected individuals are passing through the node in question, rather than originating from that node. A city that contains a highly trafficked airport but is itself a poor environment for reproduction of the pathogen might be mistaken for a superspreader if the true origins of infected individuals expanding from that node are not considered [5]. This does imply that in a city with high degree of connectivity but low R, restricting travel only from other high risk locations may be sufficient mitigation to reduce spreading.

Previous research has extensively examined either network factors or R_0 heterogeneity as a mechanistic pathway to superspreading events, while very few efforts have attempted to combine them. Our study combines both with a particular focus on the network and habitat suitability properties of individual nodes, as opposed to more complex models that holistically consider the local and global network structure surrounding each node [10–13]. The random forest model tends to overestimate the risk index of low-risk nodes, and underestimate the risk index of high-risk nodes, by an average of 10%. The precision of the model could likely be improved by considering properties such as meta-centrality in conjunction with R and would be relatively straightforward to implement, at the potential cost of narrowing the applicability of the model. The advantage of the model developed in this paper is that the mechanisms that drive superspreader capacity are easily interpreted, and therefore can be applied to a wide variety of epidemic case studies. For example, the relevant balance of superspreader risk factors depends on the nature of the infectious disease in question. Malaria, tends to be more prevalent in rural environments [53], so population centers may have an inverse correlation between degree of connectivity and R, and there may be fewer nodes in the high-risk threshold. Directly transmitted diseases such as influenza and COVID-19, on the other hand, tend to demonstrate a positive correlation between R and population density [54]. In this case, the distribution of superspreader capacity among nodes may be more bimodal, as the highly connected nodes will have extremely high superspreader capacity, and vice versa. For severely contagious diseases with $R_0 > 2$ throughout the entire network, small differences in R may have no bearing on risk when compared to network connectivity. This method of predicting

superspreaders works equally well with other epidemic models, such as Susceptible-Infected-Susceptible (SIS), because the superspreader event typically occurs on a faster time scale than recovery or reinfection.

In addition to predicting superspreader risks within a metapopulation network, the model discussed in this paper may help predict the relative effectiveness of various epidemic mitigation strategies [51, 55]. For example, closing airports or blocking highways has the effect of reducing a node's degree of connectivity and centrality, although the effectiveness of these methods is under question [56], and they require substantial social data to develop an accurate mobility model [57]. Improved hygienic practices and medical resources may reduce the effective R value in regions that follow these practices, which may reduce superspreader risk if R can be reduced to less than 3 [58, 59]. This model can also be used to predict the effectiveness of mitigation strategies if they are not applied uniformly across the network; for example, if only a subset of counties enforce social distancing guidelines to reduce the spread of a directly transmitted pathogen [60]. Future research should also take into consideration nonhomogeneous distributions of population density. Within a given population center, it must be taken into consideration that only a certain fraction of its population will engage in proper mitigation strategies to prevent an epidemic outbreak, which would contribute to a variable R within a node. It is also necessary to test how the shut-down of certain community hubs, such as a school or a shopping center, affects the superspreader risk of the entire network. This analysis may require a more advanced algorithm to generate random metapopulation networks based on multiple spatial scales [25].

Conclusion

This study demonstrates how spatially heterogeneous disease reproduction rates can affect the superspreader potential of nodes across a human metapopulation network. The statistical model, based on a random forest algorithm, could be deployed to predict the risk indices of epidemics and analyze the relative effectiveness of containment and mitigation strategies. The model could be improved by considering other network properties of the network, or the superspreader capacity of neighboring nodes evaluated recursively, but this model strikes a balance between effectiveness and simplicity. A key finding of this study is that the risk of a certain population node becoming a superspreader increases convergently with R and with degree of connectivity. A value of R or degree of connectivity individually above a certain threshold are sufficient for a node to be a moderate-risk superspreader, but both are necessary to be a high-risk superspreader. This finding suggests a balanced approach to addressing both R_0 and network connectivity to achieve optimal epidemic management scenarios.

Acknowledgments

Thank you to our collaborators Brian Allan, Sandra De Urioste-Stone, Andrew Mackay, Aiman Soliman, and Shaowen Wang for their advice and guidance throughout the course of this project.

Author Contributions

Conceptualization: Brandon Lieberthal, Allison M. Gardner.

Formal analysis: Brandon Lieberthal.

Funding acquisition: Allison M. Gardner.

Investigation: Brandon Lieberthal.

Methodology: Brandon Lieberthal, Allison M. Gardner.

Project administration: Allison M. Gardner.

Resources: Brandon Lieberthal.

Software: Brandon Lieberthal.

Supervision: Allison M. Gardner.

Validation: Brandon Lieberthal.

Visualization: Brandon Lieberthal.

Writing – original draft: Brandon Lieberthal.

Writing – review & editing: Allison M. Gardner.

References

1. Stein RA. Super-spreaders in infectious diseases. International Journal of Infectious Diseases. 2011; 15(8):e510–e513. https://doi.org/10.1016/j.ijid.2010.06.020 PMID: 21737332
2. Riley S, Fraser C, Donnelly CA, Ghani AC, Abu-Raddad LJ, Hedley AJ, et al. Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions. Science. 2003; 300 (5627):1961–1966. https://doi.org/10.1126/science.1086478 PMID: 12766206
3. Yang CH, Jung H. Topological dynamics of the 2015 South Korea MERS-CoV spread-on-contact networks. Scientific Reports. 2020; 10(1):1–11. https://doi.org/10.1038/s41598-020-61133-9 PMID: 32152361
4. Endo A, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. Wellcome Open Research. 2020; 5:67. https://doi.org/10.12688/wellcomeopenres.15842.3 PMID: 32685698
5. Nicolaides C, Cueto-Felgueroso L, González MC, Juanes R. A Metric of Influential Spreading during Contagion Dynamics through the Air Transportation Network. PLoS ONE. 2012; 7(7):40961. https://doi.org/10.1371/journal.pone.0040961 PMID: 22829902
6. Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A. Epidemic processes in complex networks. Reviews of Modern Physics. 2015; 87(3). https://doi.org/10.1103/RevModPhys.87.925
7. Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. Physical Review Letters. 2001; 86(14):3200–3203. https://doi.org/10.1103/PhysRevLett.86.3200 PMID: 11290142
8. Zhang D, Wang Y, Zhang Z. Identifying and quantifying potential super-spreaders in social networks. Scientific Reports. 2019; 9(1):1–11. https://doi.org/10.1038/s41598-019-51153-5 PMID: 31616035
9. Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, & Makse HA. Identification of influential spreaders in complex networks. Nature Physics. 2010; 6(11):888–893. https://doi.org/10.1038/nphys1746
10. Bonacich P, Lloyd P. Eigenvector-like measures of centrality for asymmetric relations. Social Networks. 2001; 23(3):191–201. https://doi.org/10.1016/S0378-8733(01)00038-7
11. Barthélemy M, Barrat A, Pastor-Satorras R, Vespignani A. Characterization and modeling of weighted networks. In Physica A: Statistical Mechanics and its Applications. vol. 346. North-Holland; 2005. p. 34–43.
12. Liu JG, Lin JH, Guo Q, Zhou T. Locating influential nodes via dynamics-sensitive centrality. Scientific Reports. 2016; 6(February):1–8. https://doi.org/10.1038/srep21380 PMID: 26905891
13. Madotto A, Liu J. Super-Spreader Identification Using Meta-Centrality. Scientific Reports. 2016; 6(1):38994. https://doi.org/10.1038/srep38994 PMID: 28008949
14. Caminade C, Turner J, Metelmann S, Hesson JC, Blagrove MSC, Solomon T, et al. Global risk model for vector-borne transmission of Zika virus reveals the role of El Niño 2015. Proceedings of the National Academy of Sciences of the United States of America. 2017; 114(7):E1301–E1302. https://doi.org/10.1073/pnas.1614303114 PMID: 27994145
15. Guillera-Arroita G, Lahoz-Monfort JJ, Elith J, Gordon A, Kujala H, Lentini PE, et al. Is my species distribution model fit for purpose? Matching data and models to applications. Global Ecology and Biogeography. 2015; 24(3):276–292. https://doi.org/10.1111/geb.12268

16. Lowen AC, Mubareka S, Steel J, Palese P. Influenza Virus Transmission Is Dependent on Relative Humidity and Temperature. PLoS Pathogens. 2007; 3(10):e151. https://doi.org/10.1371/journal.ppat.0030151 PMID: 17953482

17. Ridenhour B, Kowalik JM, Shay DK. Unraveling R0: Considerations for public health applications. American Journal of Public Health. 2018; 108(2):S445–S454. https://doi.org/10.2105/AJPH.2013.301704r

18. Fu YH, Huang CY, Sun CT. Identifying super-spreader nodes in complex networks. Mathematical Problems in Engineering. 2015; 2015. https://doi.org/10.1155/2015/675713

19. Liaw A, Wiener M. Classification and Regression by randomForest; 2002. 3. Available from: https://www.researchgate.net/publication/228451484.

20. Anderson RM, May RM. Spatial, temporal, and genetic heterogeneity in host populations and the design of immunization programmes. Mathematical Medicine and Biology. 1984; 1(3):233–266. https://doi.org/10.1093/imammb/1.3.233 PMID: 6600104

21. Colizza V, Barthélemy M, Barrat A, Vespignani A. Epidemic modeling in complex realities. Comptes Rendus—Biologies. 2007; 330(4):364–374. https://doi.org/10.1016/j.crvi.2007.02.014 PMID: 17502293

22. Albert R, Barabási AL. Statistical mechanics of complex networks. Reviews of Modern Physics. 2002; 74(1):47–97. https://doi.org/10.1103/RevModPhys.74.47

23. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature. 1998; 393(6684):440–2. https://doi.org/10.1038/30918 PMID: 9623998

24. Lee DT, Schachter BJ. Two algorithms for constructing a Delaunay triangulation. International Journal of Computer & Information Sciences. 1980; 9(3):219–242. https://doi.org/10.1007/BF00977785

25. Staudt CL, Hamann M, Gutfraind A, Safro I, Meyerhenke H. Generating realistic scaled complex networks. Applied Network Science. 2016; 2(1):36. https://doi.org/10.1007/s41109-017-0054-z

26. Biggerstaff M, Cauchemez S, Reed C, Gambhir M, Finelli L. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: A systematic review of the literature; 2014. 1. Available from: http://www.biomedcentral.com/1471-2334/14/480.

27. Matsuki A, Tanaka G. Intervention threshold for epidemic control in susceptible-infected-recovered metapopulation models. Physical Review E. 2019; 100:22302. https://doi.org/10.1103/PhysRevE.100.022302 PMID: 31574659

28. Baranov O. Resource allocation and risk assessment in pandemic situations. Humboldt University; 2019.

29. R Development Core Team 3 5 1. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.; 2018. Available from: http://www.r-project.org.

30. Murray JD. Mathematical Biology I: An Introduction. 3rd ed. Springer; 2007.

31. Colizza V, Vespignani A. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. Journal of Theoretical Biology. 2008; 251(3):450–467. https://doi.org/10.1016/j.jtbi.2007.11.028 PMID: 18222487

32. Dekker AH. Network centrality and super-spreaders in infectious disease epidemiology. Proceedings—20th International Congress on Modelling and Simulation, MODSIM 2013. 2013;(December 2013):331–337.

33. Woolhouse MEJ, Dye C, Etard JF, Smith T, Charlwood JD, Garnett GP, et al. Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. Proceedings of the National Academy of Sciences of the United States of America. 1997; 94(1):338–342. https://doi.org/10.1073/pnas.94.1.338 PMID: 8990210

34. Paull SH, Song S, McClure KM, Sackett LC, Kilpatrick AM, Johnson PTJ. From superspreaders to disease hotspots: Linking transmission across hosts and space. Frontiers in Ecology and the Environment. 2012; Vol. 10, pp. 75–82. https://doi.org/10.1890/110111

35. VanderWaal KL, Ezenwa VO. Heterogeneity in pathogen transmission: mechanisms and methodology Journal of the Royal Society Interface. 2016; 13(121)

36. Van Den Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Mathematical Biosciences. 2002; 180(1-2):29–48. https://doi.org/10.1016/S0025-5564(02)00108-6 PMID: 12387915

37. OpenStreetMap Contributors. OpenStreetMap; 2020. Available from: http://www.openstreetmap.org.

38. OpenFlights Contributors. OpenFlights; 2020. Available from: http://www.openflights.org.

39. Poletto C, Meloni S, Colizza V, Moreno Y, Vespignani A. Host Mobility Drives Pathogen Competition in Spatially Structured Populations. PLoS Computational Biology. 2013; 9(8):e1003169. https://doi.org/10.1371/journal.pcbi.1003169 PMID: 23966843

40. Salathé M, Jones JH. Dynamics and Control of Diseases in Networks with Community Structure. PLoS Computational Biology. 2010; 6(4):e1000736. https://doi.org/10.1371/journal.pcbi.1000736 PMID: 20386735

41. Valdez LD, Braunstein LA, Havlin S. Epidemic spreading on modular networks: The fear to declare a pandemic; In Physical Review E 2020;(Vol. 101) https://doi.org/10.1103/PhysRevE.101.032309 PMID: 32289896

42. Feng S, Jin Z. Infectious diseases spreading on a metapopulation network coupled with its second-neighbor network. Applied Mathematics and Computation. 2019; 361:87–97. https://doi.org/10.1016/j.amc.2019.05.005 PMID: 32287503

43. Mkhatshwa T, Mummert A. Modeling super-spreading events for infectious diseases: Case study SARS. IAENG International Journal of Applied Mathematics. 2011; 41(2):82–88.

44. Ng TC, Wen TH. Spatially Adjusted Time-varying Reproductive Numbers: Understanding the Geographical Expansion of Urban Dengue Outbreaks. Scientific Reports. 2019; 9(1):1–12. https://doi.org/10.1038/s41598-019-55574-0 PMID: 31844099

45. Prakash MK. Eat, Pray, Work: A meta-analysis of COVID-19 Transmission Risk in Common Activities of Work and Leisure. medRxiv. 2020; p. 2020.05.22.20110726.

46. Curran PJ, Atkinson PM, Foody GM, Milton EJ. Linking remote sensing, land cover and disease. Advances in Parasitology. 2000; 47:37–78. https://doi.org/10.1016/S0065-308X(00)47006-5 PMID: 10997204

47. Viboud C, Pakdaman K, Boëlle PY, Wilson ML, Myers MF, Valleron AJ, et al. Association of influenza epidemics with global climate variability. European Journal of Epidemiology. 2004; 19(11):1055–1059. https://doi.org/10.1007/s10654-004-2450-9 PMID: 15648600

48. Bonds MH, Tesla B, Mordecai EA, Demakovsky LR, Murdock CC, Ryan SJ, et al. Temperature drives Zika virus transmission: evidence from empirical and mathematical models. Proceedings of the Royal Society B: Biological Sciences. 2018; 285(1884):20180795. https://doi.org/10.1098/rspb.2018.0795 PMID: 30111605

49. Vlahov D, Galea S. Urbanization, urbanicity, and health. Journal of Urban Health 2002 79:1. 2002; 79 (1):S1–S12. https://doi.org/10.1093/jurban/79.suppl_1.S1 PMID: 12473694

50. Jia J, Ding J, Liu S, Liao G, Li J, Duan B, et al. Modeling the Control of COVID-19: Impact of Policy Interventions and Meteorological Factors. Electronic Journal of Differential Equations. 2020; 2020.

51. Feng S, Jin Z. Infectious Diseases Spreading on an Adaptive Metapopulation Network. IEEE Access. 2020; 8:153425–153435. https://doi.org/10.1109/ACCESS.2020.3016016

52. Hamer GL, Kitron UD, Goldberg TL, Brawn JD, Loss SR, Ruiz MO, et al. Host selection by Culex pipiens mosquitoes and west nile virus amplification. American Journal of Tropical Medicine and Hygiene. 2009; 80(2):268–278. https://doi.org/10.4269/ajtmh.2009.80.268 PMID: 19190226

53. Mathanga DP, Tembo AK, Mzilahowa T, Bauleni A, Mtimaukenena K, Taylor TE, et al. Patterns and determinants of malaria risk in urban and peri-urban areas of Blantyre, Malawi. Malaria Journal. 2016; 15(1):590. https://doi.org/10.1186/s12936-016-1623-9 PMID: 27931234

54. Delamater PL, Street EJ, Leslie TF, Yang YT, Jacobsen KH. Complexity of the basic reproduction number (R0). Emerging Infectious Diseases. 2019; 25(1):1–4. https://doi.org/10.3201/eid2501.171901 PMID: 30560777

55. Hollingsworth TD, Klinkenberg D, Heesterbeek H, Anderson RM. Mitigation strategies for pandemic influenza a: Balancing conflicting policy objectives. PLoS Computational Biology. 2011; 7(2). https://doi.org/10.1371/journal.pcbi.1001076 PMID: 21347316

56. Errett NA, Sauer LM, Rutkow L. An integrative review of the limited evidence on international travel bans as an emerging infectious disease disaster control measure. Journal of emergency management 2020; 18.1: 7–14. https://doi.org/10.5055/jem.2020.0446 PMID: 32031668

57. Panigutti C, Tizzoni M, Bajardi P, Smoreda Z, Colizza V. Assessing the use of mobile phone data to describe recurrent mobility patterns in spatial epidemic models. Royal Society Open Science. 2017; 4(5):160950. https://doi.org/10.1098/rsos.160950 PMID: 28572990

58. Ochoche MJ, Madubueze EC, Akaabo TB. A mathematical model on the control of cholera: hygiene consciousness as a strategy. J Math Comput Sci. 2015; 5(2):172–187.

59. Mbuthia FK, Chepkwony I. Mathematical Modelling of Tungiasis Disease Dynamics Incorporating Hygiene as a Control Strategy. Journal of Advances in Mathematics and Computer Science. 2019; p. 1–8. https://doi.org/10.9734/jamcs/2019/v33i530190

60. Matrajt L, Leung T. Evaluating the Effectiveness of Social Distancing Interventions to Delay or Flatten the Epidemic Curve of Coronavirus Disease. Emerging Infectious Diseases. 2020; 26(8). https://doi.org/10.3201/eid2608.201093 PMID: 32343222