Macroscopic patterns of interacting contagions are indistinguishable from social reinforcement
Supplementary Information

Laurent Hébert-Dufresne,[1,2,3] Samuel V. Scarpino,[4,5,6,7,8] and Jean-Gabriel Young[9]

[1]*Vermont Complex Systems Center, University of Vermont, Burlington, VT 05405, USA*
[2]*Department of Computer Science, University of Vermont, Burlington, VT 05405, USA*
[3]*Département de physique, de génie physique et d'optique, Université Laval, Québec (Québec), Canada G1V 0A6*
[4]*Network Science Institute, Northeastern University, Boston, MA 02115, USA*
[5]*Marine & Environmental Sciences, Northeastern University, Boston, MA 02115, USA*
[6]*Physics, Northeastern University, Boston, MA 02115, USA*
[7]*Health Sciences, Northeastern University, Boston, MA 02115, USA*
[8]*ISI Foundation, 10126 Turin, Italy*
[9]*Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109, USA*

CONTENTS

I. INFERRING THE COMPLEX CONTAGION FUNCTION

The result of large scale epidemiological surveillance efforts are typically reported as times series $Y_{1:T} := (Y_1, Y_2, ..., Y_T)$ of the number of infected or newly infected individuals in a monitored populations. One may think of these time series as coarse-grained observations of some detailed spreading process, taking place on hidden contact networks. We now obtain a principled inference procedure to determine the parameters of the spreading processes, from the coarse grained time series alone (assuming that the family of possible contact networks is known).

For the sake of concreteness, we will hereafter assume that the spreading process is a complex Susceptible-Infected-Recovered (SIR) dynamics, and model the population as well-mixed, but the principles derived apply more broadly. The complex SIR dynamics is defined as one where nodes in the infectious state recover at a constant rate $\gamma > 0$ (i.e. transition from I to R), but where susceptible individuals (S) get infected at a rate $\beta(I)$ that is a function of the density of infected individuals, in the neighbourhood of infected individuals. In a well-mixed population, these densities are governed by the mean-field system:

$$\frac{d}{dt}S(t) = -\beta\big(I(t)\big)I(t)S(t) \tag{1a}$$

$$\frac{d}{dt}I(t) = \beta\big(I(t)\big)I(t)S(t) - \gamma I(t) \tag{1b}$$

$$\frac{d}{dt}R(t) = \gamma I(t) \tag{1c}$$

There are no known closed form solutions, but it is fairly straightforward to generate time series by integrating from some initial condition (S_0, I_0, R_0).

As is argued in the main text, this variable infection rate $\beta(I)$ helps capture exogenous processes such as social reinforcement and disease interactions. Our inference goal will be to determine $(\gamma, \beta(I))$.

A. Inference under perfect conditions

If we get to observe the output of the mean-field dynamics at an infinite time-resolution, without any noise or misspecification whatsoever, the inverse problem of determining $\big(\gamma, \beta(I)\big)$ can be solved exactly, without much difficulties. We compute the empirical derivatives $\big(\hat{\dot{S}}(t), \hat{\dot{I}}(t), \hat{\dot{R}}(t)\big)$ of the time series, and invert Eqs. (1) to get the estimators:

$$\hat{\beta}(t) = \frac{-\hat{\dot{S}}(t)}{I(t)S(t)}\,, \tag{2a}$$

$$\hat{\gamma}(t) = \frac{\hat{\dot{R}}(t)}{I(t)}\,. \tag{2b}$$

Because the time series are directly generated by the dynamics appearing in Eq. (1), $\hat{\gamma}(t)$ does not actually depend on time, and will therefore be a constant. Any pair $(\hat{\dot{R}}(\tau), I(\tau))$ at some time τ is enough to evaluate the rate γ, although checking many pairs can in practice help average numerical errors out. In contrast, we expect $\hat{\beta}(t)$ to change with t, but only because density of infected individuals also varies with time. Making this dependency explicit yields

$$\hat{\beta}\big(I(t)\big) = \frac{-\hat{\dot{S}}\big(I(t)\big)}{I(t)\big(I(t)\big)}\,, \tag{3}$$

which gives a prescription for computing $\beta(I)$ exactly. Figure 1 confirms that the procedure works for various combinations of rates $\big(\gamma, \beta(I)\big)$.

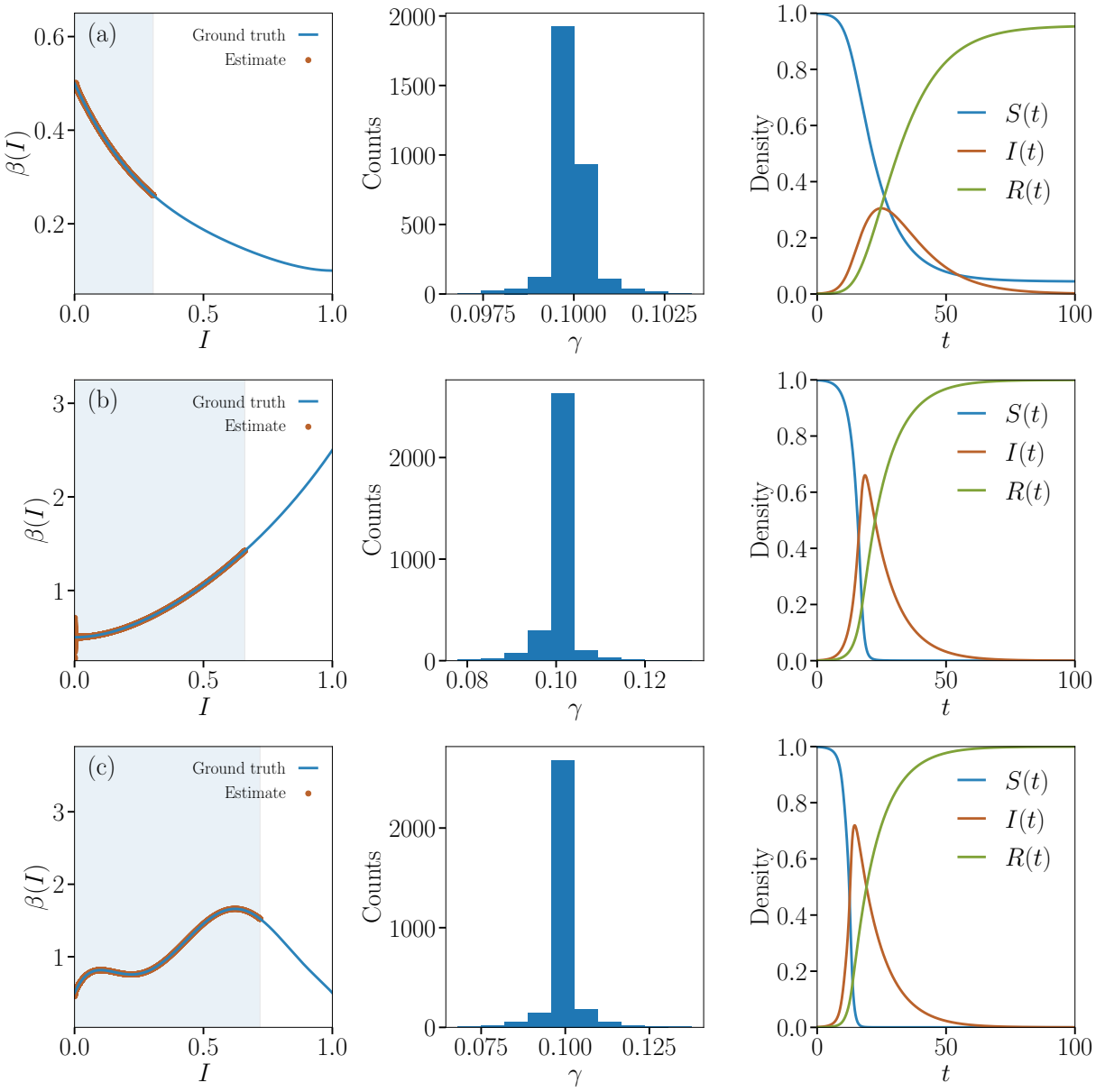

FIG. 1. **Inference under perfect conditions.** Inferred complex contagion function $\beta(I)$ (left column) and distribution of inferred recovery rate (central column) computed from time series generated by Eqs. (1), starting from the initial condition $(S_0, I_0, R_0 = (1-\varepsilon, \varepsilon, 0)$ with $\varepsilon = 0.001$ with $\gamma = 0.1$ and various functions $\beta(I)$. Note that it is only possible to infer the function $\beta(I)$ for observed values of I, indicated by a shaded region in the left column. The tested functions are *(a)* a decelerating contagion, *(b)* an accelerating contagion, and *(c)* some complicated contagion function with valleys and peaks.

B. Inference under noisy conditions

Real epidemiological time series are noisier, less time-resolved, and often not as high dimensional[1] than the ones considered in Sec. I A. As a result, applying the inversion procedure of Eqs. (2) blindly to these series would seriously overfit the data. A more principled approach is needed to handle real datasets.

[1] We do not typically have—even noisy—measurements of $S(t)$ and $R(t)$.

1. Bayesian inference: Likelihood

We model the observed series as an imperfect and partial measurement of the macroscopic state of a detailed spreading process. That is, we assume that the detailed configuration $X_t \in \mathcal{X}$ of susceptible, infected and recovered individuals at time t is measured only through a macroscopic quantity. We will consider two macroscopic measurements: (1) the density of infected individuals or prevalence, a real number in $[0, 1]$ which we will denote Y_t to emphasize that the measurement is noisy and different from $I(t)$; and (2) the number of newly infected individuals or incidence, an integer in $[0, n]$ where n is the population size. We will denote the incidence between time $t-1$ and t as Z_t. The well-mixed assumption is translated by a population structure drawn from the Erdős-Rényi-Gilbert model of density $\rho \in [0, 1]$.

Model for prevalence measurements. The likelihood of the time series $Y_{1:T}$ is given by

$$P(Y_{1:T}|\beta, \sigma, \gamma, \rho, x_0) = \sum_G \int_{\mathcal{X}^T} P(Y_{1:T}|x_{1:T}, \sigma) P(x_{1:T}|G, \beta, \gamma, x_0) P(G|\rho) dx_{1:T} \ . \tag{4}$$

where the integral is over all sequences of hidden states starting at $x_0 \in \mathcal{X}$, where $P(G|\rho)$ is a distribution over possible population structures parametrized by the density ρ, and where

$$P(Y_{1:T}|x_{1:T}, \sigma) = \prod_t q(Y_t|I[x_t], \sigma) \tag{5}$$

is the noise component of the model, with $I[\cdot]$ the density of infected individuals in configuration $x_t \in \mathcal{X}$, and $q(Y|\mu, \sigma)$ the p.d.f. of the normal distribution of mean μ and standard deviation σ.

If we could compute Eq. (4) exactly in closed form, then we could compute the posterior distribution easily, and make inference on the parameters of the model. But both the sum over graphs and the integral over hidden states turn out to be complicated in all but the simplest cases (e.g., when the dynamics occurs on the fully connected population $\rho = 1$, with no noise). We must therefore resort to approximations. Noting that only the density of infected matters in Eq. (5), we approximate the integrals over states by their largest contribution, the mean-field solutions of Eqs. (1). This leads to

$$P(Y_{1:T}|\beta, \gamma, \rho, \sigma, y_0) = \prod_{t \in \tau_{\text{obs}}} q(Y_t|\tilde{y}_t(\beta, \gamma, \rho; y_0), \sigma) \ , \tag{6}$$

where τ_{obs} is the set of time steps for which we have an observation, and where $\tilde{y}_{1:T}(\beta, \gamma, \rho; y_0)$ is a time series of infectious density, given as the solution of Eq. (1) integrated from the initial condition $y_0 = (S_0, I_0, R_0)$.

Model for incidence measurements. In most real epidemics, we do not get to measure the density of infected. Instead, we often measure an incidence rate, e.g., the number new cases in a week. It turns out that we can carry out the inference highlighted above, with little modifications. We denote the time series of average incidence rate as $Z_{1:T}$. In the mean-field system of Eqs. (1) the number of new infected between times t and $t' > t$ is $n[S(t) - S(t')]$ where n is the population (assumed constant). Let us denote by $\tilde{z}_{1:T}(\beta, \gamma, \rho, n; y_0)$ the time series of these differences, i.e.:

$$\tilde{z}_t(\beta, \gamma, \rho, n; y_0) = n\Big[\tilde{y}_t(\beta, \gamma, \rho, y_0) - \tilde{y}_{t+1}(\beta, \gamma, \rho; y_0)\Big] \ , \tag{7}$$

The likelihood of the time series $Z_{1:T}$ under a normal noise is then:

$$P(Z_{1:T}|\beta, \gamma, \rho, \sigma, y_0, n) = \prod_{t \in \tau_{\text{obs}}} q(Z_t|\tilde{z}_t(\beta, \gamma, \rho, n; y_0), \sigma) \ , \tag{8}$$

where q is, again, the p.d.f. of the normal distribution.

2. Bayesian inference: Posterior distribution

For both time series, we can obtain a posterior distribution over parameters by adding priors on the parameters of the likelihood. For example, for $Y_{1:T}$, we get:

$$P(\beta, \gamma, \sigma, y_0|Y_{1:T}) = \frac{P(Y_{1:T}|\beta, \gamma, \sigma, y_0) P(\beta, \gamma, \sigma, y_0)}{P(Y_{1:T})} \ , \tag{9}$$

and a similar equation holds for $Z_{1:T}$. Note that we have removed the dependency on ρ since the density of a well-mixed population only influences the relative timescale of events as far as the SIR dynamics is concerned—we can simply *impose* a timescale and ignore ρ.

3. Bayesian inference: Priors

We opt for simple independent priors. For the scale of the noise, we choose a half-Cauchy prior on $\sigma > 0$ (centered on 0). For the initial conditions of the mean-field equations, we place a uniform prior on $I_0 \in [0, I_{\max}]$ where $I_{\max}$ is a preset upper bound[2], and a uniform prior on S_0 that preserves the normalization $S_0 + I_0 + R_0 = 1$, namely $S_0|I_0 \sim \mathrm{Unif}(0, I_0)$. For γ, we place a very weak truncated normal prior centered at 0 (where $P(\gamma) = 0$ for $\gamma < 0$). Note that since incidence is essentially a discretized derivative of $nS(t)$, the data does not fully identify γ (see Eq. (1)). Too tight or too diffuse of a prior is likely to misguide inference in that case.

Placing a prior on β requires that we parametrize the function first. In principle, complex contagion is defined for *any* function $\beta : [0, 1] \to [0, \infty)$. Hence, we would ideally like to consider as broad a family of functions β as possible a priori. But we also need to be careful not to overfit the data with overly flexible functions. We find that low degree Bernstein polynomials are perfectly suited to our purpose.

A Bernstein polynomial $B_N(I; \boldsymbol{\xi})$ of degree N is a linear combination of the basis polynomials

$$b_{\nu,N}(I) = \binom{N}{\nu} I^\nu (1 - I)^{N-\nu}, \tag{10}$$

with coefficients ξ_ν,

$$B_N(I; \boldsymbol{\xi}) = \sum_{\nu=0}^{N} \xi_\nu b_{\nu,N}(I), \tag{11}$$

that maps $[0, 1]$ to $(-\infty, \infty)$ in general. If $\xi_\nu > 0 \; \forall \; \nu$ however, $B_N(I; \boldsymbol{\xi})$ is non-negative on $[0, 1]$. This makes the Bernstein polynomials of some fixed degree N with non-negative coefficients an *almost* perfect parametrization of β, as $\beta(I) = B_N(I; \boldsymbol{\xi})$.

The only "problem" with the unaltered Bernstein polynomials is that they are defined on the domain $[0, 1]$, whereas the observed densities $I(t)$ usually stay close to 0, never reaching 1. The net consequence is that we do not get to observe enough data to determine $\beta(I)$ on its whole domain. In theory, this should be reflected by an uninformative posterior distribution past $\max Y_{1:T}$; but in practice, sampling / optimization procedures suffer from such an ill-determined posterior. Thus we instead parametrize $\beta(I)$ with *rescaled* Bernstein polynomials, by mapping the domain of the standard Bernstein polynomials (the unit interval) to $[-\alpha I_{\max}, I_{\max}(1 + \alpha)]$, where α is an overshoot parameter that extends the boundary of the function a little bit out of the allowed region for I_0[3].

We find that in practice, low degree polynomials strikes a good balance between expressiveness, computational complexity, and regularization. We choose $N = O(1)$ in all our experiments because the model gets harder to sample as N increases, due to the appearance of local minima. To capture the fact that we expect simple contagion a priori, we parametrize the coefficient of β as $\xi_i = \mu_\xi + \Delta_i$ where μ_ξ is a baseline infection rate and Δ_i is a deviation. If all the $\{\Delta_i\}$ are close to 0, then the contagion rate is basically flat. Conversely, large $\{\Delta_i\}$ lead to a non-constant contagion function.

We use a very weak half-normal prior centered at 0 for μ_ξ, and a Cauchy prior of scale σ_Δ for all the deviations Δ_i.

4. Data normalization

To simplify the choice of priors and streamline inference, we normalize every time series before we start making our analyses. Specifically, we rescale the time interval to $\tau \in [0, 1]$. The inferred values can be transformed back to the original timescale via

---

[2] We set $I_{\max} = \max Y_{1:T}$ when we have prevalence data, and use an arbitrary upper bound when we have incidence data, see Sec. I C below for details on the latter case. Note that because the prior probability is uniform on $[0, I_{\max}]$ and zero everywhere else, the posterior probability associated to values of I_0 outside of this range will also be 0. Therefore, with the incidence data, we verify that the initial condition of the latent time series $\tilde{y}$ never "saturates"'and reaches values close to the chosen $I_{\max}$ a posteriori.

[3] The same considerations as for I_0 apply. We verify that the inferred time series never "saturate" a posteriori

the transformation $\tau = t/t_{\max}$, where $t_{\max}$ is the original running time of the time series. This is due to the fact that the ODEs

$$\frac{d}{d\tau}S = -\beta(I)IS, \qquad \frac{d}{d\tau}I = \beta(I)IS - \gamma I,$$

and

$$\frac{d}{dt}S = -t_{\max}\beta(I)IS, \qquad \frac{d}{dt}I = t_{\max}\beta(I)IS - t_{\max}\gamma I$$

have the same solutions. We present the results in their natural timescales for the sake of easy interpretability.

5. *Model summary*

The full model for prevalence data is $Y_{1:T}$

$$\begin{aligned}
Y_t &\sim N(\tilde{y}_t, \sigma^2) && t = 1, ..., T \\
\sigma^2 &\sim \text{Half-Cauchy}(0, \sigma_n) \\
\tilde{y}_{1:t} &= \text{SIR}(\beta(\boldsymbol{\xi}_{1..N}, \alpha), y_0, \gamma) \\
I_0 &\sim \text{Unif}(0, I_{\max}) \\
S_0|I_0 &\sim \text{Unif}(0, I_0) \\
\xi_i &= \mu_\xi + \Delta_i \\
\gamma &\sim \text{Half-Normal}(0, \sigma_\gamma^2) \\
\mu_\xi &\sim \text{Half-Normal}(0, \sigma_\mu^2) \\
\Delta_i &\sim \text{Cauchy}(0, \sigma_\Delta^2) && i = 0, ..., N
\end{aligned}$$

where $\text{SIR}(\beta(\boldsymbol{\xi}_{1..N}, \alpha), y_0, \gamma)$ returns a mean-field time series of prevalence with initial conditions $y_0 = (S_0, I_0, 1 - S_0 - I_0)$, recovery rate γ, and a contagion function β parametrized by the degree N Bernstein polynomials of coefficients $\boldsymbol{\xi}$. The full model for incidence data $Z_{1:T}$ is almost identical, with the differences:

$$\begin{aligned}
Z_t &\sim N(\tilde{z}_t, \sigma^2) && t = 1, ..., T \\
\tilde{z}_t &= n(\tilde{y}_t - \tilde{y}_{t+1}) \\
\gamma &\sim \text{Normal}(\mu_\gamma, \sigma_\gamma^2)
\end{aligned}$$

where n is a population size given a priori, μ_γ is used to inform the inference, and where γ is constrained to $\mathbb{R}^+$.

6. *Inference algorithm and diagnosis*

We approximate the posterior distribution with samples generated by an Hamiltonian Monte-Carlo (HMC) sampler. This process can be carried out automatically by implementing the model[4] in the `STAN` probabilistic programming language [1].

There are a number of post-sampling check that allows one to verify whether HMC samples are well behaved [1]. Of all the possible diagnostics, the presence of divergent transitions is the most important one: It indicates whether there are regions of the posterior that are hard to sample due to high curvature. When we use Bernstein polynomials of a high degree N, the posterior tends to become multimodal—even though one of the modes dominates the others. Divergent transition, can happen as a result of this peculiar geometry. To see why, consider a highly non-linear ground truth $\beta(I)$. There are at least two "explanations" of the data, and they are separated by steep chasms: A flat averaged out $\beta(I)$ with significant noise σ^2, and a posterior close to the ground-truth with next to no noise σ^2. We avoid the divergent transitions brought about by this geometry by (a) sticking to relatively low order polynomials, and (b) running long burn-in periods to make sure that the divergent transitions encountered

[4] Our implementation is available online.

on the way to the dominant mode do not pollute our inference. Note, however, that some unavoidable divergent transitions occur when the data determines the model extremely well—i.e., when the data is measured without noise—such that any description of the data but the best is a comparatively bad one. These can be safely ignored.

7. *Validation*

In Figs. 2–6, we investigate the performance of the inference method by generating artificial data with the model itself.

Figure 2 shows noisy time series, sampled more or less frequently. Even when most data points are absent, the posterior distribution correctly captures the shape of the function $\beta(I)$. Importantly, point estimates are not as accurate in recovering the contagion function as the full distribution; sampling is not only useful, and it is necessary.

Figure 3 shows noisy time series, under various level of noise σ^2. When there is no noise, even a sparse time series is enough to perfectly reconstruct the time series. Furthermore, reconstruction remains qualitatively possible under large level of noise.

Figures 4 and 5 show the effect of prior misspecification on inference with incidence data. Imprecise but vague prior tend to yield good results (Fig. 4). While precise but incorrect priors, lead to bad results (Fig. 5) This shows that incidence partially identifies γ, and suggest that we are better off using vague priors when dealing with real data.

Figure 6 shows the inference we would have made, had we kept only a (random) fraction of all the Monte Carlo samples. The figure tells us that only a few samples are really necessary.

C. Computational details

In all experiments (main text and SI alike), we set the scale of the prior on the noise to $\sigma_n = 1$ and the scale of the prior on the deviations $\{\Delta_i\}$ to $\sigma_\Delta = 1$. When we deal with prevalence data, we set the scale of the prior on γ to $\sigma_\gamma = 100$, and the overshoot parameter α on the range of β to $\alpha = 0.1$. When we deal with incidence data, we set a case specific prior for γ (see below) and use no overshoot ($\alpha = 0$) because the prevalence time series is latent in the first place.

The other parameters are varied from experiments to experiments[5].

1. *Supplementary Information figures*

In all SI figures investigating complex contagions on well-mixed populations (SI Figure 2–6), we use polynomials of degree $N = 8$, and a scale parameter $\sigma_\mu = 100$. When dealing with incidence data, we use a tight prior on γ with $\mu_\gamma = \tilde{\gamma}$ and $\sigma_\gamma = 0.1$ (where $\tilde{\gamma} = 0.1$ is the ground-truth), except in Figs. 4–5 where we intentionally give a bad prior. The population size is always set to its true value, $n = 100\,000$.

For most experiments shown in the Supplementary Information, we make our inference using 4 chains of 500 samples, all initialized with a burn-in period of 500 iterations. In figure 3, however, we use shorter chains when the data determines the posterior very well. Specifically, for an artificial noise level of $\sigma^2 = 0$, we only use 100 warm-up iterations and 100 samples per chains. For an artificial noise level of $\sigma^2 = 0.01$, we use 250 warm-up iterations and 250 samples per chains.

For the analysis of interacting contagions on real networks (Fig. 8 and 9), we use polynomials of degree $N = 6$, $\sigma_\gamma = \sigma_\mu = 100$ and $\sigma_\Delta = 1$.

2. *Figure 3*

In Figure 3 of the main text, we use polynomials of degree $N = 6$, and a scale parameter $\sigma_\mu = 100$. The population size is set to its true value, $n = 10\,000$. We make our inference using 4 chains of 1000 samples, all initialized with a burn-in period of 1000 iterations. The true recovery rate is given as a prior with noise $\sigma_\gamma = 1$ for the fit on incidence data.

[5] Note that all of the details below are also available in our replication files.

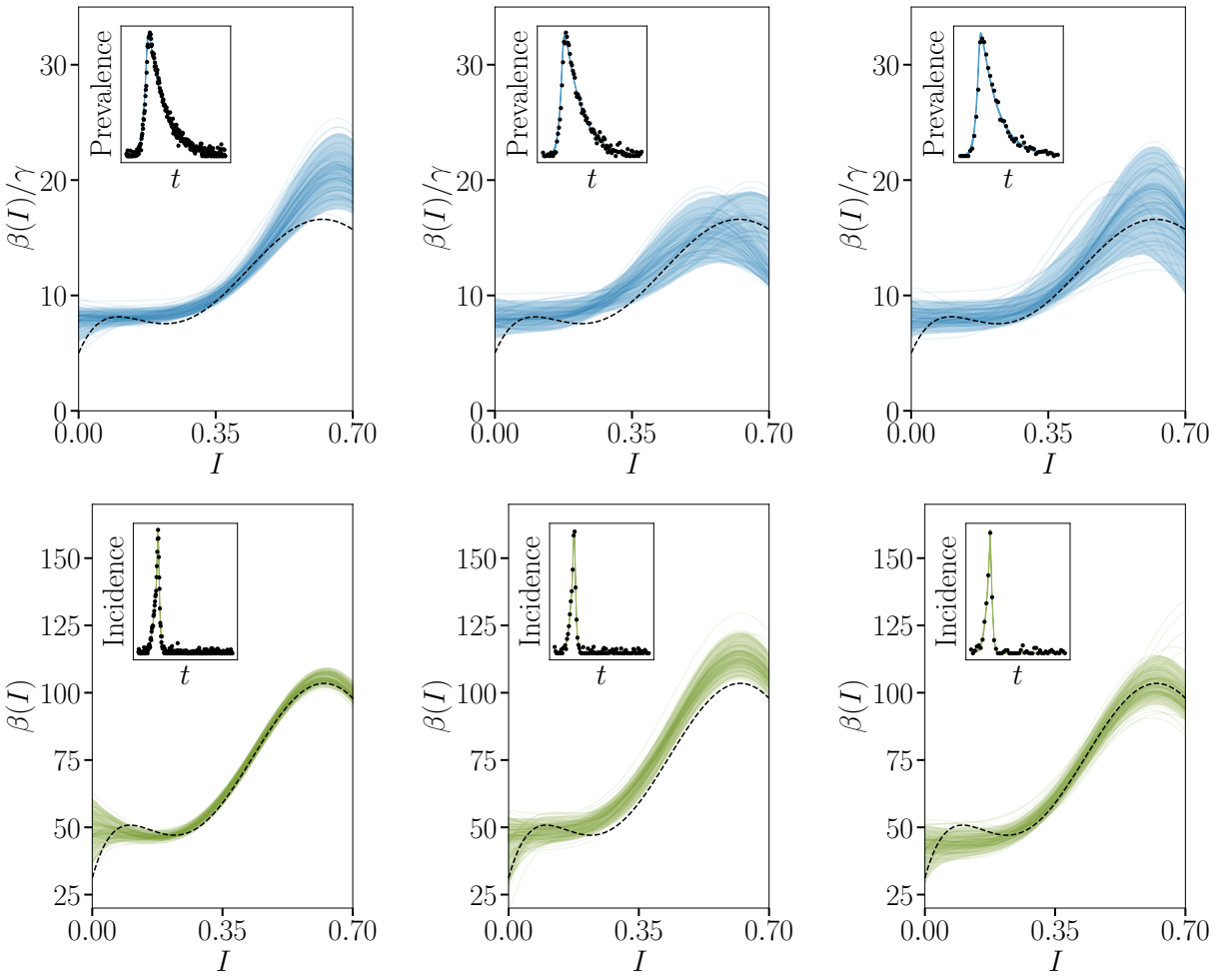

FIG. 2. **Inference at different levels of sampling.** (top row) Inference on prevalence data. (bottom row) Inference on incidence data. We use a population size of $n = 100\,000$. In all cases, the same base time series is used. It is generated with Eqs. (1) using $\gamma = 0.1$, $(S_0, I_0, R_0 = (1-\varepsilon, \varepsilon, 0)$ where $\varepsilon = 0.001$, and the complicated $\beta(I)$ function shown as the ground-truth in the main plot. We keep (from left to right): $T = 250$, $T = 100$ and $T = 50$ regularly spaced data points. An independent Gaussian noise of variance $\sigma^2 = 0.2$ is applied to the prevalence measurements, while the variance is set to $\sigma^2 = 2 \times n \times T$ for the incidence measurements. The resulting series are shown in the inset, alongside a posterior fit (95% of the a posteriori incidence curves). The main plots show the inference results for the three level of data sparsity. The ground truth is shown with a dotted line, alongside with 95% of the data (shaded region), and 100 random posterior samples to emphasize sample-to-sample variability.

3. *Figure 4*

In Figure 4 of the main text (real epidemics), we use polynomials of degree $N = 6$, and a scale parameter $\sigma_\mu = 1000$ (social) and $\sigma_\mu = 100$ (Dengue). The dispersion prios is set to $\sigma_\Delta = 10$ in both case. The population size is set to its true value, $n = 1\,000\,000$ (social contagion) and $n = 3\,761\,000$ (Dengue in Puerto Rico). We make our inference using 4 chains of 1000 samples, all initialized with a burn-in period of 1000 iterations. The posterior is multimodal so we keep two chains in each case (the ones that converge to the best optimum). The prior on the recovery rate is centered at 1 with scale $\sigma_\gamma = 10$ (social), and on 0 with scale $\sigma_\gamma = 1$ (dengue). For the social contagion, we use a scale $\sigma_n = 0.001$ on noise, and constrain $I(t)$ to 0.005 and $I(t = 0)$ to 0.0002. For Dengue, we use a scale $\sigma_n = 0.001$ on noise, and constrain $I(t)$ to 0.001 and $I(t = 0)$ to 0.00001. All these values are chosen to obtain a reliable fit of the data a posteriori.

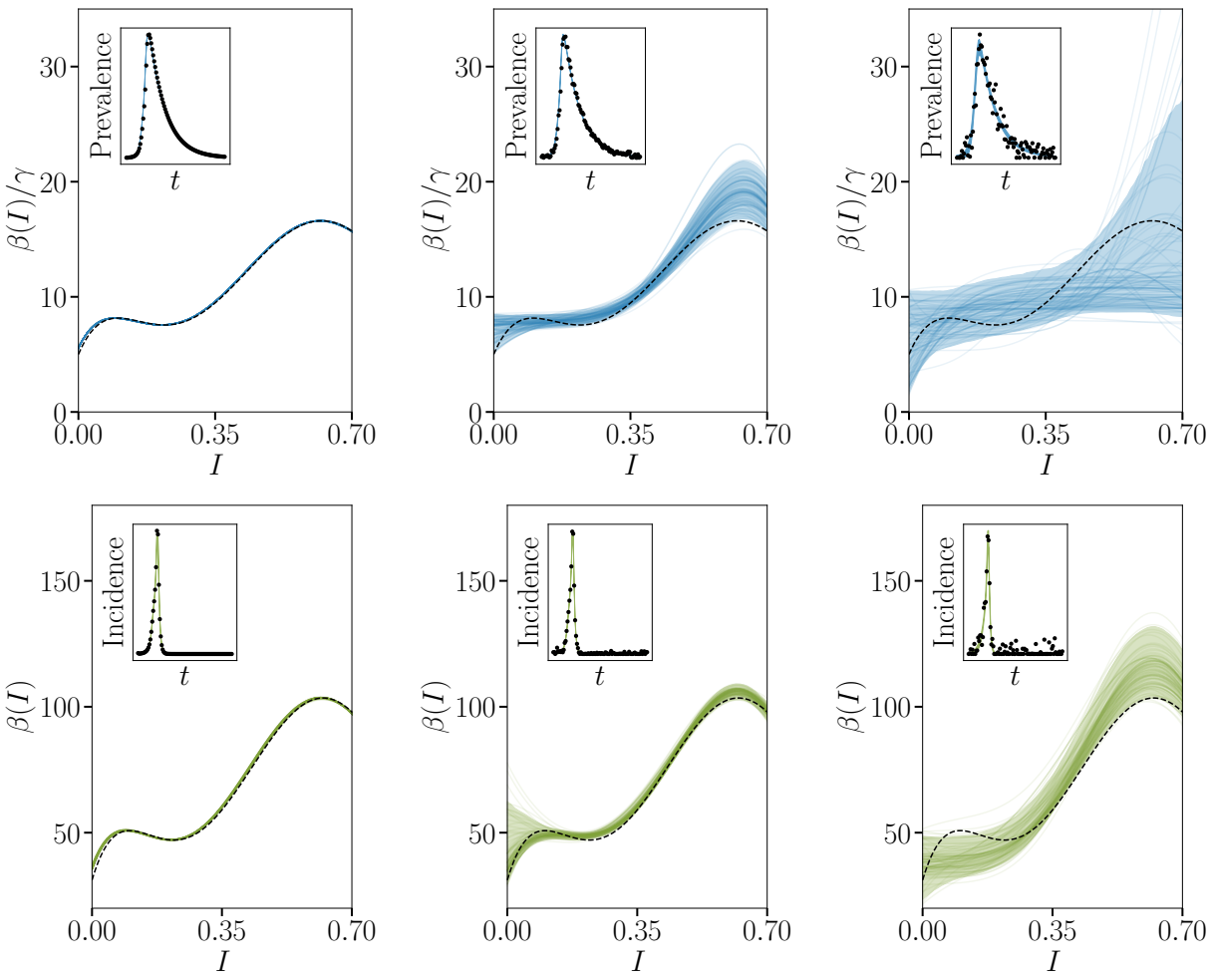

FIG. 3. **Inference at different levels of noise.** See the caption of 2 for how to read this figure. The artificial time series are obtained by sampling the latent time series at $T = 100$ regularly spaced points. An independent Gaussian noise of variance (from left-to-right) $\sigma^2 = 0, 0.1$ and 0.5 is applied to the prevalence measurements. The variance is set to $10 \times \sigma^2 \times n \times T$ for the incidence measurements.

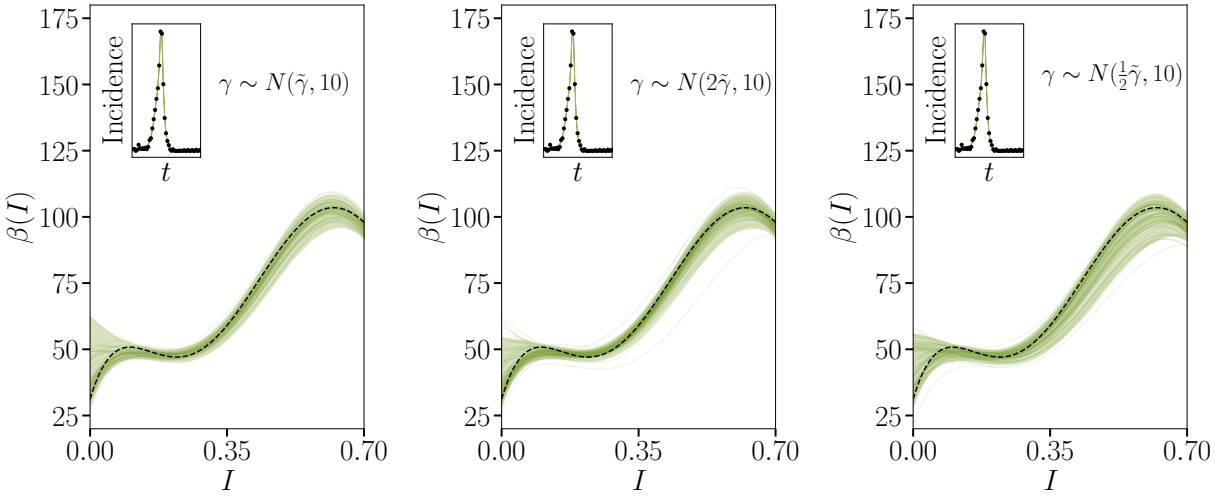

FIG. 4. **Inference with incidence data and vague imprecise priors.** See the caption of 2 for how to read this figure. The ground-truth is $\tilde{\gamma} = 0.1 t_{\max}$ where $t_{\max} \approx 7$ is a constant introduced in the data normalization step (see Sec. I B 4). In the 3 cases shown, the location of the prior on γ is inaccurate, but the the variance is comparatively large.

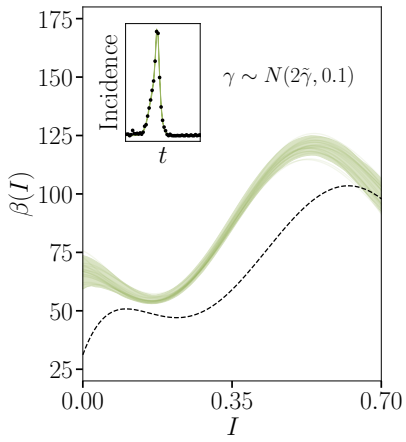

FIG. 5. **Inference with incidence data and tight imprecise priors.** Same experiment as a in Fig. 4, but this time we use a much narrower prior. The posterior moves away from the ground-truth. In the case shown, the infection rate is shifted to compensate for the increased infection rate, leaving the a posteriori incidence curves basically unaltered.

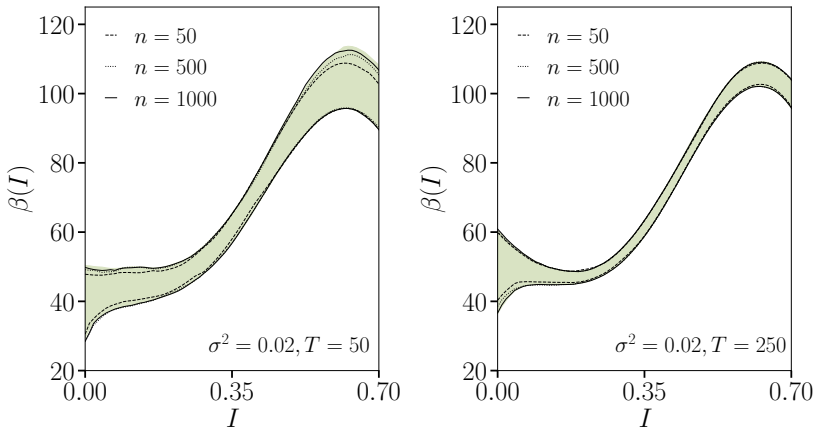

FIG. 6. **Impact of the number of samples.** Regions covered by 95% of the posterior draws, when n samples are used, for two different levels of temporal sampling T. The shaded regions shows this range for $n = 2000$ samples (the actual number of sample used in all our experiments). Even for temporally sparse data ($T = 50$), a few samples suffice to determine the confidence interval.

II. SIMULATIONS OF CONTAGION MODELS ON CONTACT NETWORKS

To simulate all contagion models, we use a discrete time process with transmission and recovery probabilities p_t and p_r lesser than 10^{-3} to mimic the continuous time process used in our ODEs. At every time step, a contact between an infectious and susceptible node can transmit the contagion with probability p_t and infectious nodes can recover with probability p_r. In all simulations, we initiate the system by infecting a random fraction 10^{-3} of nodes with the simple contagion, the complex contagion, or both interacting contagions.

A. Simulations in the main text

All simulations from the main text take place on two simple network topologies: clustered network structures and equivalent random networks. Clustered networks are obtained by assigning every node to m of cliques of size n, with k additional random neighbors. Equivalent random networks are obtained by keeping the same degree sequence but randomizing every connection to destroy cliques and reduce clustering close to zero. As discussed in the main text, we use regular networks where all nodes have the same degree to avoid conflicting the role of clustering with that of degree correlations.

The results of our simulations reported in the main text concern the global fraction of infected nodes at a given time (prevalence), number of new infections since the last observation (incidence) and local state correlations around infection events. To summarize the results shown in the main text: even though we parametrized all simulations to have a similar prevalence after 5000 time steps, complex and interacting contagions are more spatially correlated and therefore benefit from network clustering; all of these results can be used in practice to distinguish complex and interacting contagions from simple contagions, but not from each other. Additionally to the results shown in the main text, we find similar results for the number of infectious neighbours of individuals upon recovery. Indeed, on clustered networks, that number increases to 145% (median, 50% CI [130%,160%]) that of random networks for simple contagions, and 188% (median, 50% CI [166% 213%]) and 180% (median, 50% CI [144% 205%]) for interacting and complex contagions, respectively.

B. Simulations on complex network data

We now reproduce some of the results from the main text using an empirical complex network where both ideas, behaviours and biological contagions can potentially spread: the arXiv co-authorship network. In the data, nodes are authors of arXiv preprint and edges between them represent co-authorship status. This network is therefore the projection of a bipartite network between authors and papers, such that there is a significant amount of clustering but also highly heterogeneous distributions of both degree and clique size [2]. Using this heterogeneous network, we mostly want to show that the approaches used in the main text to compare contagions lead to similar, albeit noisier, results on very heterogeneous networks.

On the one hand, Fig. 7 shows that correlations, especially in the number of infectious neighbours upon recovery, can again be a useful way of distinguishing interacting contagions from simple contagions on real network data. However, simulations of complex contagions are too noisy on heterogeneous networks such as the arXiv to make any definitive statement. This is mostly due to the fact that a majority of nodes have a degree 1, and therefore simply cannot benefit from any reinforcement of the complex contagions.

On the other hand, we do not expect the results of Fig. 3 from the main text to translate directly to heterogeneous systems. Indeed, our inference procedure assumes an homogeneous structure such that we know a simple contagion on such a structure will result in a mostly flat $\beta(I)$. Based on the results from Ref. [3], we know that using our inference procedure out of the box on time series from simple contagion on heterogeneous structures will not result in a flat $\beta(I)$, and instead will capture structural heterogeneities through the $\beta(I)$ function. Most importantly, we expect high degree nodes to be infected earlier in the contagion process such that the spread should slow down as hubs are depleted, leading to a $\beta(I)$ decreasing with I. We could extend our model to account for a heterogeneous degree distribution, either through a heterogeneous mean-field description or by including the first few moments of the degree distribution in moment-closure approximations. Yet, the simplest approach is to leverage the results from Ref. [3] and run a simple contagion as a baseline to capture structural heterogeneity, and then compare other $\beta(I)$ functions to that baseline to detect interactions.

We apply this simple approach to the time series of Fig. 7(a). The results of the inference procedure are given in Fig. 8. As expected, our inference procedure captures the early depletion of high degree individuals through a monotonously decreasing $\beta(I)$ for simple contagions. Contrarily, for interacting contagions, a relatively low interaction factor is enough to upset this depletion and lead to an initial increase in $\beta(I)$, followed by a slower decrease near the epidemic peak, driven by depletion.

The impact of interactions in counteracting the decrease in $\beta(I)$ due to depletion of high degree individuals is further highlighted in Fig. 9.

Altogether, one can use a baseline relying on a known simple contagion to capture some structural features of the system and then distinguish contagions, without having to rely to higher-order mean-field models.

C. Model selection of simple and complex contagion

In Ref. [4], the authors use a different approach to distinguish simple and complex contagions. Briefly, they fit specific models of both simple and complex contagions; the latter is parametrized with a sigmoid form for $\beta(i)$ around nodes with i infectious neighbours, as we used in Fig. 7. They then calculate the likelihood of some observational data given those specific models and compare the results using a Bayesian Information Criterion (BIC).

We here apply a similar approach. We use a local network model where nodes belong to two cliques of fixed size and we track the density $S_{i,j}$ of nodes with i and j infectious neighbours in their two cliques. This is straightforwardly done using the following set of ODEs, now using a simpler SI dynamics for simplicity (recovery rate $\gamma = 0$):

$$\frac{d}{dt}S_{i,j} = -\beta(i+j)S_{i,j} - (n-1-i)\beta(i+\Theta_i)S_{i,j} - (n-1-j)\beta(j+\Theta_j)S_{i,j} + (n-i)\beta(i-1+\Theta_{i-1})S_{i-1,j} + (n-j)\beta(j-1+\Theta_{j-1})S_{i,j-1}$$

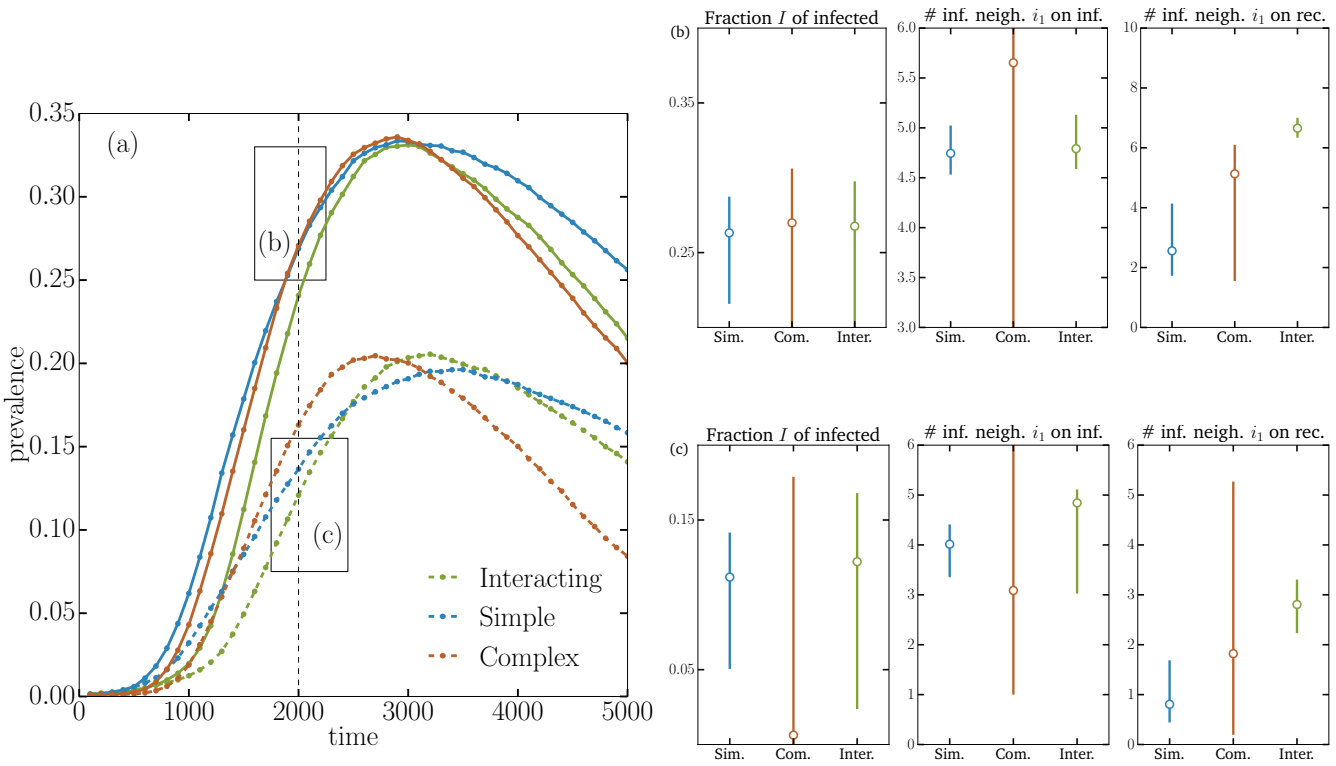

FIG. 7. **Time series produced with different contagion models and their local statistics.** Time series are shown in panel (a). The time series leading to a high epidemic peak were produced by a simple contagion with transmission rate 0.00027 and recovery rate 0.00037, interacting contagions with transmission rate 0.0001 and recovery rate 0.001 boosted by a symmetric interaction factor of 5 increasing the transmission rate and decreasing the recovery rate, and a complex contagion produced with a transmission rate $\beta(i) = 0.0002 * [1 + 0.75/(1 + \exp(-2 * i - 4))]$ and recovery rate $\gamma = 0.002/i$ around indivduals with i infectious neighbours. Their local statistics are highlighted in panel (b). The time series leading to a lower epidemic peak were produced by a simple contagion with infection rate 0.00019 and recovery rate 0.00005, interacting contagions with transmission rate 0.0001 and recovery rate 0.001 and *asymmetric* interaction factors where disease 1 sees its infection rate boosted and recovery rate decreased by a factor of 5 while disease 2 gets half the benefit, and a complex contagion with the same parametrization as before but with a base transmission rate of $5/3 \times 10^{-4}$ instead of 0.0002. Their local statistics are highlighted in panel (c). Unlike those shown in the main text, the complex contagions were not parametrized to reproduce the full statistics of the interacting contagions but simply similar time series. They do highlight, however, the high levels of noise noise expected in local statistics on real complex networks, which is why more homogeneous systems were used in the main text.

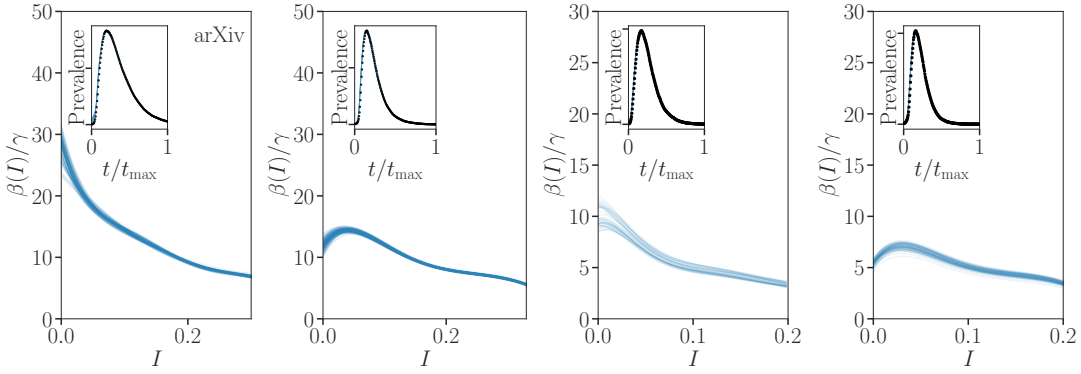

FIG. 8. **Inferring a complex transmission function on the simple and interacting time series of Fig. 7(a).** The time series correspond, from left to right, to the simple and symmetric interacting contagions with higher peak prevalence whose statistics are shown in Fig. 7(b), followed by the simple and asymmetric interacting contagions with lower peak prevalence whose statistics are shown in Fig. 7(c). The key signature of interaction here is the initially increasing $\beta(I)$ caused by the fact that interactions counteract the depletion of highly connected individuals which cause a decrease in $\beta(I)$ for simple contagions. This is further highlighted below on Fig. 9.

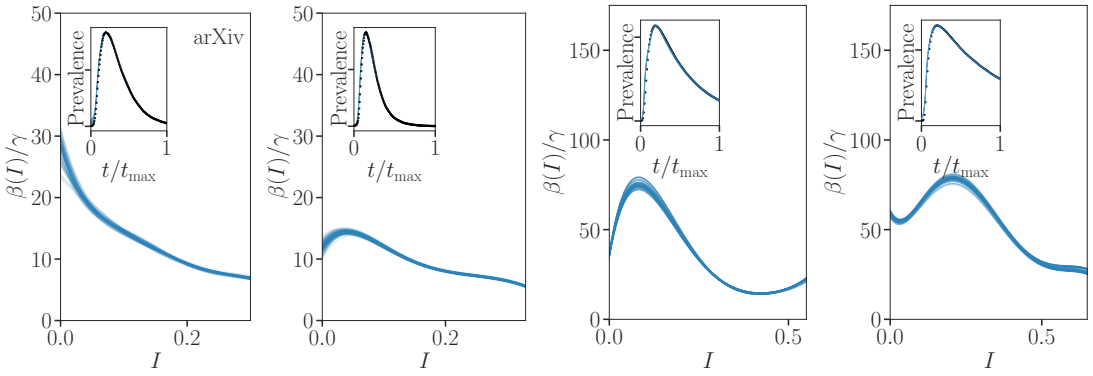

FIG. 9. **Observing increasing interaction factors through the complex transmission function.** We infer $\beta(I)$ on time series produced with different levels of interaction factors. The leftmost panel correspond to the simple contagion used in previous figures. The second panel correspond to the symmetric interacting contagions from the previous figures, which use an interaction factor of 5. The next two panels use the same parametrization but with an interaction factor of 10 and 15, respectively. Note that the regime of increase $\beta(I)$ takes a larger and larger fraction of the I space as the interaction factor increases.

where $\Theta_x = \sum_y y(S_{x,y}+S_{y,x})/\sum_y (S_{x,y}+S_{y,x})$ are mean-field values around susceptible nodes in a clique with x infectious individuals, describing the average number of infectious nodes in their other clique.

Using time series produced by interacting contagions with 10 different interaction factors (while keeping other parameters fixed), we fit this network model using either a simple contagion $\beta(i) = \beta_0$ with one parameter or a complex contagion with $\beta(i) = \beta_- + \Delta/\left[1+\exp(-wi+wi_0)\right]$ with 4 parameters. Note that both models are wrong, but one can still be chosen to model the data by minimizing the BIC. The results of this experiment are shown in Fig. 10.

For low interaction factors, the approach appropriately identifies simple contagions as the preferred model. Indeed, not only is it close to correct for low interaction factors, but it also contains less parameters and is therefore less penalized. As the interaction factor increases, the complex contagion model is eventually preferred. Therefore, if we were not to consider the potential for interacting contagions, we might wrongly categorize these time series as complex contagions, even if the simulation code that produced them follows a pairwise contagion model.

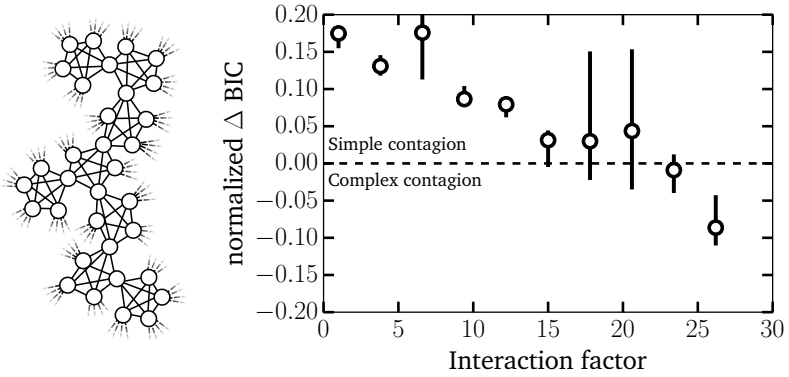

FIG. 10. **Selecting from simpler models with a Bayesian Information Criterion to fit time series from interacting contagions.** (left) The network used in these simulations: every node belongs to 2 cliques of size 5. (right) We produce time series using a SI model of interacting contagions (no recovery), and fit them with local network models using ODEs of both simple and complex contagions. The simulations use two diseases with transmission rate $\beta_1 = \beta_2 = 1.2 \cdot 10^{-5}$ with interactions such that their transmission rate is increased by a fixed interaction factor (horizontal axis) whenever the other disease is present in any way on a given edge. While the simple contagion model is initially preferred because it uses less parameters, the complex contagion is eventually preferred once the interactions in the underlying simulations become strong enough.

[1] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell, Journal of Statistical Software **76** (2017).

[2] L. Hébert-Dufresne, A. Allard, V. Marceau, P.-A. Noël, and L. Dubé, Phys. Rev. Lett. **107**, 158702 (2011).
[3] Q.-H. Liu, M. Ajelli, A. Aleta, S. Merler, Y. Moreno, and A. Vespignani, Proceedings of the National Academy of Sciences **115**, 12680 (2018).
[4] B. Mønsted, P. Sapieżyński, E. Ferrara, and S. Lehmann, PloS one **12**, e0184148 (2017).