Mathematical Biology

A note on species richness and the variance of epidemic severity

Peter Shaffery[1] · Bret D. Elderd[2] · Vanja Dukic[1]

Abstract

The commonly observed negative correlation between the number of species in an ecological community and disease risk, typically referred to as "the dilution effect", has received a substantial amount of attention over the past decade. Attempts to test this relationship experimentally have revealed that, in addition to the mean disease risk decreasing with species number, so too does the *variance* of disease risk. This is referred to as the "variance reduction effect", and has received relatively little attention in the disease-diversity literature. Here, we set out to clarify and quantify some of these relationships in an idealized model of a randomly assembled multi-species community undergoing an epidemic. We specifically investigate the variance of the community disease reproductive ratio, a multi-species extension of the basic reproductive ratio R_0, for a family of random-parameter community SIR models, and show how the variance of community R_0 varies depending on whether transmission is density or frequency-dependent. We finally outline areas of further research on how changes in variance affect transmission dynamics in other systems.

Keywords Disease ecology · Biodiversity · Modeling

Mathematics Subject Classification 92 · 15 · 37

✉ Peter Shaffery
   pesh5067@colorado.edu

1   Department of Applied Mathematics, University of Colorado Boulder, Engineering Center, ECOT 225, 526 UCB Boulder, CO 80309-0526 Boulder, USA

2   Louisiana State University, Baton Rouge, LA, USA

1 Introduction

A major focus of research efforts in the field of disease ecology is the effect of biodiversity on epidemic behavior or pathogen prevalence ("disease-diversity" relationships). The most commonly discussed such relationship is the "dilution effect", which is an observed negative correlation between the number of different host species (species richness) and the incidence or risk of a specific disease (Ostfeld and Keesing 2000; Civitello et al. 2015). The idea that greater biodiversity inhibits disease risk is attractive, intuitive, and easily reinforced by occasional catastrophic outbreaks in monocultures (Curl 1963; Vandermeer 1992). However, evidence exists to suggest that opposite "amplification effects" can sometimes also occur, where increased species richness results in an increase in disease risk (Salkeld et al. 2013; Randolph and Dobson 2012).

While dilution and amplification have received the bulk of attention, higher-order relationships between disease incidence have recently started to become of interest as well (Buhnerkempe et al. 2015; Johnson et al. 2015). One such property, the focus of this work, is the "variance reduction effect", which refers to a phenomenon in which the variability of disease risk (as measured by either pathogen prevalence or the community disease reproductive ratio) decreases as species richness increases. Understanding this and other higher-order phenomena is important as the focus on mean behavior in disease ecology may not only miss important diversity relationships, but also result in biased inference and predictions. Within a single population, susceptibility to the pathogen or infectivity often varies a great deal amongst individuals (Dwyer et al. 1997; Lloyd-Smith et al. 2005). When considering susceptibility, individual variability leads to a decidedly non-linear relationship in infection dynamics such that as pathogen levels increase transmission asymptotes at ever higher pathogen amounts (Elderd et al. 2013). Standard models that do not account for variation in susceptibility show a linear relationship to an increase in pathogen, which does not always hold when confronted with the empirical data (Dwyer et al. 1997). At the between-population scale, variance reduction suggests that diverse host communities may be less susceptible to extreme epidemics, and that their dynamics are more predictable and perhaps controllable. In addition, understanding when to expect variance reduction, or if other variance effects are possible or even likely at play, can help to determine the role that diversity plays in epidemic modeling and inference as well as how variance reduction effects may help stabilize communities.

While the variance reduction effect has been observed in randomized experiments, so far the underlying mechanism is not precisely understood (Mitchell et al. 2002; Rottstock et al. 2014; Mihaljevic et al. 2014). The most popular explanation argues that the variance reduction effect is a "selection effect" (Huston 1997); that communities with increased numbers of species are more likely to have a high degree of overlap, and therefore share properties which result in a similar response to disease. This results in a decrease in the variability of associated disease risk in species rich communities as compared to species poor communities.

Nevertheless, controlling for the effect of community composition (e.g. average pathogen susceptibility of the member host species) in experimental communities does not appear to remove the variance reduction effect, as would be expected if a selection effect were the driving mechanism (Mitchell et al. 2002). This suggests that

the selection effect explanation of variance reduction is incomplete and that the reduced variance in pathogen prevalence among species rich communities is due to some other factors besides increased similarity among host populations. Furthermore, variance reduction also appears *in silico*, in simulations of dynamical models of epidemic spread with random parameters (Mihaljevic et al. 2014). As these models allow for explicit control of all variables, it is therefore unlikely that the unexplained variance reduction is due to a confounding variable, as was proposed in Mitchell et al. (2002).

In this work we set out to explore the mathematical properties of variance reduction using a community, Susceptible-Infected-Recovered (SIR) model of an epidemic which infects a community of randomly selected species. For this model we derive and analyze the "Next Generation Matrix" (NGM) and its spectral norm, which we term the "community invasion factor". This definition of community R_0 approximates the average rate of new infections per individual, generalizing the reproductive ratio from classical epidemic modeling (Diekmann et al. 1990).

Due to the randomization of community host species, our NGM is a random matrix and so its spectral norm will also be a random variable. We derive bounds for the variance of this distribution and thus the community R_0, which depend on the species richness of the host community. In doing so, we show that the presence and strength of a variance reduction phenomenon hinges on the pathogen transmission mode (density or frequency-dependent), and that when disease is density dependent a variance amplification (an increase in variance with species richness) is also possible.

2 Model formulation

2.1 Epidemic model

We consider a random-parameter community epidemic model, with SIR dynamics describing the course of the disease within and between n host species. Each species within the community is undergoing an epidemic of the same pathogen, according to the SIR model. Our choice of model and assumptions follows Dobson (2004), and reflects a balance of realism and tractability.

Our multi-host SIR epidemic model is:

$$\dot{S}_i = -p_i \sum_{j=1}^{n} \beta_{ij} I_j$$
$$\dot{I}_i = p_i \sum_{j=1}^{n} \beta_{ij} I_j - \alpha_i I_i \qquad (1)$$
$$\dot{D}_i = \alpha_i I_i$$

and it assumes that for host species i, living individuals fall into one of three categories: susceptible (S_i), infected (I_i), or recovered (D_i). Additionally we assume that species abundances N_i are constant for the duration of the epidemic, i.e. we do not account for population dynamics, effectively assuming that the population dynamics happens

Table 1 Definitions of parameters used in the model in Eq. 1

Parameter	Definition
β_{ij}	Per capita rate of pathogen transmission from host species i to host species j
α_i	Recovery rate of host species i when infected with the pathogen
K_i	Population size of species i in the absence of the pathogen
N_i	Total population size ($N_i = S_i + I_i + Di$)
c_{ij}	Parameter that modifies between-species transmission from an arithmetic mean of their within-species rates

at a slower rate than the epidemiological dynamics, which is common Anderson and May (1982); Keeling and Rohani (2011).

For the between-species transmission rates, we use the convention that $\beta_{ij} = c_{ij}\frac{\beta_{ii}+\beta_{jj}}{2}$ where $c_{ij} \geq 0$. Note that we let $c_{ij} \neq c_{ji}$, which allows for transmission to be asymmetric between species. The p_i factors reflect whether transmission is density-dependent ($p_i = S_i$) or frequency-dependent ($p_i = \frac{S_i}{\sum_i N_i}$). Density dependence refers to transmission dynamics where the rate of infectious contacts increases in proportion to population size, whereas frequency dependence assumes that it is constant. A canonical example of the former would be the flu, and of the latter would be a sexually transmitted infection. See Table 1 for the definitions of all model parameters in the model Eq. 1.

To account for the randomization employed in experiments, we treat the parameters of the ODEs in (1) as random variables, with the exception of the c_{ij}, which we assume to be fixed givens. Reserving the c_{ij} as fixed allows us to account for different regimes of between-species transmission explicitly. In our proof and simulation we will assume that all $c_{ij} = c$, and by varying c we can look at variance reduction in communities with overall low- and high- between-species infection. Furthermore, allowing for large numbers of the c_{ij} to be set equal to 0 (as would be the case with a vector-borne pathogens) may constrain variance reduction more than the bound derived below, although we do not deal with this possibility here.

Based on this random ODE community SIR model, we will derive bounds for the variance of disease risk as a function of species richness, under different assumptions about the transmission mode of the disease.

2.2 Measures of epidemic severity

2.2.1 The next generation matrix

For many classes of multi-host epidemic models the severity of the epidemic can be captured through the "Next Generation Matrix" (NGM) (Diekmann et al. 1990). Denoting the NGM matrix as G, its elements G_{ij} can be interpreted as the expected number of new infections an individual of type j produces among susceptibles of type i, over its entire infectious duration. Letting $\phi_i^{(1)}$ denote the number of infected individuals of type i in the first generation during the initial disease outbreak, then

the expected number of infected individuals of type i in the second generation is $\phi_i^{(2)} = \sum_j G_{ij}\phi_j^{(1)}$. For the entire community, we have that G pushes the initial outbreak vector forward to the second generation vector, $\boldsymbol{\phi}^{(2)} = G\boldsymbol{\phi}^{(1)}$, where $\boldsymbol{\phi}^{(m)} = [\phi_1^{(m)}, \ldots, \phi_n^{(m)}]^T$.

For our model, the NGM can be derived from the Jacobian J of the $\dot{I}_i$ with respect to all I_j, evaluated at $I_j = 0$, the so-called disease-free equilibrium. Letting $s(J) = \max(\{\text{Re}(\lambda), \lambda \in \sigma(J)\})$ where $\sigma(J)$ is the spectrum of J, the disease-free equilibrium is unstable whenever $s(J) > 0$. When the disease-free equilibrium is unstable then small outbreaks of the disease will spread into full epidemics.

For the system in Eq. (1), direct calculation gives that $J_{ij} = \beta_{ij}p_i - \alpha_i\delta(i-j)$, where $\delta(i-j)$ is the Kronecker delta function. This implies that $J = T - \Sigma$, where T is referred to as the "transmission matrix" and represents the rate at which new infections are being created. Σ is a diagonal matrix whose nonzero elements are α_i and is referred to as the "recovery matrix", and represents that rate at which infected individuals recover. Given this partition of J we define the NGM as $G = T\Sigma^{-1}$; intuitively this can be thought of as "rate of infection over rate of recovery".

It can be quickly verified that the elements of G for Eq. (1) are $G_{ij} = \frac{\beta_{ij}}{\alpha_i}$. However, given our assumption that $\beta_{ij} = \frac{c_{ij}}{2}(\beta_{ii} + \beta_{jj})$ it is convenient to reparameterize G in terms of its diagonal elements $R_i = \frac{\beta_{ii}}{\alpha_i}$. The NGM for a community with n host species then has the form:

$$G_n = \begin{bmatrix} p_1 R_1 & \cdots & \frac{c_{1n}p_1}{2}\left(R_1 + \frac{\alpha_n}{\alpha_1}R_n\right) \\ \vdots & \ddots & \vdots \\ \frac{c_{n1}p_n}{2}\left(R_n + \frac{\alpha_1}{\alpha_n}R_1\right) & \cdots & p_n R_n \end{bmatrix} \tag{2}$$

Observe that, at the disease-free equilibrium, $p_i = K_i$ when transmission is density-dependent and $p_i = \frac{K_i}{\sum_i K_i}$ when transmission is frequency-dependent.

2.2.2 Community R_0 and community invasion factor

Having calculated the NGM G_n, the question now is how best to extract information from it about the severity of the epidemic. The canonical choice would be to take its spectral radius, $\rho(G_n)$, as was done by Diekmann et al. in their 1990 paper on the definition of R_0 for heterogeneous populations. This choice was motivated by the observation that, for any matrix norm $||\cdot||$, $\rho(G_n) = \lim_{m\to\infty} ||G_n^m||^{\frac{1}{m}}$, so in a certain sense the spectral radius of G_n captures the average, long-term growth factor of the initial outbreak. Furthermore, it can be shown that $\text{sign}(s(J)) = \text{sign}(\rho(G_n) - 1)$, where $s(J)$ denotes the modulus of the largest eigenvalue of J. This means that the spectral radius of G_n also captures the stability of the disease-free equilibrium. For brevity we do not include a proof of this fact, but a straightforward one can be found in the appendix of Diekmann et al. (2009).

In its original derivation, however, the NGM is derived as an infinite-dimensional forward operator $\mathcal{G}_n$ for the dynamical system underlying the epidemic model (Diek-

mann et al. 1990). Its repeated application, denoted $\mathcal{G}_n^m$, produces the exact trajectory of the epidemic for discrete time-steps m. However in 2.2.1 our derivation of $\boldsymbol{G}_n$ is based on a linearized version of the underlying model, and its iterated application does not produce the true epidemic, only a short-time approximation of it. From that perspective, the large-m behavior of $||\boldsymbol{G}_n^m||^{\frac{1}{m}}$ is less relevant than when $m = 1$.

Furthermore, the spectral radius $\rho(\boldsymbol{G}_n)$ is a mathematically inconvenient definition of epidemic severity. The spectral radius lacks a number of properties found in matrix norms, which make it challenging to derive results about them for general random matrices. For example: spectral radii do not obey a triangle inequality; it is possible that $\rho(\boldsymbol{M} + \boldsymbol{N}) > \rho(\boldsymbol{M}) + \rho(\boldsymbol{N})$ for matrices $\boldsymbol{M}$ and $\boldsymbol{N}$. As we show, properties like a triangle inequality will be important in proving our results.

For these reasons we will instead rely on the spectral norm $||G_n||_2$, to measure epidemic severity. This can be understood as measuring the maximum second generation growth factor of the infective population, since $||G_n||_2 = \max_\phi \frac{||G\boldsymbol{\phi}^{(1)}||}{||\boldsymbol{\phi}^{(1)}||}$. We therefore term this measure the "community invasion factor", to distinguish it from the community R_0, $\rho(G_n)$.

2.2.3 Comparing spectral norm and spectral radius for epidemic random matrices

We would like to assess whether our definition of epidemic severity $||\boldsymbol{G}_n||$ is "equivalent" to the canonical choice ($\rho(\boldsymbol{G}_n)$). Equivalency of metrics can have several meanings, but here we will consider just two. First, we test the extent to which exchanging $\rho(\boldsymbol{G}_n)$ and $||\boldsymbol{G}_n||_2$ will preserve the ordering; i.e., whether for two random matrices $\boldsymbol{G}_n$ and $\boldsymbol{G}_n'$ with $\rho(\boldsymbol{G}_n) \leq \rho(\boldsymbol{G}_n')$ we also have that $||\boldsymbol{G}_n||_2 \leq ||\boldsymbol{G}_n'||_2$ (termed "monotonicity"). Second, we compare how effectively $\rho(\boldsymbol{G}_n)$ and $||\boldsymbol{G}_n||_2$ predict total epidemic size for numerically solved trajectories of epidemic model (1) (termed "predictiveness"). If $\rho(\boldsymbol{G}_n)$ and $||\boldsymbol{G}_n||_2$ are equivalent in these ways, then we can meaningfully refer to "high" or "low" community R_0, regardless of which definition we use.

To assess monotonicity and predictiveness we will look at the sampling distributions of $\rho(\boldsymbol{G}_n)$, $||\boldsymbol{G}_n||_2$, and numerical epidemic size under four cases: assuming either a small community ($n = 5$) or a large community ($n = 50$), and either realistic or unrealistic distributions for the R_i, K_i, and α_i. The realistic parameter distributions are drawn from natural systems and follow Mihaljevic et al. (2014). In this case about 95% of the R_i fall in the range 0–12 and we see that the spectral radius (community R_0) stays below 1 (Fig. 1). This is because we are using the frequency dependent case of the NGM, where its elements are scaled by a p_i that grows as $\mathcal{O}(\frac{1}{n})$. Therefore in the unrealistic case we assigned each parameter a Gamma density with shape and scale parameters chosen to produce a wide range of community R_0 values which covers community $R_0 = 1$. In all cases we set $c_{ij} = .05$ (following the value chosen in Mihaljevic et al. (2014)).

After drawing from these sampling distributions we quantify predictiveness using simple linear regression: we perform linear regression between $\rho(\boldsymbol{G}_n)$ and total percent infected, and between $||\boldsymbol{G}_n||_2$ and total percent infected (Fig. 1). To quantify monotonicity we use the Kendall Tau rank correlation coefficient (Fig. 2). Kendall Tau

tests the extent to which two random variables have a monotonic relationship, without specifying the functional form of that relationship. It is calculated by comparing pairs $(\rho(G_n), ||G_n||_2)$ and $(\rho(G'_n), ||G'_n||_2)$. If $\rho(G_n) \geq \rho(G'_n)$ and $||G_n||_2 \geq ||G'_n||_2$ the comparison is said to be "cordant", otherwise they are termed "discordant". All pairs of samples are compared (for n samples we make $n(n-1)/2$ comparisons) and the Kendall Tau coefficient is then calculated by $\tau := \frac{n_c - n_d}{n_c + n_d}$, where n_c is the number of cordant comparisons and n_d is the number of discordant ones. A value of $\tau = \pm 1$ implies a perfectly monotonic relationship (either increasing or decreasing, respectively).

From Fig. 1 we see that, for the realistic parameter distributions, both metrics predict epidemic severity fairly well, however the canonical choice $\rho(G_n)$ outperforms our candidate somewhat, for all cases. We attribute this to the stability interpretation of $\rho(G_n)$, which provides some information about the long-time behavior of the epidemic, whereas $||G_n||$ provides information purely regarding the short term growth rate. Figure 2 indicates that the relationship between $\rho(G_n)$ and $||G_n||_2$ is fairly monotonic (and increasing), as desired. We note that Kendall Tau does decrease somewhat as species richness n grows large, we understand this behavior by observing that when $n = 1$ the spectral radius and spectral norm are exactly equal and so the Kendall Tau is 1. As n grows, the increasing volume of possible matrices G_n means that it becomes more probable for spectral norm to decrease while the radius increases, or vice versa.

While these results do not indicate that our proposed metric of epidemic severity should replace $\rho(G_n)$, they do suggest that it will suffice for our analysis of variance reduction. Clearly $||G_n||_2$ is not as predictive of epidemic size as $\rho(G_n)$, however the correlation is strong enough that variance reduction effects present in $||G_n||_2$ should still manifest in direct measurement of epidemic spread. Furthermore, given its relationship with $\rho(G_n)$ it may be possible to reformulate our results for the canonical metric as well, although we will not do so here.

3 Results and proof

Going forward we will use the notation $r(G_n) = ||G_n||_2$. Recall that our model represents an approximation of natural community assembly processes wherein a host community is populated with randomly selected species from a global pool of potential members (Mitchell et al. 2002). In our model each host species is uniquely defined by the random triple of positive reals (R_i, α_i, K_i). While we allow for dependence between the members of a single triple, we assume that any two distinct triples (i.e. species) are independent. With this randomness, our NGM (2) is thus a random matrix, and its spectral norm is a random variable, whose distribution is induced by the distribution of (R_i, α_i, K_i). We can now define a variance reduction effect as occurring if $\mathrm{Var}_{R_i, \alpha_i, K_i}[r(G_n)]$ is decreasing in n, the number of unique host species.

Our main result is the following:

Theorem 1 *Let $R = [R_1, \ldots, R_n]^T$, $\alpha = [\alpha_1, \ldots, \alpha_n]^T$, and $K = [K_1, \ldots, K_n]^T$ be random, real valued vectors with independent elements. Assume that for all n, the support of R, α, and K is such that $R_i \leq \Gamma$, $0 < A_1 \leq \alpha_i \leq A_2$, and $K_i \leq \kappa$. Let*
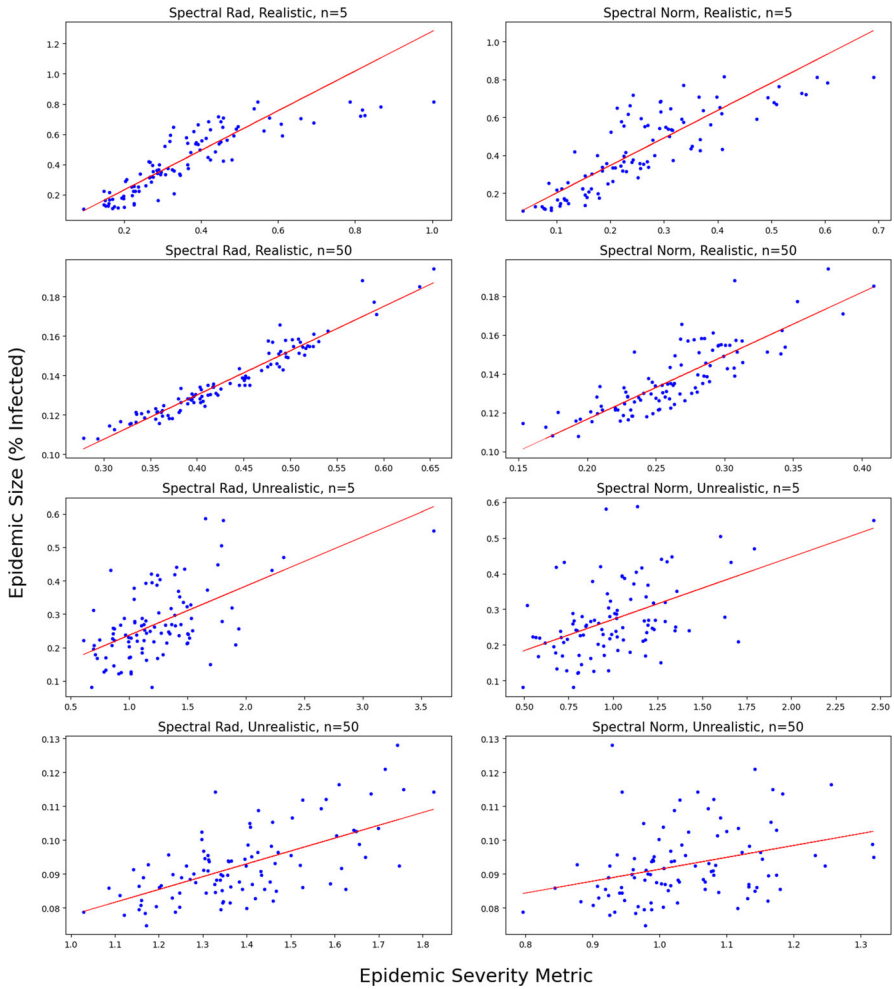
Fig. 1 Estimated relationship between total epidemic size and $\rho(\boldsymbol{G}_n)$ (spectral radius) and $||\boldsymbol{G}_n||_2$ (spectral norm). Epidemic size was defined as total number of individuals infected over the course of the epidemic $(\sum_i \int I_i(t)dt)$ as a percentage of community abundance $(\sum_i N_i)$, and was calculated with $M = 100$ random draws of $\rho(\boldsymbol{G}_n)$, $||\boldsymbol{G}_n||_2$, and percent infected. Epidemic size was calculated by numerically solving Eq. (1) using Runge-Kutta (4,5) for 100 time steps of size $\delta t = 1$. Subplot headers indicate epidemic metric (canonical "spectral radius" or proposed "spectral norm") and underlying parameter distribution (realistic or unrealistic)

$\boldsymbol{G}(\boldsymbol{R}, \boldsymbol{\alpha}, \boldsymbol{K})$ *be a matrix-valued function defined in Eq.* 2*, under frequency-dependent transmission, i.e.* $p_i = \frac{K_i}{\sum_i K_i}$. *Then* $Var[||\boldsymbol{G}(\boldsymbol{R}, \boldsymbol{\alpha}, \boldsymbol{K})||_2] \sim \mathcal{O}(n^{-1/2} + 1)$.

Before we begin the proof of this result, it is constructive to look at why we might intuitively expect it to be true. First, in the case of frequency-dependent transmission, the $p_i \sim \mathcal{O}(n^{-1})$. We can then "pull out" a factor of $\frac{1}{n}$ from $\boldsymbol{G}_n$ by writing $\boldsymbol{G}_n = \frac{1}{n}\boldsymbol{G}'_n$. This implies that $Var[r(\boldsymbol{G}_n)] = \frac{1}{n^2}Var[r(\boldsymbol{G}'_n)]$. Second, we observe that for many classes of random matrices $\boldsymbol{M}$ one has that $Var[r(\boldsymbol{M})] \sim o(n)$. Intuitively this can be
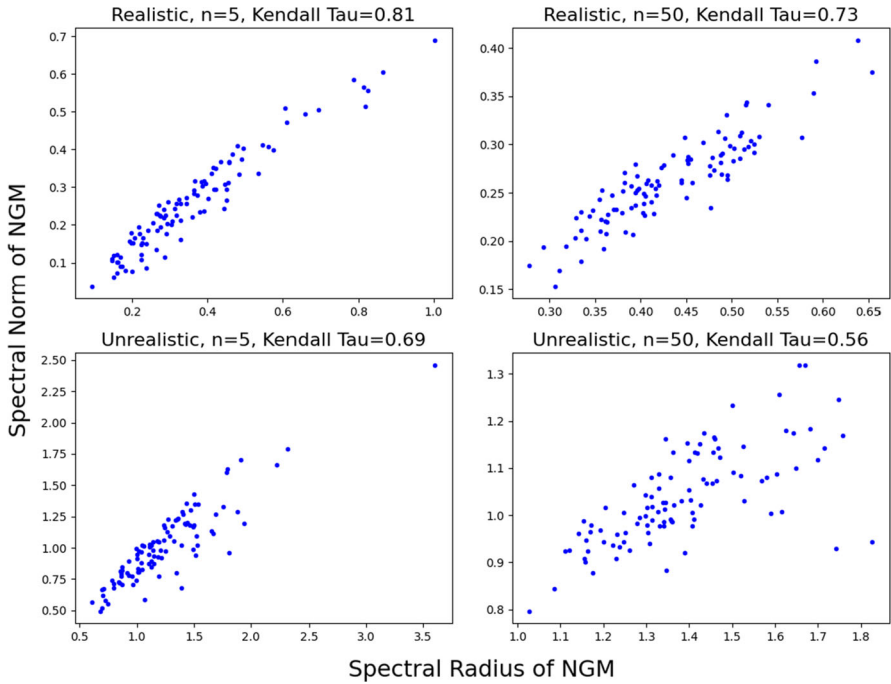
Fig. 2 Estimated Kendall Tau and joint sampling distribution for $M = 100$ samples of $\rho(\boldsymbol{G}_n)$ and $||\boldsymbol{G}_n||_2$. Subplot headers indicate underlying parameter distribution (realistic or unrealistic)

understood as a concentration of measure result (Tao 2012). An $n \times n$ matrix is a high-dimensional object, even for moderate n, and matrix norms are continuous in the matrix elements. As n grows, a large change in the norm (or radius) requires increasingly large numbers of elements to experience substantial, "coordinated" deviations. Such deviations are improbable for sufficiently large n, and so the bulk of the probability mass concentrates near the mean.

If such a result were to hold for $\boldsymbol{G}'$, then asymptotically we would see $\text{Var}[r(\boldsymbol{G}_n)] \sim \mathcal{O}(n^{-1})$. However these results typically hold only for specific classes of random matrices, such as Hermitian matrices with independent, mean-zero random entries (Tao 2012) or sums of fixed matrices with random coefficients (Anderson et al. 2010). The NGM in (2) is not Hermitian, and it has dependence between its entries; random matrices with correlated elements can behave idiosyncratically (Schenker and Schulz-Baldes 2005). Our proof then, amounts to showing that these shortcomings are not sufficient enough to overwhelm the concentration of measure (i.e. the variance reduction effect) entirely, although they do reduce its severity.

3.1 Outline of proof

We begin by properly defining G'_n:

$$G'_n = \frac{1}{\bar{K}} \begin{bmatrix} K_1 R_1 & \cdots & \frac{c_{1n} K_1}{2} \left(R_1 + \frac{\alpha_n}{\alpha_1} R_n \right) \\ \vdots & \ddots & \vdots \\ \frac{c_{n1} K_n}{2} \left(R_n + \frac{\alpha_1}{\alpha_n} R_1 \right) & \cdots & K_n R_n \end{bmatrix}. \tag{3}$$

Where we have defined $\bar{K} = \frac{1}{n} \sum_{i=1}^{n} K_i$, which is nearly constant for large n. Observe that G'_n is effectively the "density-dependent version" of 2 and therefore the effect of switching from density to frequency dependence is to induce the scaling $G_n = \frac{1}{n} G'_n$, and subsequently $\text{Var}[\rho(G_n)] = \frac{1}{n^2} \text{Var}[\rho(G'_n)]$. This $\frac{1}{n}$ scaling is what produces the variance reduction effect; we therefore expect that a variance reduction effect will occur only in the frequency-dependent case. Absent this scaling, i.e. when disease is density dependent, a variance amplification should occur.

For convenience we will adopt the notation $r(R, \alpha, K) = ||G'_n(R, \alpha, K)||_2$. Next we observe that, by the Law of Total Variance:

$$\text{Var}[r(R, \alpha, K)] = E_{\alpha, K} \left[\text{Var}_R[r(R, \alpha, K)|\alpha, K] \right] + \text{Var}_{\alpha, K} \left[E_R[r(R, \alpha, K)|\alpha, K] \right] \tag{4}$$

Our proof will show that the first term on the right hand side of Eq. 4 grows no faster than $\mathcal{O}(n^{3/2})$, while the second term is bounded by $\mathcal{O}(n^2)$. The desired result then follows from the $\frac{1}{n^2}$ scaling due to frequency dependence.

3.2 Bounding $E_{\alpha, K} [\text{Var}_R[r(R, \alpha, K)|\alpha, K]]$

Our result follows from the following theorem:

Theorem 2 (Concentration Inequality for Lipschitz Functions) *Let $X_1, \ldots, X_n$ be independent, real-valued random variables with $|X_i| \leq A$, for all $1 \leq i \leq n$. Let $F : \mathbf{R}^n \to \mathbf{R}$ be a convex, Lipschitz function with Lipschitz constant L. Then for any t one has that:*

$$P[(F(X) - E[F(X)])^2 \geq t] \leq \exp \left(\frac{t}{2A^2 L^2} \right)$$

This result is a well-known consequence of Talagrand for 1-Lipschitz functions (Boucheron et al. 2013), that has here been scaled to accommodate L-Lipschitz functions. We observe that, since $G(R + R', \alpha, K) = G(R, \alpha, K) + G(R', \alpha, K)$, then $r(R, \alpha, K)$ satisfies the norm axioms over R. Since norms are equivalent over finite dimensions this immediately implies that $r(R - R', \alpha, K) \leq L_{\alpha, K} ||R - R'||_2$, i.e. that $r(R, \alpha, K)$ is Lipschitz in R with respect to $||R||_2$. Simple integration of the right hand side of Theorem 2, together with $R_i \leq \Gamma$, gives the following:

$$\text{Var}_R[r(R, \alpha, K)|\alpha, K] \leq 2\Gamma^2 L^2_{\alpha, K}$$

A straightforward calculation shows that $L^2_{\boldsymbol{\alpha}, \boldsymbol{K}}$ is bounded by $C \sim \mathcal{O}(n^{3/2})$ (see Appendix 5.1), where C is independent of $\boldsymbol{\alpha}$ and $\boldsymbol{K}$. We can therefore bound $E_{\boldsymbol{\alpha}, \boldsymbol{K}}[\mathrm{Var}_{\boldsymbol{R}}[r(\boldsymbol{R}, \boldsymbol{\alpha}, \boldsymbol{K})|\boldsymbol{\alpha}, \boldsymbol{K}]]$ by the same term, which concludes the proof.

3.3 Bounding $\mathrm{Var}[E[r(R, \alpha, K)|\alpha, K]]$

Here we will use the fact that $r(\boldsymbol{R}, \boldsymbol{\alpha}, \boldsymbol{K}) \leq ||\boldsymbol{G}'(\boldsymbol{R}, \boldsymbol{\alpha}, \boldsymbol{K})||_F$, where $||\boldsymbol{M}||_F = \sqrt{\sum_{i,j} M^2_{ij}}$ is the Frobenius matrix norm. Therefore $Var[E[r(\boldsymbol{R}, \boldsymbol{\alpha}, \boldsymbol{K})|\boldsymbol{\alpha}, \boldsymbol{K}]] \leq E[E[r(\boldsymbol{R}, \boldsymbol{\alpha}, \boldsymbol{K})|\boldsymbol{\alpha}, \boldsymbol{K}]^2] \leq E[||\boldsymbol{G}'(\boldsymbol{R}, \boldsymbol{\alpha}, \boldsymbol{K})||^2_F]$. The last term on the right hand side is simply the expectation of a sum of bounded terms, and hence grows no faster than $\mathcal{O}(n^2)$. We therefore have that $Var[E[r(\boldsymbol{R}, \boldsymbol{\alpha}, \boldsymbol{K})|\boldsymbol{\alpha}, \boldsymbol{K}]] \sim \mathcal{O}(n^2)$, as desired.

3.4 Numerical experiments

To test the extent to which the above derived bounds are indeed useful for understanding of the actual behavior of $\mathrm{Var}[||\boldsymbol{G}(\boldsymbol{R}, \boldsymbol{\alpha}, \boldsymbol{K})||_2]$, we performed some simple Monte Carlo simulations of random realizations of the spectral norm of our NGM (2) with realistic distributions of parameter values for the case of frequency-dependent transmission, following Mihaljevic et al. (2014).

We assumed that the joint distribution of K_i, R_i, and α_i was such that R_i and α_i were conditionally independent, given K_i. K_i was then sampled from a log-normal distribution, in accordance with Preston's law of abundance distributions (Verberk 2011). Given K_i we then sampled α_i and R_i from truncated normal distributions (with support over [.1, 5] and [0, 10], respectively), conditioned on the values $k_i = \ln(K_i)$. The mode of the α_i was proportional to $\exp(k_i - 2)$, while for the R_i it is simply proportional to k_i. The constant of proportionality between α_i and $\exp(k_i - 2)$ is well established, following relationships between species abundance and life span. In addition, we allow for correlations between species susceptibility R_i and log-log-abundance $\ln(k_i)$, which we denote a_R. Although we did not explicitly encode "pace-of-life" traits into this model, the parameter a_R allows us to implictly consider their impact on variance reduction as fast-lived species often have high equilibrium abundances (Dobson 2004)

Using these sampling distributions we estimated separately each term in Eq. (4): the expectation of the conditional variance ($E_{\boldsymbol{\alpha}, \boldsymbol{K}}[\mathrm{Var}_{\boldsymbol{R}}[r(\boldsymbol{R}, \boldsymbol{\alpha}, \boldsymbol{K})|\boldsymbol{\alpha}, \boldsymbol{K}]]$) and the variance of the conditional expectation ($\mathrm{Var}[E[r(\boldsymbol{R}, \boldsymbol{\alpha}, \boldsymbol{K})|\boldsymbol{\alpha}, \boldsymbol{K}]]$) for both frequency- and density-dependent NGMs. Furthermore, for each type of NGM we considered four test cases: both a low and medium susceptibility-abundance correlation case ($a_R = .5$ and $a_R = 2$), as well as a low and high between-species infection case using $c_{ij} = .05$, and $c_{ij} = 1$. These are the values used in Mihaljevic et al. (2014) and the upper limit for the c_{ij} discussed in Dobson (2004), respectively.

Finally, for the low a_R, low c_{ij} case, and for each transmission mode, we consider two cases where the susceptibility of added host species increases or decreases with n, to better characterize how variance reduction behaves when community growth is not entirely random (Fig. 5). In this case we set $a_R = 0$, and instead sampled R_i

with from a truncated normal distribution (with support on $[0, 2]$) with mean $E[R_i] \propto c_R n + 1$ where c_R was set to $\pm .05$ for the positively and negatively correlated cases (respectively). Figure 6 shows an example of the behavior of the total variance, as well as the coefficient of variation for the low c, low a_R case for both transmission modes.

For the frequency-dependent NGM we see a clear variance reduction effect in every case (Fig. 3). While the actual estimated values of the terms were substantially lower than the bounds given in Theorem 1 (which were not plotted for legibility), we do see that the predicted qualitative behavior holds. Notably, in every case the mean conditional variance is strictly decreasing for $n > 10$. Thus, under relatively low species diversity, one sees a variance amplification effect. Yet, in each case the variance of the conditional mean (which produced the constant term in Theorem 1) is smaller than the mean of the conditional variance. At higher diversity levels ($n > 10$), the sum of these two terms, the total variance of the community R_0, will therefore decrease rapidly in n: a variance reduction effect, as we can see in Fig. 6. As expected, we see that an opposite, variance amplification effect occurs for the density-dependent NGM (Fig. 4).

This occurs because the effect of switching from frequency to density dependence is to multiply the NGM by a factor of n, which scales the variance by n^2. The concentration of measure is overwhelmed by this scaling, so the variance increases with n. We see a similar effect in Fig. 5 when susceptibility and richness are positively correlated. Since the positive correlation case assumes that $R_i \sim \mathcal{O}(n)$, the reason for variance amplification is the same as for Fig. 4. Note, however, that in the density dependent, negative correlation case we see that the variance amplification tapers for large n. This is because R_i is bounded below by 0, and so as n grows that R_i become small with high probability, counteracting the effect of density-dependence on the elements of the NGM.

Finally, in both Figs. 3 and 4 we see that the qualitative scaling predicted by our bounds appears to hold for large n, even when the actual numerical value of our bound is much larger than the observed variance values. Importantly, however, in the frequency dependent case we see that the mean conditional variance decays slower than our bound would predict, and that for large c the mis-fit becomes worse. This is predictable from the viewpoint of random matrix theory, as the off-diagonal terms of the NGM G_n exhibit correlation with each other. Were all matrix elements independent, variance reduction of both the community invasion factor and the community R_0 would be guaranteed by the Wigner Semicircular law (Tao and Vu 2014). The parameter c has the effect of reducing the impact of these correlated parameters, so when c is very small the matrix is more well-behaved, and the scaling behavior appears at smaller n.

4 Discussion

Our results have shown that, for frequency-dependent diseases, there exists terms $C_1 \sim \mathcal{O}(n^{-1/2})$ and $C_2 \sim \mathcal{O}(1)$ such that $\text{Var}[r(\mathbf{R}, \boldsymbol{\alpha}, \mathbf{K})] \leq C_1 + C_2$. It is challenging to make claims about the relative size of C_1 and C_2 for a general randomized NGM, and in cases where $\mathbf{R}$ is tightly determined by $\boldsymbol{\alpha}$ and $\mathbf{K}$, C_2 may dominate the bound, so variance reduction may not appear. However, in numerical experiments conducted
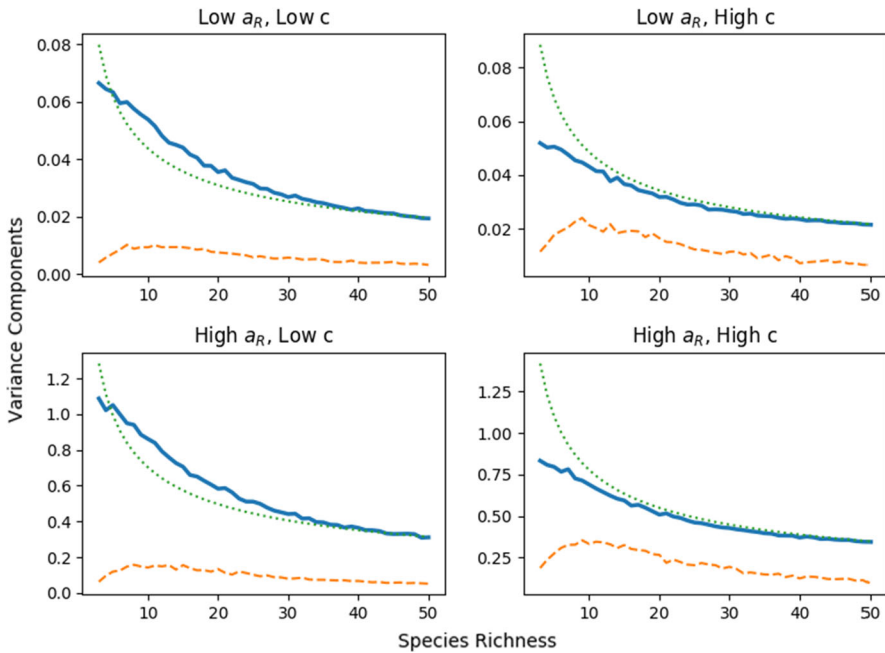
Fig. 3 Monte Carlo estimates of the mean conditional variance (blue, solid line) and the variance of the conditional mean (yellow, dashed line), the first and second terms (respectively) in Eq. (4), under the assumptions of the **frequency-dependent** NGM. In addition we plot the scaling behavior of the mean conditional variance predicted by its bound (green, dotted line). We see that in all cases a variance reduction effect occurs in the mean conditional variance, as is predicted by Theorem 1. Furthermore, in all cases we see that the mean conditional variance dominates the variance of the conditional mean, which explains why variance reduction is so pronounced in the sum of the two terms (color figure online)

with realistic distributions of R, α, and K we observed that for values of n less than 50, C_1 dominates C_2, and so variance reduction occurs. On the other hand, when disease is density-dependent our upper bound increases as $\mathcal{O}(n^{3/2})$ accordingly. In this case it is possible for either variance reduction or amplification to occur without violating our bound, although our numerical experiments suggests that variance amplification occurs for realistic biological parameter values.

These results can be contrasted with the explanation of variance reduction found in Mitchell et al. (2002), which observed a variance reduction in their experimental data, and attributed it to decreasing variability in community composition, i.e. a selection effect. Since the variance in community composition would presumably behave the same under either frequency or density dependent transmission, our results suggest that this explanation is incomplete. Rather, disease transmission mode mediates the relationship between community composition. In the case of frequency dependent transmission and high total community abundance, the introduction of a highly-competent host species will have a similar effect to the addition of a less competent host species, since the pathogen contact rate experienced by either will be fairly low. On the other hand, under density dependence the behavior of an epidemic will be more sensitive to community composition, and so a wider range of epidemic outcomes will
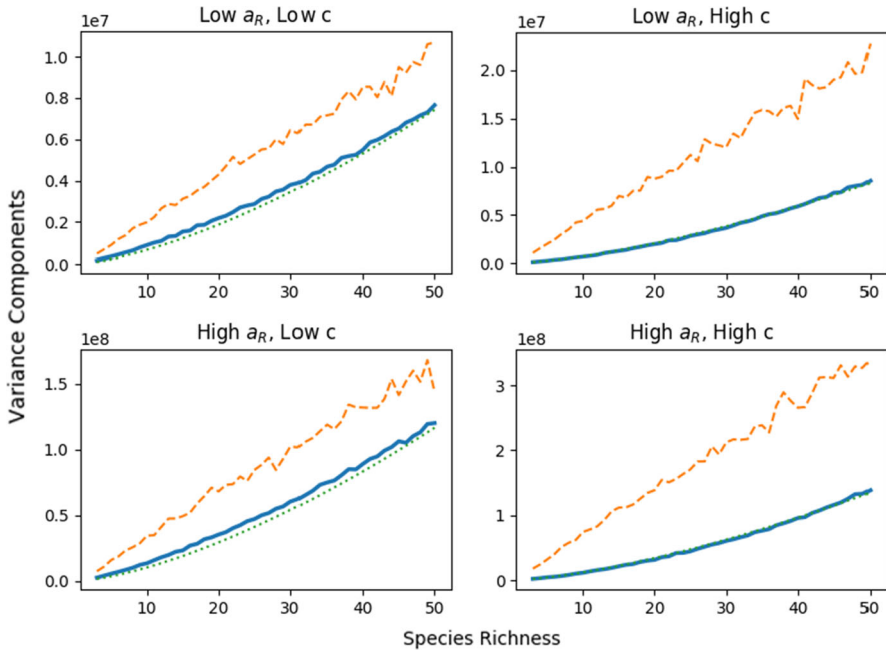
Fig. 4 Monte Carlo estimates of the mean conditional variance (blue, solid line) and the variance of the conditional mean (yellow, dashed line), the first and second terms (respectively) in Eq. (4), under the assumptions of the **density-dependent** NGM, as well as the predicted behavior of the mean conditional variance (green, dotted line, partially covered by the conditional variance in some plots). In this case variance reduction is overwhelmed by the $\mathcal{O}(n)$ scaling of the NGM due to density-dependence, and a variance amplification results (color figure online)

be likely. The outcome will in part be determined by the probability of including one or more highly-competent hosts.

Mathematically this difference manifests in the presence or absence of an $\mathcal{O}(n^{-1})$ scaling factor in the NGM. The core of our theorem demonstrates that, in the density dependent case the matrix norm of our NGM concentrates in such a way that variance goes as $\mathcal{O}(n^{3/2})$. Switching from density to frequency dependence then introduces the $\mathcal{O}(n^{-1})$, scaling the variance of the norm of the NGM such that it goes as $\mathcal{O}(n^{-1/2})$. This also suggests that any models where the NGM has elements which scale as $\mathcal{O}(n^{-1})$ (and where the elements are suitably independent) will lead to variance reduction. In Fig. 5 we see that scaling the R_i with n is sufficient to produce variance amplification, even when transmission is frequency-dependent, and that decreasing the R_i with n can counteract variance amplification when transmission is density-dependent. Furthermore, Mihaljevic et al. (2014) consider an identical model to ours but in addition distinguish between "compensatory" population growth, where $\sum_i K_i$ is held constant (and so $K_i \sim \mathcal{O}(n^{-1})$), and "additive" growth, where it is not. In numerical simulations they found that whenever population scaling was compensatory, variance reduction occurred, even if transmission was density dependent.

There are a number of opportunities to improve or extend our results. A simple improvement may be possible through a more careful calculation of the bound of
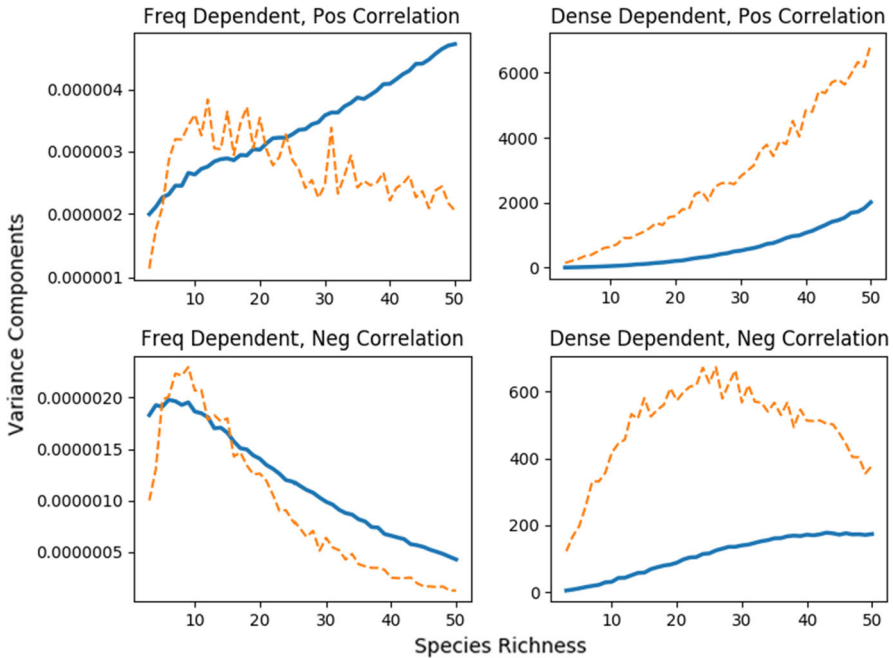
Fig. 5 Monte Carlo estimates of the mean conditional variance (blue, solid line) and the variance of the conditional mean (yellow, dashed line), the first and second terms (respectively) in Eq. (4), for the case where susceptibility exhibits positive or negative correlation with increases in species richness (color figure online)

$L^2_{\alpha,K}$ (see Appendix). Other improvements are possible by considering a less general model of between-species infections. In the case of vector borne diseases, for example, a number of the c_{ij} will go to zero, leaving only a handful of rows and columns nonzero which may allow for tighter control of the eigenvalues through the Gershgorin Circle Theorem. Alternately this could provide a route to derive a variance reduction effect for the community R_0. By setting enough c_{ij} to zero such that the underlying host community is broken into non-interacting "subcommunities" that do not grow with n to guarantee variance reduction in the spectral radius of G_n (Schenker and Schulz-Baldes 2005). The results of 2.2.3 suggest that variance reduction may be expected for community R_0 in a more general setting purely by correlation with the community invasion factor (and similarly for metrics of epidemic severity such as pathogen prevalence) (Fig. 6).

In addition to improving our bounds, they can be extended to cover several model variants. While the presented model ignores population dyanmics, a more realistic extension would be to assume density-dependence in birth or death rates, for example, by including a $(b_i - \delta N_i)S_i$ term in Eq. (1), where b_i and d_i denoted birth- and death-rates, respectively. However, as long as the population size is at equilibrium when the epidemic begins, introducing density dependence into birth rate is not expected to affect our results, since this would only modifies $\dot{S}$, and the NGM at the core of our theorem depends only on $\dot{I}$. On the other hand, if we instead used density-dependent
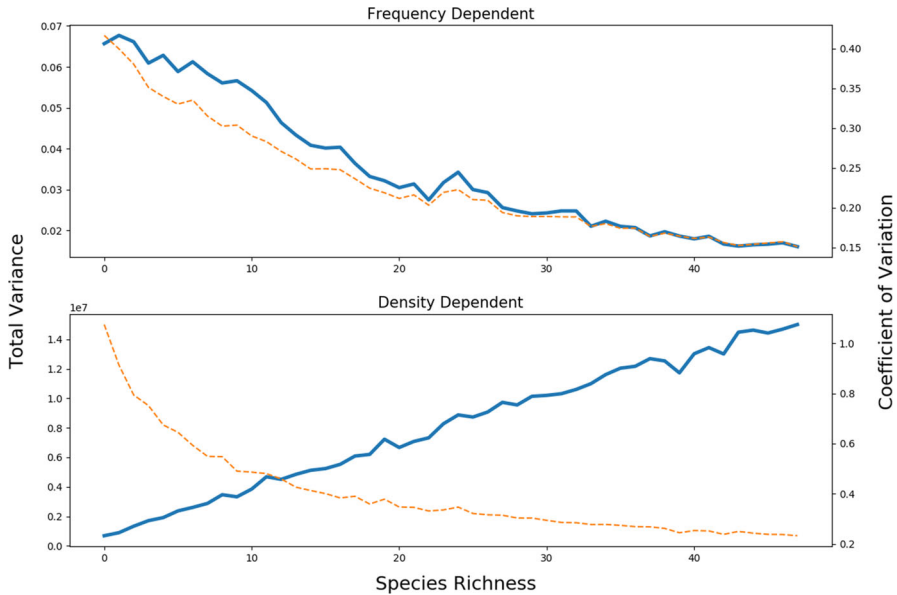
Fig. 6 Monte Carlo estimates of the total variance (blue, solid line) and the coefficient of variation (yellow, dashed line) of the community invasion factor, for frequency- and density-dependent transmission in the low a_R, low c case (color figure online)

death rates (instead adding terms $(d_i + \delta N_i)S_i$, $(d_i + \delta N_i)I_i$, and $(d_i + \delta N_i)D_i$) this would also change the "denominator" term of the NGM, resulting in correlation between the α_i and the K_i factors. Depending on the severity of this correlation it may reduce or reverse the variance reduction, although the numerical experiments conducted in Fig. 3 suggest that variance reduction will occur even with some modest correlation between these parameters.

It is also possible to consider model variants where our bounds do not hold. For example, if we reject our assumption that the triples (R_i, α_i, K_i) are independent for different i. While this assumption holds well for some experimentally designed host communities, it almost certainly does not hold for most natural ecosystems. Ecological relationships between host species such as mutualism or predation can easily create correlations between the K_i, which may have implications for the relationships between the α_i and R_i as well. While these correlations do not rule out a variance reduction effect necessarily, our bounds will not hold as proven and it would be difficult to make predictions about the ultimate behavior of the variance of community R_0.

While our model treats diversity as existing purely between homogeneous species, within-species heterogeneity due to genetic variation also can have substantial effects on epidemic dynamics (Elderd 2013; FlemingDavies et al. 2015; Dwyer et al. 1997). Given that the community R_0 is still definable in these cases (indeed its original definition was explicitly for the case of a single, heterogeneous host species), we expect that phenomena similar to variance reduction can also occur due to genetic diversity, and that our results could be adapted to understand this case.

Using a dynamical system model of epidemic spread we have provided a mathematical framework for understanding the variance reduction effect in epidemics with multiple host species. Our results suggest that variance reduction is driven more by pathogen transmission mode (e.g. density or frequency dependence), than by increased overlap between communities, i.e. a selection effect. This work also shows the utility of random matrix theory to disease ecology more generally. Where the behavior of an epidemic in a diverse ecological community can be modeled by a dynamical system, we expect that randomized linearizations of that model can be a rich source of insight into disease-diversity phenomena.

Compliance with ethical standards

5 Appendix

5.1 Bounding $L_{\alpha,K}^2$

To bound $L_{\alpha,K}^2$ we first observe that $\max_{||R||_2=1} r(R, \alpha, K) = L_{\alpha,K}$. We can therefore bound $L_{\alpha,K}^2$ by bounding $r(R, \alpha, K)^2$ for R on the unit ball. Since $r(R, \alpha, K)^2 \leq ||G'(R, \alpha, K)||_F^2$, where $||M||_F = \sqrt{\sum_{i,j} M_{ij}^2}$ is the Frobenius matrix norm, we can compute an upper bound in terms of a sum of the matrix elements. Furthermore, we will use the fact that $||R||_1 \leq \sqrt{n}||R||_2$, and define $\eta = \max(\{1, \frac{c^2}{4}\}) * \frac{A_2}{A_1}$. This definition of η is simple, and produces very loose bounds. Improvements may be made through a more careful consideration of this factor.

Now, from (3) and keeping in mind that $\sum_i R_i^2 = 1$, we have:

$$
\begin{aligned}
||G'(R, \alpha, K)||_F^2 &= \sum_i \left(R_i^2 + \sum_j \frac{c_{ij}^2}{4}\left(R_i + \frac{\alpha_j}{\alpha_i} R_j\right)^2 \right) \\
&\leq \eta \left(\sum_i \left(R_i^2 + \sum_j (R_i + \eta_1 R_j)^2 \right) \right) \\
&= \eta \left(\sum_i R_i^2 + \sum_{i,j\neq i} (R_i + R_j)^2 \right) \\
&= \eta \left(\sum_i R_i^2 + \sum_{i,j\neq i} (R_i^2 + 2R_i R_j + R_j^2) \right) \\
&\leq \eta \left(\sum_i R_i^2 + \sum_{i,j\neq i} (R_i^2 + 2R_j + R_j^2) \right)
\end{aligned}
$$

$$= \eta \left(\sum_i R_i^2 + (n-1)R_i^2 + 2(||R||_1 - R_i) + ||R||_2^2 - R_i^2 \right)$$

$$= \eta \left(||R||_2^2 + (n-1)||R||_2^2 + 2n||R||_1 - 2||R||_1 + n||R||_2^2 - ||R||_2^2 \right)$$

$$\leq \eta \left(1 + (n-1) + 2n\sqrt{n} - 2\sqrt{n} + n - 1 \right)$$

$$= \eta \left(2n\sqrt{n} + 2n - 2\sqrt{n} - 1 \right)$$
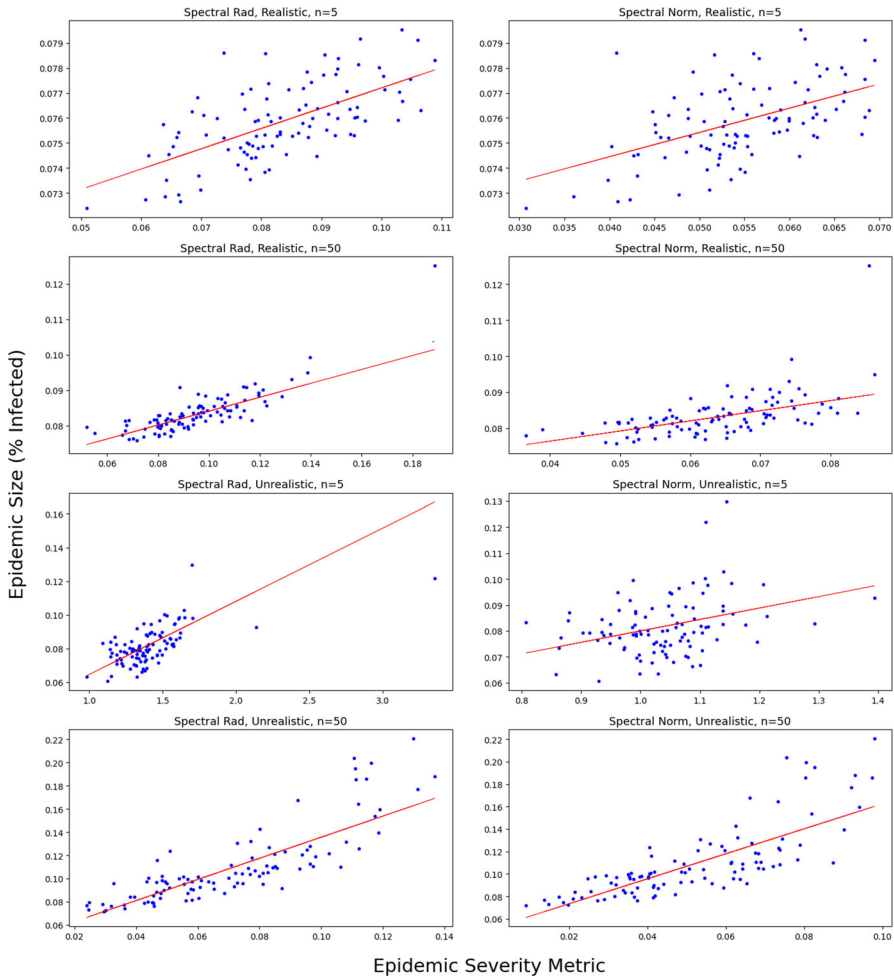
Which proves the claim.

Fig. 7 Relationships between epidemic size and $\rho(G_n)$ and $||G_n||$ with a model that includes infection-related deaths 5. In all cases $c = .05$ and $a_R = .5$, and we set $\sigma_i = .5d_i$ and $\alpha_i = 10d_i$. Epidemic size was defined as total number of individuals infected over the course of the epidemic ($\sum_i \int I_i(t)dt$) as a percentage of community abundance ($\sum_i N_i$)

5.2 Infection-related deaths

In the simulations plotted in Fig. 1 we had assumed that no births or deaths occurred over the course of the epidemic. In particular, this excluded the possibility of infection-related deaths, a feature present in most models of multi-host infections. Here we assess the effect of this modeling choice on the results of our numerical simulations.

We first modify Eq. (1) to include both a Malthusian birth term ($b_i(S_i + I_i + D_i)$), a base death term (e.g. $d_i S_i$) as well as an infection related deaths term, $(d_i + \alpha_i + \sigma_i)I_i$: Our multi-host SIR epidemic model is:

$$
\begin{aligned}
\dot{S}_i &= b_i(S_i + I_i + D_i) - p_i \sum_{j=1}^{n} \beta_{ij} I_j - d_i S_i \\
\dot{I}_i &= p_i \sum_{j=1}^{n} \beta_{ij} I_j - (d_i + \alpha_i + \sigma_i) I_i \\
\dot{D}_i &= \alpha_i I_i - d_i D_i
\end{aligned}
\tag{5}
$$

We assumed that $b_i = d_i$, that $\alpha_i = 10d_i$, and that $\sigma_i = .5d_i$. The results of this simulation are show in Fig. 7. We see that the primary effect of introducing σ_i is to reduce average infection prevalence when $n =$, and to reduce the variability of prevalence in all cases. This has the effect of reducing the relationship between prevalence and both $\rho(G_n)$ and $||G_n$. Again we see that both perform similarly in predicting infection size, although $||G_n||$ still performs slightly worse, particularly when $n = 50$.

References

Anderson GW, Guionnet A, Zeitouni O (2010) An Introduction to Random Matrices. Cambridge University Press, Cambridge

Anderson RM, May RM (1982) Directly transmitted infectious diseases: control by vaccination, 9p

Boucheron S, Lugosi G, Massart P (2013) Concentration Inequalities: a Nonasymptotic theory of independence. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199535255.001.0001; http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199535255.001.0001/acprof-9780199535255

Buhnerkempe MG, Roberts MG, Dobson AP, Heesterbeek H, Hudson PJ, Lloyd-Smith JO (2015) Eight challenges in modelling disease ecology in multi-host, multi-agent systems. Epidemics 10:26–30, https://doi.org/10.1016/j.epidem.2014.10.001; http://www.sciencedirect.com/science/article/pii/S1755436514000693

Civitello DJ, Cohen J, Fatima H, Halstead NT, Liriano J, McMahon TA, Ortega CN, Sauer EL, Sehgal T, Young S, Rohr JR (2015) Biodiversity inhibits parasites: broad evidence for the dilution effect. Procof Natl Acad Sci 112(28):8667–8671. https://doi.org/10.1073/pnas.1506279112; http://www.pnas.org/content/112/28/8667

Curl EA (1963) Control of plant diseases by crop rotation. Bot Rev 29(4):413–479. https://doi.org/10.1007/BF02860813

Diekmann O, JaP Heesterbeek, JaJ Metz (1990) On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. J Math Biol 28(4):365–382. https://doi.org/10.1007/BF00178324

Diekmann O, Heesterbeek JaP, Roberts MG (2009) The construction of next-generation matrices for compartmental epidemic models. J R Soc Interface p rsif20090386. https://doi.org/10.1098/rsif.2009.0386

Dobson A (2004) Population dynamics of pathogens with multiple host species. Am Nat 164(S5):S64–S78, http://www.journals.uchicago.edu/doi/abs/10.1086/424681

Dwyer G, Elkinton J, Buonaccorsi J (1997) Host heterogeneity in susceptibility and disease dynamics: tests of a mathematical model. Am Nat 150(6):685–707. https://doi.org/10.1086/286089

Elderd BD (2013) Developing models of disease transmission: insights from ecological studies of insects and their baculoviruses. PLoS Pathogens 9(6):e1003372. https://doi.org/10.1371/journal.ppat.1003372

Elderd BD, Dwyer G, Dukic V (2013) Population-level differences in disease transmission: a Bayesian analysis of multiple smallpox epidemics. Epidemics 5(3):146–156. https://doi.org/10.1016/j.epidem.2013.07.001; http://linkinghub.elsevier.com/retrieve/pii/S1755436513000315

FlemingDavies AE, Dukic V, Andreasen V, Dwyer G (2015) Effects of host heterogeneity on pathogen diversity and evolution. Ecol Lett 18(11):1252–1261. https://doi.org/10.1111/ele.12506

Huston MA (1997) Hidden treatments in ecological experiments: re-evaluating the ecosystem function of biodiversity. Oecologia 110(4):449–460. https://doi.org/10.1007/s004420050180

Johnson PTJ, Ostfeld RS, Keesing F (2015) Frontiers in research on biodiversity and disease. Ecol Lett 18(10):1119–1133. https://doi.org/10.1111/ele.12479

Keeling MJ, Rohani P (2011) Modeling infectious diseases in humans and animals. Princeton University Press, Princeton

Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM (2005) Superspreading and the effect of individual variation on disease emergence. Nature 438(7066):355–359. https://doi.org/10.1038/nature04153

Mihaljevic JR, Joseph MB, Orlofske SA, Paull SH (2014) The scaling of host density with richness affects the direction, shape, and detectability of diversity-disease relationships. PLoS ONE 9(5):e97812. https://doi.org/10.1371/journal.pone.0097812

Mitchell CE, Tilman D, Groth JV (2002) Effects of grassland plant species diversity, abundance, and composition on foliar fungal disease. Ecology 83(6):1713–1726. https://doi.org/10.1890/0012-9658(2002)083[1713:EOGPSD]2.0.CO;2

Ostfeld RS, Keesing F (2000) Biodiversity and disease risk: the case of lyme disease. Conserv Biol 14(3):722–728. https://doi.org/10.1046/j.1523-1739.2000.99014.x

Randolph SE, Dobson ADM (2012) Pangloss revisited: a critique of the dilution effect and the biodiversity-buffers-disease paradigm. Parasitology Camb 139(7):847–63. https://doi.org/10.1017/S0031182012000200; http://search.proquest.com.colorado.idm.oclc.org/docview/1017536663/abstract/BC0121ED6FC04C59PQ/1

Rottstock T, Joshi J, Kummer V, Fischer M (2014) Higher plant diversity promotes higher diversity of fungal pathogens, while it decreases pathogen infection per plant. Ecology 95(7):1907–1917. http://onlinelibrary.wiley.com/doi/10.1890/13-2317.1/full

Salkeld DJ, Padgett KA, Jones JH (2013) A meta-analysis suggesting that the relationship between biodiversity and risk of zoonotic pathogen transmission is idiosyncratic. Ecol Lett 16(5):679–686. https://doi.org/10.1111/ele.12101

Schenker JH, Schulz-Baldes H (2005) Semicircle law and freeness for random matrices with symmetries or correlations. arXiv preprint arXiv:math-ph/0505003

Tao T (2012) Topics in random matrix theory. American Mathematical Society. https://bookstore.ams.org/gsm-132/

Tao T, Vu V (2014) Random matrices have simple spectrum. arXiv:1412.1438 [math]

Vandermeer JH (1992) The ecology of intercropping. Cambridge University Press, Cambridge

Verberk W (2011) Explaining general patterns in species abundance and distributions. Nat Educ Knowl 3(10):38