Noname manuscript No.
(will be inserted by the editor)

A Mean-Field Approximation of SIR Epidemics on an Erdös-Rényi Network Model

Christina Durón · Alex Farrell

Received: date / Revised version: date

Abstract The stochastic nature of epidemic dynamics on a network makes their direct study very challenging. One avenue to reduce the complexity is a mean-field approximation (or mean-field equation) of the dynamics; however, the classic mean-field equation has been shown to perform sub-optimally in many applications. Here, we adapt a recently developed mean-field equation for SIR epidemics on a network in continuous time to the discrete time case. With this new discrete mean-field approximation, this proof-of-concept study shows that, given the density of the network, there is a strong correspondence between the epidemics on an Erdös-Rényi network and a system of discrete equations. Through this connection, we developed a parameter fitting procedure that allowed us to use synthetic daily SIR data to approximate the underlying SIR epidemic parameters on the network. This procedure has improved accuracy in the estimation of the network epidemic parameters as the network density increases, and is extremely cheap computationally.

Keywords Mean-field approximation, SIR epidemic model, Discrete time, Difference equations, Erdös-Rényi networks, Parameter estimation

1 Introduction

Computational modeling has been an important tool for understanding the spread of epidemics, predicting the severity of an outbreak, evaluating the effectiveness of interventions, and optimizing the deployment of new control policies (Wan et al., 2014). While the most commonly used framework to describe the dynamics of such diffusion processes are compartmental models,

Christina Durón
Department of Mathematics, University of Arizona, Tucson, AZ, USA
E-mail: duronc@math.arizona.edu

Alex Farrell
Department of Mathematics, University of Arizona, Tucson, AZ, USA

e.g., susceptible-infective-recovered (SIR) (Kermack and McKendrick, 1927), due to their simplicity and generality (Wang et al., 2021; Tolles and Luong, 2020; Zhang et al., 2017; Bjørnstad et al., 2020), network epidemic models can provide a more precise alternative. Since epidemics reflect the dynamics of complex systems, network models provide a natural framework to study their spread on a population where the individuals are represented as nodes and their interactions as edges.

Unfortunately, the direct study of epidemic dynamics on networks is very challenging due to the stochastic nature of their behavior, and becomes intractable as the number of nodes grows (Sahneh et al., 2013). Consequently, the complexity of the system can be reduced by applying the mean-field approximation (MFA) which assumes that the individuals within the population are well-mixed and interact with each other completely at random (Pastor-Satorras et al., 2015). For example, in SIS epidemics on networks, the MFA replaces the actual infection rate towards an individual node by its average rate (Van Mieghem and van de Bovenkamp, 2015).

While the study of MFA epidemic models has led to various results (see Van Mieghem et al. (2008); Wu and Fu (2016) for some examples), our focus is on its reduction of the complexity of the system. However, this simplification comes at the expense of exactness (Sahneh et al., 2013), e.g., the accuracy of the mean-field model is highly dependent upon the range of the epidemic parameters.

To address this disparity between the classic mean-field equation and simulations of the SIS model on various network models, Zhang et al. (2017) proposed a modified mean-field equation for continuous time epidemic models, such as SIS, SIR, and SEIR. In particular, Zhang et al. replaced $\beta(t)\cdot<k>\cdot\rho(t)$, the term depicting the probability that a susceptible node is infected by its neighbor, with $1-(1-\beta(t))^{<k>\cdot\rho(t)}$, where $\beta(t)$ is the infection probability at time t, $\rho(t)$ is the infection density at time t, and $<k>$ is the average degree of the network. By comparing calculations of the classic and their modified MFA equation, Zhang et al. showed that their modified equation outperformed the classic MFA on regular random networks, Erdös-Rényi networks, and small-world networks.

In this paper, we extend the results of Zhang et al. to a discrete time SIR model on undirected and unweighted Erdös-Rényi (ER) networks of *known* density d, where the ER network $G(N,m)$ is built from a set of N nodes and m edges in which each of the $N(N-1)/2$ possible edges is present with probability p (Erdős and Rényi, 1960). The main contributions of our proof-of-concept study are a proposed MFA infection function and a parameter fitting procedure for SIR epidemics on ER networks that utilizes this infection function and a system of difference equations. Thus, given synthetic daily SIR data and the network density, the discrete SIR model can uncover the underlying SIR epidemic parameters on the network in a quick and reliable fashion in many instances. Furthermore, given the density of the network, we show analytical and numerical agreement with the mean-field and SIR epidemic simulations on the Erdös-Rényi network as the density increases.

The paper is organized as follows; Section 2 begins with the construction of the discrete SIR model, along with the proposal and validation of the MFA infection function, before concluding with a discussion on the procedure to create a one-to-one correspondence between the discrete model's SIR epidemic parameters (i.e., infective, recovery, and death probabilities) and the ER network model's SIR epidemic parameters. The results of the procedure applied to ER networks of various size, density, and epidemic parameters are detailed in Section 3, while our concluding remarks are in Section 4.

2 Methods

2.1 The Discrete SIR Model

The discrete SIR model has four possible states: susceptible S, infective I, recovered R, or dead D. From first principles, we have

$$\begin{aligned} S(t+1) &= S(t) - f(S(t), I(t), \beta, d), \\ I(t+1) &= I(t) + f(S(t), I(t), \beta, d) - (\mu + \rho)I(t), \\ R(t+1) &= R(t) + \rho I(t), \\ D(t+1) &= D(t) + \mu I(t). \end{aligned} \tag{1}$$

where $f(S(t), I(t), \beta, d)$ is an infection function that describes how many susceptible individuals, on average, are infected during one time step, and is dependent upon the number of susceptible and infective individuals (S and I, respectively), as well as the infection probability β and the network density d (i.e., the portion of all possible edges that are realized in the network). The remaining parameters ρ and μ in Equation (1) are the probability of recovery from and succumbing to the disease, respectively.

Since the number of nodes in the network N is fixed and known, where

$$N = S(t) + I(t) + R(t) + D(t). \tag{2}$$

we can remove any one compartment (S, I, R, or D) to reduce the system. For example, to obtain the value of $D(t)$, we rearrange Equation (2) such that $D(t) = N - S(t) - I(t) - R(t)$.

Without loss of generality, we have removed $D(t)$ and only need to consider the following SIR system of three equations:

$$\begin{aligned} S(t+1) &= S(t) - f(S(t), I(t), \beta, d), \\ I(t+1) &= I(t) + f(S(t), I(t), \beta, d) - (\mu + \rho)I(t), \\ R(t+1) &= R(t) + \rho I(t). \end{aligned} \tag{3}$$

We note the decision to exclude D is purely to create an 'SIR' model; we could have chosen any one of S, I, R, or D to be removed and not lose any information. Depending on the data at hand, it may be more useful to exclude another compartment.

2.2 The Updated Discrete SIR Model with Proposed Infection Function

We define our MFA infection function f in our discrete SIR model as

$$f(S, I, \beta, d) = S(t)\left(1 - (1-\beta)^{d \cdot I(t)}\right). \tag{4}$$

After some simplification, this yields the updated discrete SIR model:

$$\begin{aligned} S(t+1) &= S(t)(1-\beta)^{d \cdot I(t)}, \\ I(t+1) &= I(t) + S(t)\left(1 - (1-\beta)^{d \cdot I(t)}\right) - (\mu + \rho)I(t), \\ R(t+1) &= R(t) + \rho I(t). \end{aligned} \tag{5}$$

The validation of our proposed MFA infection function f, in mean terms, is provided in Section 2.3.

2.3 Discrete Mean-Field Approximation of Susceptible Population

In a SIR epidemic process, the state of the node v_i at time t is specified by a state variable $X_i(t) \in \{0, 1, 2, -1\}$: $X_i(t) = 0$ for a susceptible node, $X_i(t) = 1$ for an infective node, $X_i(t) = 2$ for a recovered node, and $X_i(t) = -1$ for a dead node.

We define $E[S(t+1)]$, the expectation of the total number of susceptible nodes at time $t+1$, as

$$E[S(t+1)] = E\left[\sum_{i=1}^{N} 1_{\{X_i(t)=0\}} \cdot (1-\beta)^{\sum_{j=1}^{N} a_{i,j} \cdot 1_{\{X_j(t)=1\}}}\right] \tag{6}$$

where $1_{\{\cdot\}}$ is the indicator function, and for an undirected and unweighted network of N nodes described by adjacency matrix $A = \{a_{i,j}\}$ of size $N \times N$, entry $a_{i,j} = 1$ if nodes v_i and v_j are connected by an edge, and $a_{i,j} = 0$ otherwise.

For numerical support of Equation (6), consider Figure 1 in which we simulate an epidemic with parameters (β, μ, ρ) on ER networks of size N and density d over a 30 day period. For $N = 1000$ and $N = 5000$, we simulate an epidemic with $d = 0.03$ and $(\beta, \mu, \rho) = (0.07, 0.1, 0.21)$, $d = 0.1$ and $(\beta, \mu, \rho) = (0.09, 0.19, 0)$, and $d = 0.6$ and $(\beta, \mu, \rho) = (0.05, 0.16, 0.07)$ on the ER network over 30 time steps (such that we have a total of 6 different epidemic scenarios). At each time t, we calculate both $\sum_{i=1}^{N} 1_{\{X_i(t)=0\}} \cdot (1-\beta)^{\sum_{j=1}^{N} a_{i,j} \cdot 1_{\{X_j(t)=1\}}}$ and $S(t+1)$, which are denoted by the $(x, y)-$coordinate, respectively, of the black points in Figure 1. To approximate the left and right-hand expressions of Equation (6), we average the 1000 simulations of $S(t+1)$ and $\sum_{i=1}^{N} 1_{\{X_i(t)=0\}} \cdot (1-\beta)^{\sum_{j=1}^{N} a_{i,j} \cdot 1_{\{X_j(t)=1\}}}$ at each time t over the 30 day period. Our numerical simulations indicate that the means of $S(t+1)$ and $\sum_{i=1}^{N} 1_{\{X_i(t)=0\}} \cdot (1 -$

$\beta)^{\sum_{j=1}^{N} a_{i,j} \cdot 1_{\{X_j(t)=1\}}}$ at each time t is approximated by the red line $y = x$, thus giving numerical support of Equation (6).

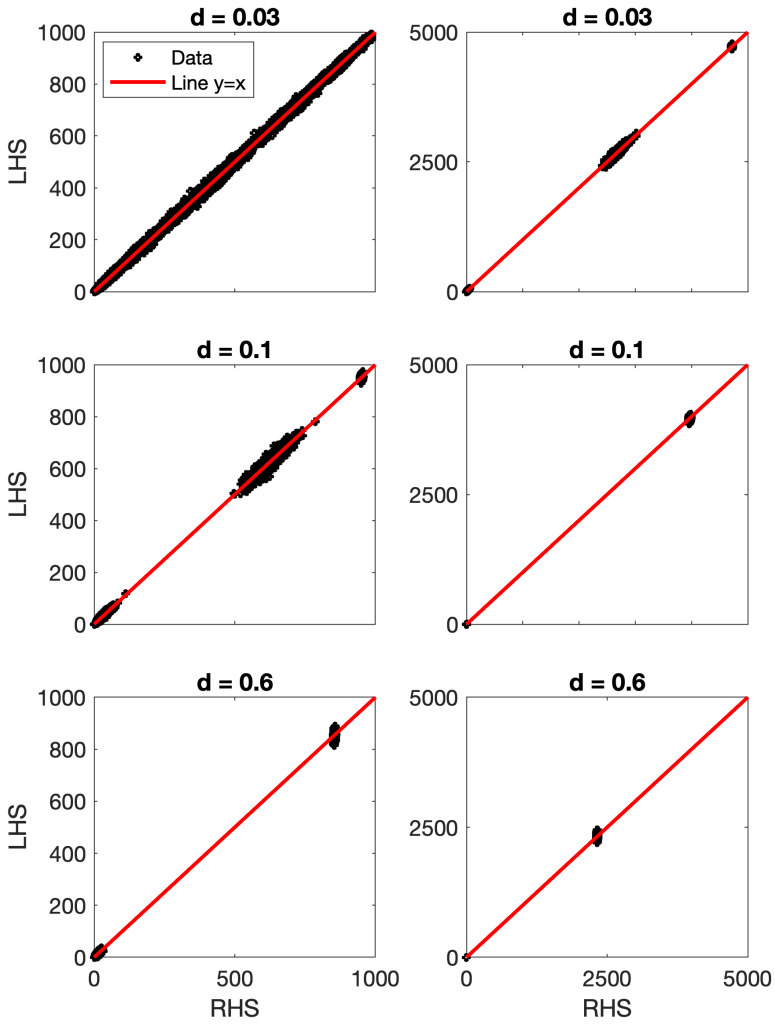

Fig. 1 A comparison of the left and right-hand expressions (LHS and RHS, respectively) of Equation (6) for ER networks of size $N = 1000$ (column 1) and $N = 5000$ (column 2), and density $d = 0.03$ (row 1), $d = 0.1$ (row 2), and $d = 0.6$ (row 3). The epidemic parameters (β, μ, ρ) equal (0.07, 0.1, 0.21) for row 1, $(0.09, 0.19, 0)$ for row 2, and $(0.05, 0.16, 0.07)$ for row 3. Each black point ('Data') denotes $\sum_{i=1}^{N} 1_{\{X_i(t)=0\}} \cdot (1-\beta)^{\sum_{j=1}^{N} a_{i,j} \cdot 1_{\{X_j(t)=1\}}}$ ($x-$coordinate) versus $S(t+1)$ ($y-$coordinate) for 1000 simulations of an epidemic over a 30 day period.

For $0 < d \leq 1$, an *approximate* upper bound on the total number of susceptible nodes at time $t+1$ is

$$E\left[S(t+1)\right] \geq E\left[S(t)\right] \cdot (1-\beta)^{d \cdot E[I(t)]} \tag{7}$$

since

$$\begin{aligned}
E\left[S(t+1)\right] &= E\left[\sum_{i=1}^{N} 1_{\{X_i(t)=0\}} \cdot (1-\beta)^{\sum_{j=1}^{N} a_{i,j} \cdot 1_{\{X_j(t)=1\}}}\right] \\
&= \sum_{i=1}^{N} E\left[1_{\{X_i(t)=0\}} \cdot (1-\beta)^{\sum_{j=1}^{N} a_{i,j} \cdot 1_{\{X_j(t)=1\}}}\right] \\
&\approx \sum_{i=1}^{N} E\left[1_{\{X_i(t)=0\}}\right] \cdot E\left[(1-\beta)^{\sum_{j=1}^{N} a_{i,j} \cdot 1_{\{X_j(t)=1\}}}\right] \qquad (8) \\
&\geq \sum_{i=1}^{N} E\left[1_{\{X_i(t)=0\}}\right] \cdot (1-\beta)^{E\left[\sum_{j=1}^{N} a_{i,j} \cdot 1_{\{X_j(t)=1\}}\right]} \qquad (9) \\
&\approx \sum_{i=1}^{N} E\left[1_{\{X_i(t)=0\}}\right] \cdot (1-\beta)^{<k> \cdot \frac{E[I(t)]}{N-1}} \qquad (10) \\
&= \sum_{i=1}^{N} E\left[1_{\{X_i(t)=0\}}\right] \cdot (1-\beta)^{d \cdot E[I(t)]} \\
&= E\left[S(t)\right] \cdot (1-\beta)^{d \cdot E[I(t)]}
\end{aligned}$$

where $<k>$ denotes the average degree of the network and $d = \frac{<k>}{N-1}$ denotes the network density.

Since spatial correlations between any pair of nodes (e.g., if one node is infected, then its neighbors are likely to be infected) make an exact analysis of the SIR model on a network difficult (Rodríguez et al., 2018; Cai et al., 2016), the mean-field approximation that the expected values of random variable pairs factorize is applied in Equation (8), as it is assumed that there are no dynamical correlations.

In Equation (9), Jensen's inequality (Jensen et al., 1906) is applied since the function

$$(1-\beta)^{\sum_{j=1}^{N} a_{i,j} \cdot 1_{\{X_j(t)=1\}}}$$

is convex for $1-\beta > 0$ with $\beta \in (0,1)$.

In Equation (10), we approximate the number of infective neighbors of node v_i, denoted by the sum $\sum_{j=1}^{N} a_{i,j} \cdot 1_{\{X_j(t)=1\}}$, with the product $<k> \cdot \frac{I(t)}{N-1}$, which holds as long as the degree-distribution of the network has low variance (which is true for ER networks) and the mean-degree is high (Vajdi et al., 2018).

We compare our MFA f to an infection function found an another discrete model (created by Lowell Reed and Wade Hampton Frost) for fully connected

networks (Abbey, 1952), i.e., ER networks with $d = 1$. By the Reed-Frost model,

$$S(t+1) = S(t) \cdot (1-\beta)^{I(t)} \tag{11}$$

such that upon substituting Equation (11) into Equation (3), we get the infection function

$$f(S, I, \beta) = S(t)\left(1 - (1-\beta)^{I(t)}\right) \tag{12}$$

which, we note, is Equation (4) with $d = 1$. This lends additional support to our MFA infection function.

2.4 Overview of the Fitting Procedure

With an MFA infection function in hand, our goal was to create a correspondence between the parameters of the updated discrete time SIR epidemic model and the network SIR epidemic model. By simulating epidemics on multiple networks, we hoped to achieve a true MFA for any ER network given only its density d and size N, and no additional information provided about the network structure. To that end, we begin this subsection with an overview of the parameter fitting procedure before detailing the methodology in full in Subsections 2.5 – 2.7.

First, we select a parameter set consisting of the network density, probability of infection, probability of dying from the disease, and probability of recovering from the disease, (d, β, μ, ρ). With these parameters, we simulate an epidemic on a ER network of size N and density d to create synthetic day-by-day data for the susceptible, infective, and recovered populations over $t_n = 30$ days (where we emphasize that although the procedure is applied to data over 30 days in this work, it may be generalized to any positive length of time t_n). For the $N = 1000$ and $N = 5000$ node networks, 100 networks are created for each density, and a single epidemic is simulated on each network. Due to computational constraints, this process is done once for the networks with 10000 and 50000 nodes.

Then, in an attempt to remove stochastic variations, we take the average of the 100 simulated epidemics to create a single 30 day set of S, I, and R synthetic data for the 1000 and 5000 node networks. For larger networks ($N > 5000$), the single 30 day set of S, I, and R synthetic data is the single epidemic simulation.

Next, we fit our updated discrete SIR model to the single 30 day data. Finally, we compare the epidemic parameters from the ER network, (β, μ, ρ), to the best fit epidemic parameters from the updated discrete time model, $(\tilde{\beta}, \tilde{\mu}, \tilde{\rho})$, to gauge the accuracy of the parameter fitting procedure.

This process is repeated for all of the parameter combinations prescribed to the networks of size N in Subsection 2.5. For computational time's sake, we reduce the parameter space explored for larger networks.

For an example of how the discrete model's SIR predictions align with the simulated network SIR data, consider Figure 2. We generate 100 ER networks

with $N = 1000$ and density $d = 0.02$. For each network, we simulate an epidemic with parameters $(\beta, \mu, \rho) = (0.05, 0.16, 0.07)$ over a 30 day period. In total, 100 simulations are performed, with one simulated epidemic per network. The graphs of the S, I, and R compartments of the 100 individual simulations performed on the networks are denoted by the black curves in Figure 2, while the average of the individual network simulations for each S, I, and R are denoted by the red diamonds. Using the fitting procedure, we determine the best fit parameters from the updated discrete time model, $(\tilde{\beta}, \tilde{\mu}, \tilde{\rho})$, and use them to generate the graphs of the S, I, and R compartments which are denoted by the green curves. A comparison of the graphs of the S, I, and R data generated by the discrete fit (green line) with those of the mean network data (red diamonds) indicates alignment between the discrete model's predictions and the simulated network data.

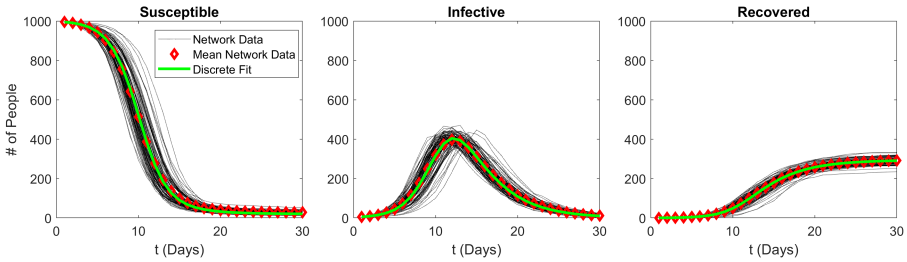

Fig. 2 An example of how the discrete model's predictions align with the mean of the simulated network data (100 simulations, one per network) for an epidemic simulated over a 30 day period. Each ER network contained $N = 1000$ nodes, had density $d = 0.02$, and network epidemic parameters $(\beta, \mu, \rho) = (0.05, 0.16, 0.07)$.

We refer the reader to Supplemental Figures S1 and S2 for additional examples of the sample graphs of the S, I, and R compartments of the individual network simulations (network data), mean of the simulations (mean network data), and the best fit provided by the discrete model (discrete fit). In these figures, 100 ER networks with $N = 1000$ are generated at densities $d = 0.01$ and $d = 0.06$, respectively, and for each network, an epidemic with parameters $(\beta, \mu, \rho) = (0.05, 0.16, 0.07)$ is simulated for a 30 day period.

We now describe the parameter fitting procedure in more detail in Subsections 2.5 – 2.7, where we have generalized the procedure to an epidemic spanning a period of t_n days. Tables and figures of the results of the procedure are provided in Section 3, while additional tables with the median and standard error of the fits are provided in the Supplementary Materials.

2.5 Network Simulations

The network simulations for $N = 1000, 5000, 10000$, and 50000 nodes are performed on a RStudio Virtual Machine. All simulations are run on a Windows PC with 16 cores, 32 threads at 4.5 GHz with 128 GB of RAM at 3200 MHz.

In `R`, ER networks are simulated using the function `sample_gnm(N, m)` from the `igraph` package (Csardi and Nepusz, 2006) by specifying the number of nodes `N` and the number of edges `m` (where $m = \frac{d \cdot N \cdot (N-1)}{2}$ for undirected networks). The function `sample_gnm` contains two additional parameters, `directed` and `loops`, which indicate whether the network will be directed and whether to add loop edges, respectively. The default value (`FALSE`) for both logical parameters is used in the ER network simulations.

Given the network parameters (β, μ, ρ), we determine the state of the network at time $t + 1$ through the following process:

- For each node that is susceptible at time t (i.e., each individual in $S(t)$), we draw a random number $u \in [0, 1]$, and count the number of neighbors that are infected at time t, which we denote $i(t)$. Each $i(t)$ has probability β to infect the node in question, thus
  - if $u < (1 - \beta)^{i(t)}$, then the node remains susceptible at time $t + 1$.
  - if $(1 - \beta)^{i(t)} \leq u$, then the node becomes infective (i.e., convert this individual to I) at time $t + 1$.
- For each node that is infective at time t (i.e., each individual in $I(t)$), we draw a random number $u \in [0, 1]$.
  - if $u < \mu$, then the infective node is dead (i.e., convert this individual to D) at time $t + 1$.
  - if $\mu \leq u < \mu + \rho$, then the node is recovered (i.e., convert this individual to R) at time $t + 1$.
  - if $\mu + \rho \leq u$, then the node remains infective at time $t + 1$.

We begin at $t = 1$ with a network containing $N \times 0.005$ randomly chosen individuals initially in the infective state. We assume the epidemic lasts for t_n days, and so we repeat the process above $t_n - 1$ times to create the network's state for $t = 2$ through $t = t_n$, where $\triangle t = 1$.

For $N = 1000$ nodes, 100 ER networks are created for each density,

$$d \in \{0.01, 0.02, 0.03, 0.06, 0.1, 0.3, 0.6, 1\}$$

and a single epidemic is simulated on each network for each possible parameter combination (β, μ, ρ),

$$\begin{aligned} \beta &\in \{0.01, 0.02, 0.03, 0.04, \ldots, 0.2\}, \\ \mu &\in \{0, 0.01, 0.02, 0.03, \ldots, 0.25\}, \\ \rho &\in \{0, 0.01, 0.02, 0.03, \ldots, 0.25\}. \end{aligned}$$

Similarly, 100 ER networks of size $N = 5000$ are created for each density,

$$d \in \{0.01, 0.02, 0.03, 0.06, 0.1, 0.3, 0.6, 1\}$$

and a single epidemic is simulated on each network for each possible parameter combination (β, μ, ρ),

$$\beta \in \{0.01, 0.02, 0.04, 0.06, \ldots, 0.2\},$$
$$\mu \in \{0, 0.01, 0.03, 0.05, \ldots, 0.25\},$$
$$\rho \in \{0, 0.01, 0.03, 0.05, \ldots, 0.25\}.$$

For both $N = 10000$ and $N = 50000$ nodes, 1 ER network is created for each density,

$$d \in \{0.001, 0.005, 0.01, 0.03\}$$

and a single epidemic is simulated on each network for each parameter set described in Table 1 (thus, a total of 3 epidemics are simulated on each network).

Table 1 The three parameter sets used to simulate an epidemic on an ER network of size $N = 10000$ and $N = 50000$.

Set	β	μ	ρ
1	0.07	0.1	0.21
2	0.09	0.19	0
3	0.05	0.16	0.07

To ensure there are no simulations that resulted in a 'non-epidemic' due to their stochastic nature, we resimulate any epidemics that have no infective individuals at time $t = 10$ and have less than 0.5% of the population infected by time $t = t_n$.

For each set of 100 simulated networks, the mean and standard deviation of the basic structural features of networks of size $N = 1000$ and $N = 5000$ with m edges are summarized in Tables S7 and S8, respectively, where the structural features include the edges m, the average degree of the network $<k>$, the clustering coefficient $<cc>$ (i.e., a measure of the degree to which nodes in a network tend to cluster together), and the *diameter* (i.e., the maximum length of all shortest paths between any two nodes). Table S9 summarizes the basic structural features for the simulated networks of size $N = 10000$ and $N = 50000$.

We remark that the range of the epidemic parameters, (β, μ, ρ), was chosen for two reasons. First, the authors are unaware of any diseases that have such high death or recovery rates (and hope to never see one in person). Lastly, computational and data storage restraints prevented the authors from performing the breadth of the described analysis.

2.6 Discrete Equation Fitting

Based upon our initial (unpublished) results, we believe that β (or d) is *required* to accurately determine d (or β) due to a lack of structural identifiability

(Denis-Vidal et al., 2001) (see Section 4 for details). Therefore, to more accurately determine β, we assume the network density d is *known*.

Then, if given the data for $N = 1000$ or $N = 5000$, which contains 100 network simulations, we first average each compartment (S, I, R, and D) at each time point to obtain a single vector of representative data points for each compartment ($\bar{S}, \bar{I}, \bar{R}$ and $\bar{D}$). More specifically, we take the average of each $S(1)$ to create $\bar{S}(1)$, the average of each $S(2)$ to create $\bar{S}(2)$, and so on until we create the representative vectors $\bar{S}(t), \bar{I}(t), \bar{R}(t)$ and $\bar{D}(t)$ for $t = 1, \ldots, t_n$ days where $\triangle t = 1$. In the case where a single simulation comprises all of the data (i.e., for $N = 10000$ or $N = 50000$), we use the data directly (e.g., $\bar{S}(t) = S(t)$, etc.).

We define the residue, which is the square of the distance between the discrete equation's output (the discrete **M**odel's components are denoted as S_M, I_M, R_M) and the averaged network data (denoted as $\bar{S}, \bar{I}, \bar{R}$), in the following fashion:

$$\mathrm{Res}(\beta, \mu, \rho) = \sum_{t=1}^{t_n} (S_M(t) - \bar{S}(t))^2 + \sum_{t=1}^{t_n} (I_M(t) - \bar{I}(t))^2 + \sum_{t=1}^{t_n} (R_M(t) - \bar{R}(t))^2 \tag{13}$$

To minimize the residue, we use the built-in MATLAB function `fminsearch`. Since this function is a 'local solver', it requires initial parameter guesses that are near the global minimum in order to correctly identify the best fit parameters $(\tilde{\beta}, \tilde{\mu}, \tilde{\rho})$. We denote the initial guesses for (β, μ, ρ) as $(\hat{\beta}, \hat{\mu}, \hat{\rho})$.

2.7 Estimating Initial Parameter Guesses

In practice, the initial parameter guesses $(\hat{\beta}, \hat{\mu}, \hat{\rho})$ are typically determined by hand by comparing the data to the discrete model's output and choosing parameters that create a 'good fit' as a starting point for a local minimizer (such as `fminsearch`). Due to the sheer amount of fitting required in our work, fitting each parameter by hand was simply not possible. We now describe our automated initial parameter guess calculator, which can still provide a useful starting guess in practice.

To calculate an initial guess for the recovery probability, $\hat{\rho}$, we first define ρ_t for each of $t = 1, 2, \ldots, t_n - 1$ by using the difference equation for R_M (or equivalently, for $\bar{R}$):

$$\begin{aligned} \bar{R}(t+1) &= \bar{R}(t) + \rho_t \bar{I}(t), \\ \bar{R}(t+1) - \bar{R}(t) &= \rho_t \bar{I}(t), \\ \rho_t &= \frac{\bar{R}(t+1) - \bar{R}(t)}{\bar{I}(t)}. \end{aligned}$$

Note that we cannot consider ρ_{t_n} as that requires data from day $t_n + 1$, which is outside of the time frame we consider. Additionally, notice that ρ_t is only defined when $\bar{I}(t) \neq 0$.

Let

$$t_I := \text{Maximum } t \in [1, 2, \ldots, t_n - 1] \quad \text{such that} \quad \bar{I}(t) \geq N \cdot 0.005. \tag{14}$$

With this, we see that all of $\rho_1, \rho_2, \ldots, \rho_{t_I - 1}$ are defined. We do not choose t_I to be the largest t below t_n such that $\bar{I}(t_I) \neq 0$ because this process may use averaged data over 100 simulations. Thus, to avoid potential statistical outliers (such as exactly one simulation out of 100 with the property $I(t) > 0$ for $t > 10$), we choose t_I to approximate the 'end of the epidemic'. By this, we mean to choose t_I such that $\bar{I}(t_I)$ is the latest time in the t_n day period where the number of infective individuals is above its initial value (which we set as 0.5% of the total number of nodes N).

We then choose our initial guess $\hat{\rho}$ for ρ to be the average of $\rho_1, \rho_2, \ldots, \rho_{t_I - 1}$, i.e.,

$$\hat{\rho} = \frac{1}{t_I - 1} \sum_{t=1}^{t_I - 1} \frac{\bar{R}(t+1) - \bar{R}(t)}{\bar{I}(t)}. \tag{15}$$

An analogous process is followed for determining our estimate $\hat{\mu}$ of the death probability μ, but using the difference equation for $\bar{D}(t)$ where

$$\hat{\mu} = \frac{1}{t_I - 1} \sum_{t=1}^{t_I - 1} \frac{\bar{D}(t+1) - \bar{D}(t)}{\bar{I}(t)}. \tag{16}$$

Throughout our work, we discovered $\hat{\rho}$ and $\hat{\mu}$ to be generally reasonable approximations of the minimization of ρ and μ, respectively. This is not surprising since the equations governing $R_M(t)$ and $D_M(t)$ are derived solely from first principles.

Lastly, we determine the estimate $\hat{\beta}$ for the infection probability β by using the least squares criterion to minimize the residual (Equation (13)) evaluated with initial guesses $\hat{\rho}$ and $\hat{\mu}$,

$$\hat{\beta} = \min_{\beta \in B} \text{Res}(\beta, \hat{\mu}, \hat{\rho}), \qquad \text{where} \qquad B = \{0.01, 0.02, \ldots, 1\}. \tag{17}$$

Therefore, using $(\hat{\beta}, \hat{\mu}, \hat{\rho})$ as initial parameter guesses, `fminsearch` minimizes the residue to produce the actual best fit parameters $(\tilde{\beta}, \tilde{\mu}, \tilde{\rho})$.

3 Results

To gauge the accuracy of the parameter fitting procedure, we compare the epidemic parameters from the ER network, (β, μ, ρ), to the best fit epidemic parameters from the updated discrete time model, $(\tilde{\beta}, \tilde{\mu}, \tilde{\rho})$. Our results for the networks with $N = 1000$ nodes are graphically summarized in Figures 3 – 5, while the results for $N = 5000$ node networks are in Figures 6 – 8. In particular, the correspondence of the infection probability between β and $\tilde{\beta}$ is depicted in Figures 3 and 6 for $N = 1000$ and $N = 5000$, respectively. Similarly, the correspondence of the death probability between μ and $\tilde{\mu}$ for $N = 1000$ and

$N = 5000$ are displayed in Figures 4 and 7, respectively, while Figures 5 and 8 highlight the correspondence of the recovery probability between ρ and $\tilde{\rho}$ for $N = 1000$ and $N = 5000$, respectively. Since the parameters of the ER network correspond directly to those of the discrete model, a perfect correspondence between the models should lie directly on the line $y = x$ for all parameters (β, μ, ρ) regardless of the density of the network or number of nodes.

To numerically quantify the accuracy of the parameter fitting procedure, we refer the reader to Tables S1 - S6 which contain the median and standard error of each set between the fitting procedure parameters $(\tilde{\beta}, \tilde{\mu}, \tilde{\rho})$ and epidemic parameters (β, μ, ρ) for networks of size $N = 1000$ and $N = 5000$. We remark that the median was used since it is not influenced by extreme values (unlike the mean).

Regarding the accuracy of the fitting procedure with the larger sized networks ($N = 10000$ and $N = 50000$), Table 2 displays both the percent and absolute differences between the estimated parameters $(\tilde{\beta}, \tilde{\mu}, \tilde{\rho})$ and actual parameters (β, μ, ρ). Here, "percent difference" is calculated by (estimated parameter - actual parameter)/(actual parameter) while the "absolute difference" is (estimated parameter - actual parameter).

As evidenced by Figures 3 – 8, and Tables 2 and S1 - S6, the parameter fitting procedure produces epidemic parameters $(\tilde{\beta}, \tilde{\mu}, \tilde{\rho})$ that are, overall, fairly close to the parameters (β, μ, ρ) used to simulate an epidemic on the network.

In particular, the infection rate determined from the fitting process, $\tilde{\beta}$, generally underestimated β, the actual value used to create the network epidemic simulations. This is not wholly unexpected. In Equation (7), the right-hand side of the inequality is a decreasing function of β; thus, a smaller value of β could transform the inequality into an equality.

Regarding the death and recovery rate, we can recover both μ and ρ as long as the average number of edges is sufficiently large. This is clearly seen in Figures 3 – 8, as the parameters determined through our fitting procedure approach the parameters used to create the synthetic epidemic network data, i.e., our data approach $y = x$ as the density or network size increases.

In addition, we observed that as the density d increases, the difference between the input network epidemic parameters (β, μ, ρ) and fitted parameters $(\tilde{\beta}, \tilde{\mu}, \tilde{\rho})$ decreases. This is also expected from the analytical viewpoint, as the approximation formulated from Equation (7) is exact at $d = 1$ (i.e., Equations (4) and (12)).

Finally, the time required to fit the epidemic parameters to a single data set is extremely fast, taking less than half of a second (for any parameter set, network density, or network size we explored, including the networks with $N = 10000$ or $N = 50000$ nodes).

4 Discussion

In this work, we show that for the parameter fitting procedure based upon the discrete time SIR model, the results for the modified mean-field equation

using the proposed MFA infection function and undirected and unweighted Erdös-Rényi network simulations are consistent as the density of the network increases. This agrees with the work of Zhang et al., albeit for discrete models instead of continuous; such a correspondence is depicted in Figures 3 – 8 for networks of size $N = 1000$ and $N = 5000$. Our mean-field approximation performs better as the density of the ER network increases (as seen in Figures 3 – 8, and Tables 2 and S1 – S6), though the fits found at lower density/network size can still be useful, minimally, as a rough estimate. This reinforces that Equation (7) is a reasonable approximation when the network density is high.

Although many real-world networks are often observed to exhibit highly heterogeneous degree distributions, we believe the results of this proof-of-concept study can be applied to real-world data since ER networks may be a good approximation to subgraphs of real-world networks (Taglieber and Freiberg, 2020).

Importantly, using the difference equations parameter fitting procedure to determine the network epidemic parameters is extremely cheap computationally, taking less than a second to complete. This is a stark contrast to the time required to fit epidemic parameters to a network model directly. In our work, for example, it took an average of 11.4 seconds to create one 50000 node network with density $d = 0.03$ and an average of 259.2 seconds to simulate a single epidemic on that network for $t_n = 30$ days using the techniques outlined in Section 2. Note that 259.2 seconds is the time to simulate a single epidemic on the network, not the time to estimate epidemic parameters on such a network (which likely will require many such simulations). In addition, the time required to simulate an epidemic on a network is highly dependent upon the network's size and density. Thus, the speed of fitting epidemic parameters using difference equations is orders of magnitude faster than directly using networks. Furthermore, we were unable to simulate any ER network with $N \geq 50000$ and $d > 0.3$ due to computational constraints. Therefore, if a network is extremely large and/or highly connected, then our method may provide a viable alternative.

While there are readily available data sets for an epidemic of the SIR model, the fitting procedure proposed within this work requires knowing *both* the size and the density of the network. Once such a data set is located, the authors are eager to apply their proof-of-concept study to real data. As a result, requiring the network density to be known proved to be one of the largest limitations to this work. Unfortunately, we believe that fitting both the infection probability and density simultaneously is not possible due to a lack of structurally identifiability (Denis-Vidal et al., 2001). In short, this unidentifiability would imply that it is mathematically impossible to uniquely identify a value for both β and d that yield the best fit. For a simple example of unidentifiability, consider the toy problem where $ab = 1$ provides the best fit and we must (uniquely) determine both a and b. Yet, even if this problem is unidentifiable, it may be possible to fit the network density if given the infection probability (e.g., if it was found experimentally).

To that end, as part of our future work, we believe that if we have an approximation of the infection probability (from experiments, for example), then we can approximate the density of the network. Finally, while the mean-field approximation of discrete SIR epidemics is only investigated on ER networks, we would like to explore this approximation on different network models, such as regular random or small-world networks, and on hypergraphs to develop both local and global mean-field approximations.

Acknowledgments

This work was supported by the Postdoctoral Collaborative Grant from the Department of Mathematics at the University of Arizona. The authors would like to thank Joceline Lega, Joseph C. Watkins, and Faryad Sahneh (all from the University of Arizona) for their helpful discussions surrounding this work.

Table 2 The relative "Percent Difference" is calculated by (Estimated Parameter - Actual Parameter)/(Actual Parameter) while the "Absolute Difference" is (Estimated Parameter - Actual Parameter) for networks of size N and density d. The estimated parameter values are those produced by the fitting procedure, while the actual parameter values (which are listed in Table 1) of the epidemic simulated on the network are denoted by Sets 1 – 3. All values are rounded to 3 decimal places. Note that since $\rho = 0$ for Set 2, we cannot compute the relative error for ρ for Set 2.

			Percent Difference			Absolute Difference		
N	d	Set	β	μ	ρ	β	μ	ρ
10000	0.001	1	-5.8%	14.1%	7.8%	-0.004	0.014	0.016
		2	-10.9%	5.7%	N/A	-0.01	0.011	0
		3	-15%	-4.9%	0.8%	-0.007	-0.008	0.001
10000	0.01	1	-4%	-2%	-0.7%	-0.003	-0.002	-0.002
		2	-3.6%	-0.1%	N/A	-0.003	0	0
		3	-0.7%	-0.6%	1.6%	0	-0.001	0.001
10000	0.03	1	-5.3%	4.3%	-2.4%	-0.004	0.004	-0.005
		2	-5.6%	-0.5%	N/A	-0.005	-0.001	0
		3	-4.6%	1.1%	1.7%	-0.002	0.002	0.001
10000	0.005	1	-5%	2.3%	-0.1%	-0.004	0.002	0
		2	-7%	0.2%	N/A	-0.006	0	0
		3	-0.8%	-0.2%	-3.4%	0	0	-0.002
50000	0.001	1	-5%	0.7%	0%	-0.003	0.001	0
		2	-5%	0.1%	N/A	-0.005	0	0
		3	-3.8%	-0.7%	-0.4%	-0.002	-0.001	0
50000	0.01	1	-4.1%	-0.9%	0.7%	-0.003	-0.001	0.001
		2	-3.6%	-0.2%	N/A	-0.003	0	0
		3	-2%	0.2%	-0.5%	-0.001	0	0
50000	0.03	1	-4.4%	-0.8%	0%	-0.003	-0.001	0
		2	-5%	-0.1%	N/A	-0.004	0	0
		3	-3.7%	0.4%	1.1%	-0.002	0.001	0.001
50000	0.005	1	-3.3%	-0.4%	-0.3%	-0.002	0	-0.001
		2	-4.8%	-0.1%	N/A	-0.004	0	0
		3	-2.4%	0.1%	-0.2%	-0.001	0	0

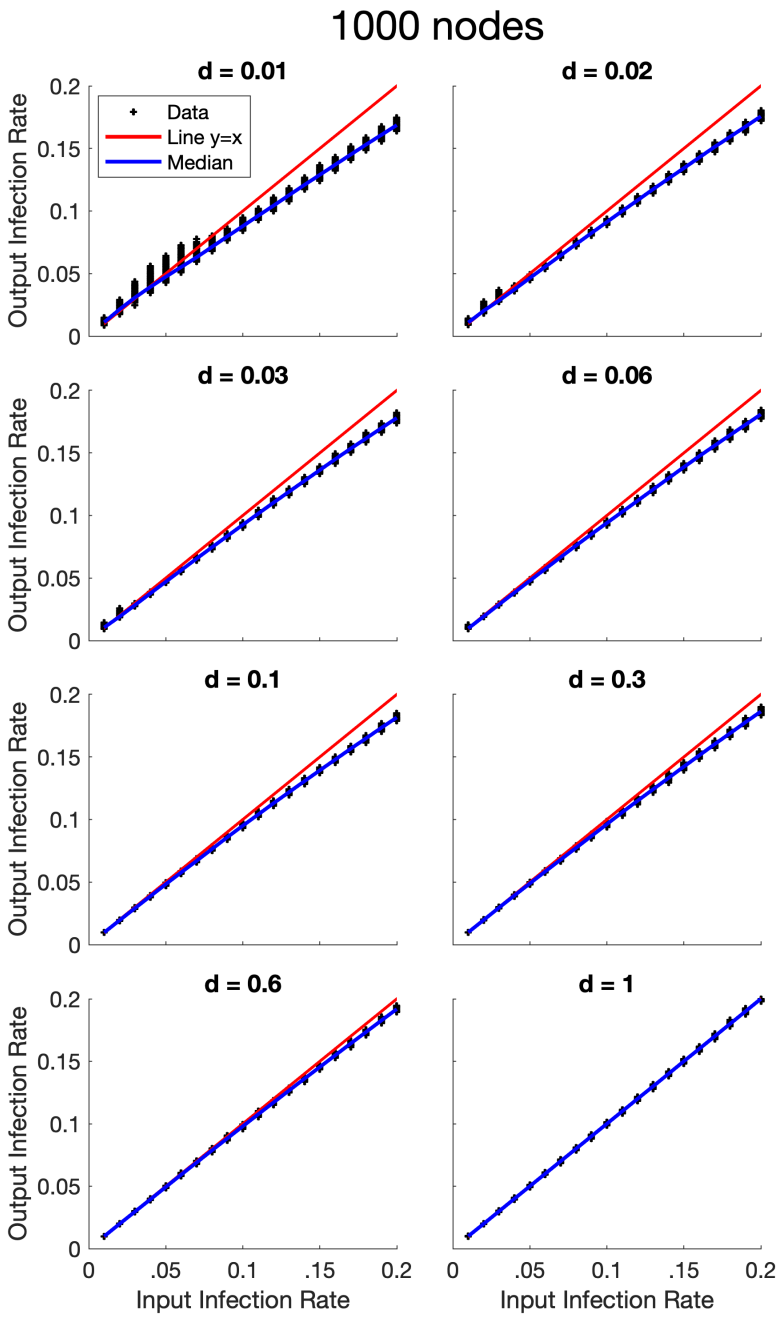

Fig. 3 The correspondence of the infection probability between the parameters used to create the network data ('input') and the parameters found through the difference equation fitting procedure ('output') is plotted in blue for various network densities with $N = 1000$ nodes.

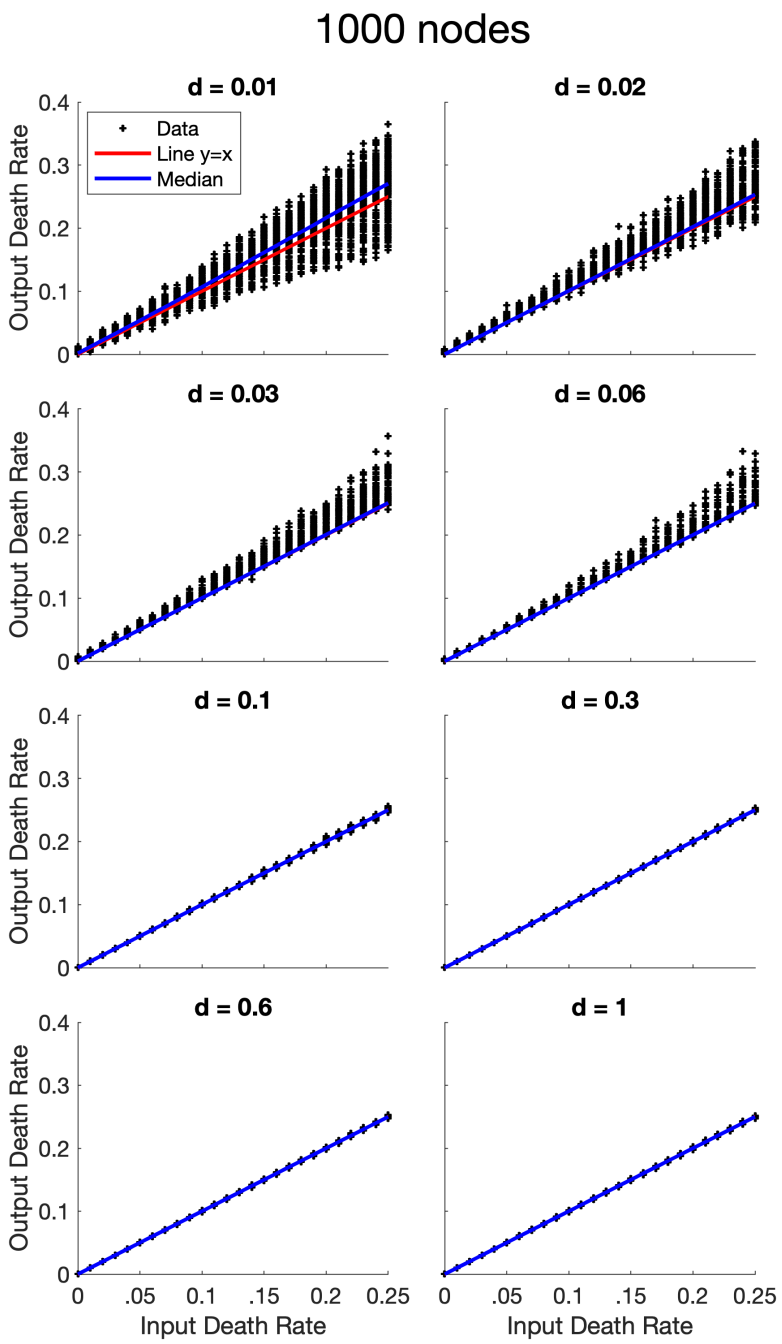

Fig. 4 The correspondence of the death probability between the parameters used to create the network data ('input') and the parameters found through the difference equation fitting procedure ('output') is plotted in blue for various network densities with $N = 1000$ nodes.

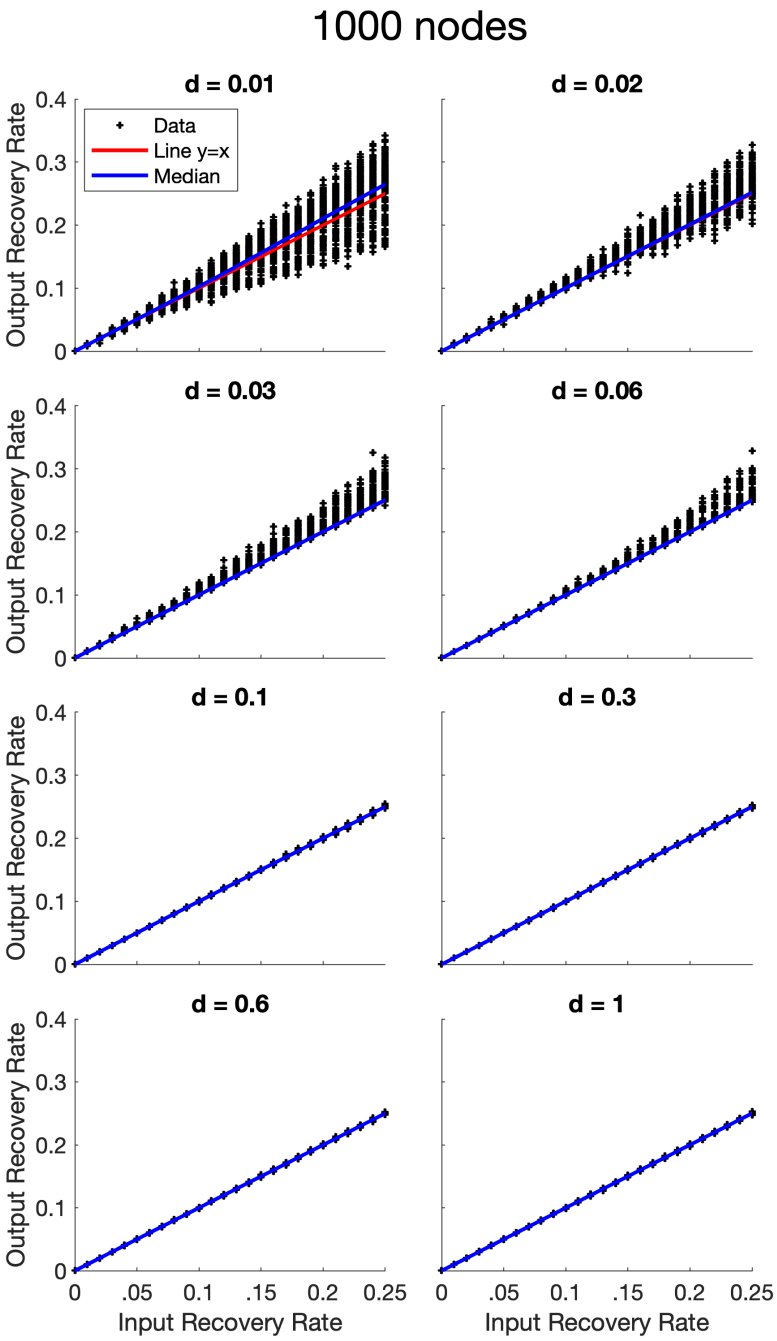

Fig. 5 The correspondence of the recovery probability between the parameters used to create the network data ('input') and the parameters found through the difference equation fitting procedure ('output') is plotted in blue for various network densities with $N = 1000$ nodes.

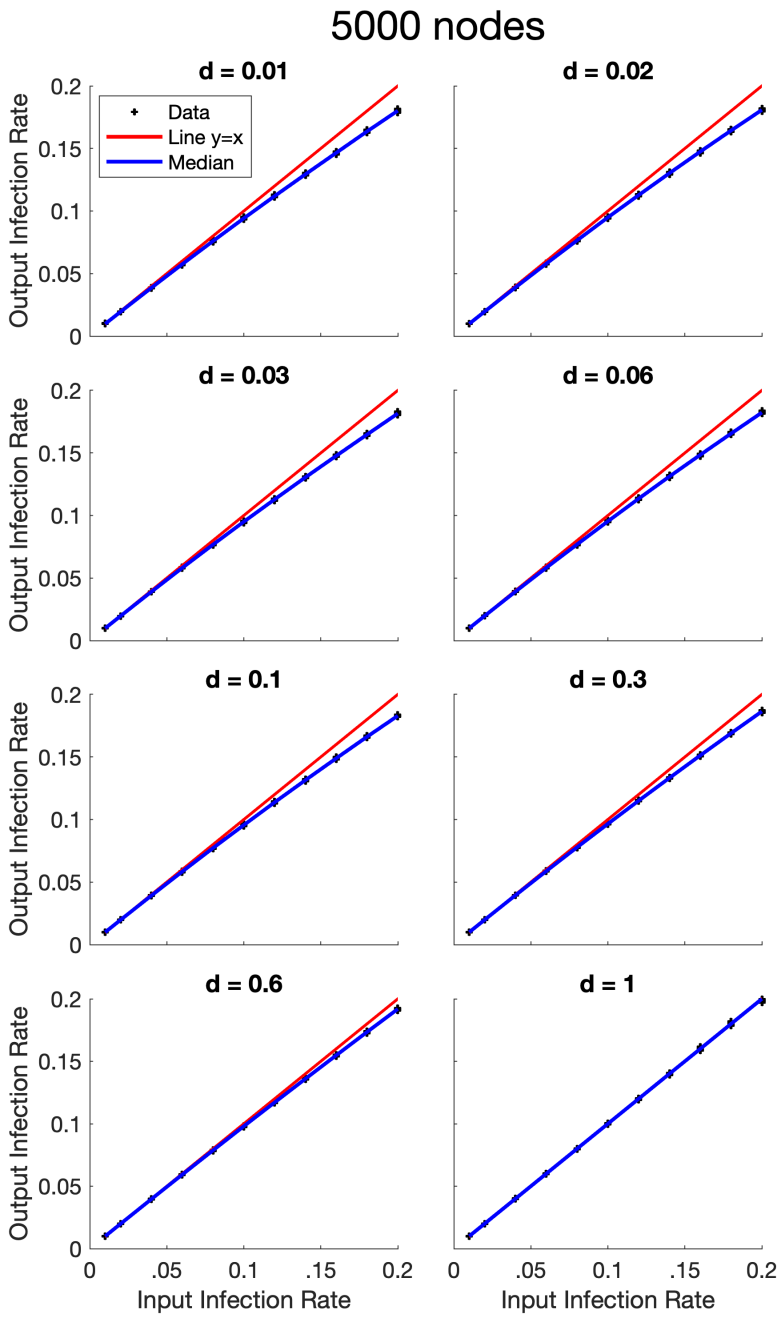

Fig. 6 The correspondence of the infection probability between the parameters used to create the network data ('input') and the parameters found through the difference equation fitting procedure ('output') is plotted in blue for various network densities with $N = 5000$ nodes.

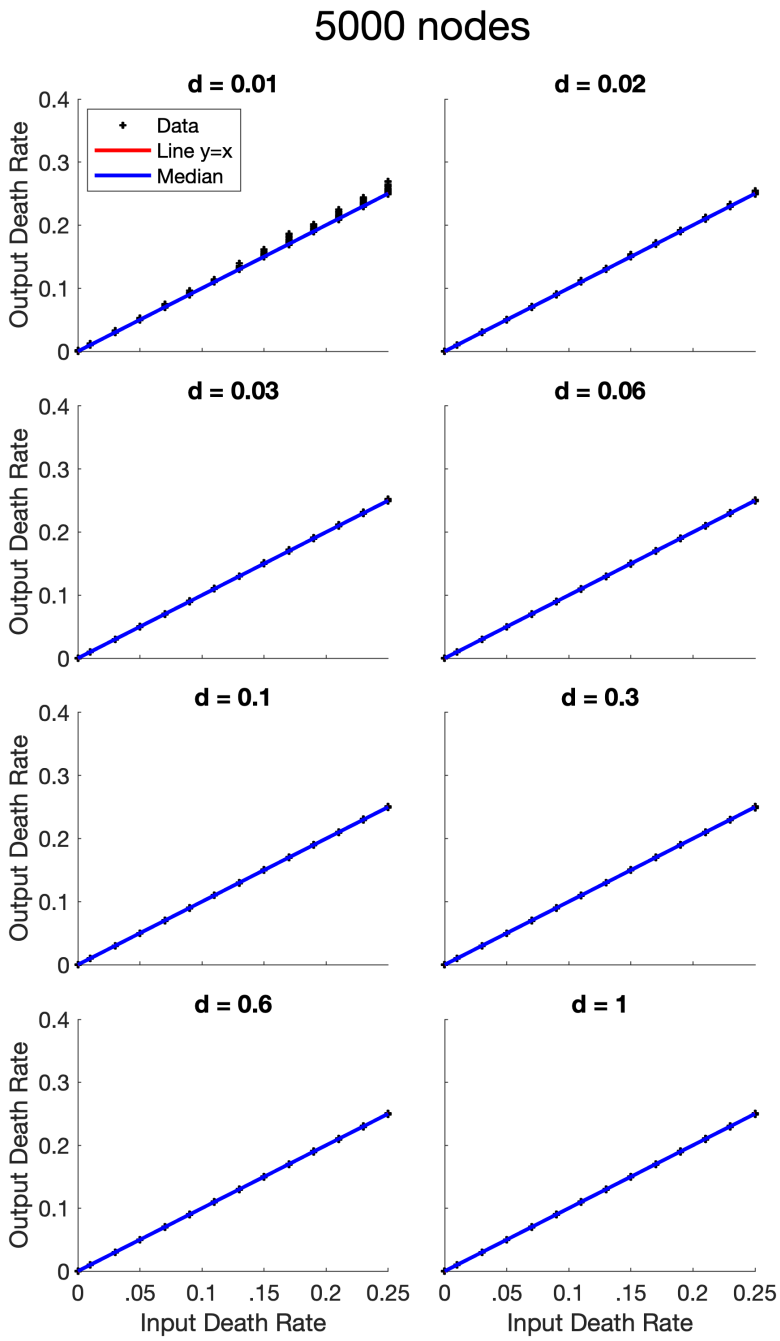

Fig. 7 The correspondence of the death probability between the parameters used to create the network data ('input') and the parameters found through the difference equation fitting procedure ('output') is plotted in blue for various network densities with $N = 5000$ nodes.

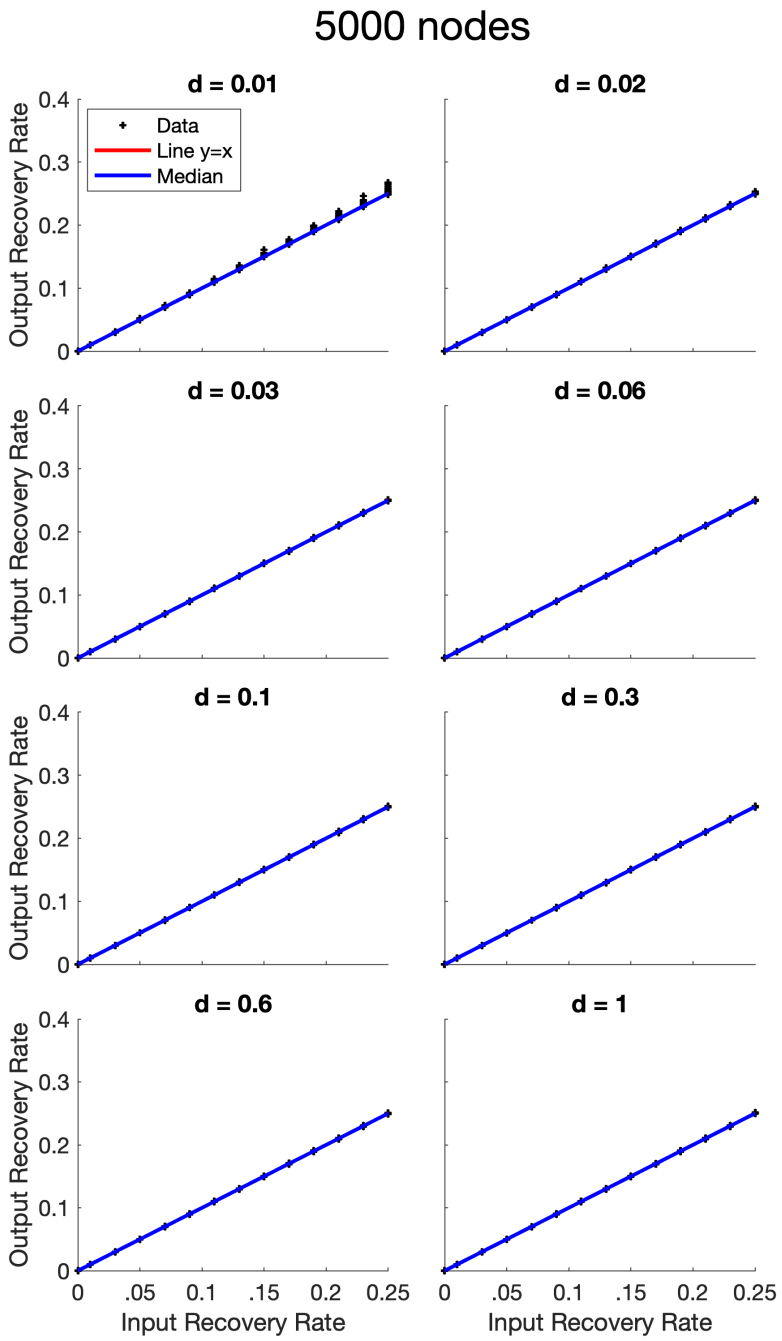

Fig. 8 The correspondence of the recovery probability between the parameters used to create the network data ('input') and the parameters found through the difference equation fitting procedure ('output') is plotted in blue for various network densities with $N = 5000$ nodes.

References

Xiang Wan, Jiming Liu, William K Cheung, and Tiejun Tong. Inferring Epidemic Network Topology from Surveillance Data. *PLoS ONE*, 9(6), 2014.

William Ogilvy Kermack and Anderson G McKendrick. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London*, 115(772):700–721, 1927.

Kai Wang, Xiao Fan Liu, and Dongchao Guo. High-Order Mean-Field Approximations for Adaptive Susceptible-Infected-Susceptible Model in Finite-Size Networks. *Complexity*, 2021, 2021.

Juliana Tolles and ThaiBinh Luong. Modeling Epidemics with Compartmental Models. *Jama*, 323(24):2515–2516, 2020.

Jiaquan Zhang, Dan Lu, and Shunkun Yang. Comparison of two mean-field based theoretical analysis methods for SIS model: Classical MF reaction equation and a modified equation. *Chaos, Solitons & Fractals*, 104:209–214, 2017.

Ottar N Bjørnstad, Katriona Shea, Martin Krzywinski, and Naomi Altman. Modeling infectious epidemics. *Nature Methods*, 17(5):455–456, 2020.

Faryad Darabi Sahneh, Caterina Scoglio, and Piet Van Mieghem. Generalized Epidemic Mean-Field Model for Spreading Processes over Multilayer Complex Networks. *IEEE/ACM Transactions on Networking*, 21(5):1609–1620, 2013.

Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87(3):925, 2015.

Piet Van Mieghem and Ruud van de Bovenkamp. Accuracy criterion for the mean-field approximation in susceptible-infected-susceptible epidemics on networks. *Physical Review E*, 91(3):032812, 2015.

Piet Van Mieghem, Jasmina Omic, and Robert Kooij. Virus Spread in Networks. *IEEE/ACM Transactions on Networking*, 17(1):1–14, 2008.

Qingchu Wu and Xinchu Fu. Immunization and epidemic threshold of an SIS model in complex networks. *Physica A: Statistical Mechanics and its Applications*, 444:576–581, 2016.

Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5(1):17–60, 1960.

Pablo M Rodríguez, Alejandro Roldán-Correa, and Leon Alexander Valencia. Comment on "Nodal infection in Markovian susceptible-infected-susceptible and susceptible-infected-removed epidemics on networks are non-negatively correlated". *Physical Review E*, 98(2):026301, 2018.

Chao-Ran Cai, Zhi-Xi Wu, Michael ZQ Chen, Petter Holme, and Jian-Yue Guan. Solving the Dynamic Correlation Problem of the Susceptible-Infected-Susceptible Model on Networks. *Physical Review Letters*, 116(25): 258301, 2016.

Johan Ludwig William Valdemar Jensen et al. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30:175–193, 1906.

Aram Vajdi, David Juher, Joan Saldana, and Caterina Scoglio. SIS Epidemics in Multilayer-based Temporal Networks. *arXiv preprint arXiv:1812.05073*, 2018.

Helen Abbey. An examination of the Reed-Frost theory of epidemics. *Human Biology*, 24(3):201, 1952.

Gabor Csardi and Tamas Nepusz. The Igraph Software Package for Complex Network Research. *InterJournal*, Complex Systems:1695, 2006.

Lilianne Denis-Vidal, Ghislaine Joly-Blanchard, and Céline Noiret. Some effective approaches to check the identifiability of uncontrolled nonlinear systems. *Mathematics and Computers in Simulation*, 57(1-2):35–44, 2001.

Klemens Taglieber and Uta Freiberg. Random graphs and their subgraphs. In *Discrete and Continuous Models in the Theory of Networks*, pages 311–340. Springer, 2020.

S5 Supplementary Materials

Table S1 The reported numbers are the median ± the standard error of the 'output' infection probability (found through the difference equation fitting procedure) for the given 'input' infection probability (used to create the network data), and density d with $N = 1000$ across all death and recovery probabilities ($26 \times 26 = 676$ simulations, see Section 2.5) To save space, all values are rounded to 2 decimal places, and 0^* is used to notate that the standard error, rounded to three decimal places, is zero.

Infection ($N = 1000$)								
Input	$d = 0.01$	$d = 0.02$	$d = 0.03$	$d = 0.06$	$d = 0.1$	$d = 0.3$	$d = 0.6$	$d = 1$
0.01	0.011±0.001	0.011±0.001	0.011±0.001	0.010±0*	0.010±0*	0.010±0*	0.010±0*	0.010±0*
0.02	0.021±0.002	0.020±0.002	0.019±0.001	0.019±0*	0.019±0*	0.020±0*	0.020±0*	0.020±0*
0.03	0.031±0.004	0.029±0.002	0.028±0*	0.029±0*	0.029±0*	0.030±0*	0.030±0*	0.030±0*
0.04	0.039±0.004	0.037±0.001	0.038±0*	0.038±0*	0.039±0*	0.039±0*	0.039±0*	0.040±0*
0.05	0.047±0.004	0.046±0.001	0.047±0*	0.048±0*	0.048±0*	0.049±0*	0.049±0*	0.050±0*
0.06	0.055±0.004	0.056±0.001	0.056±0.001	0.057±0*	0.058±0*	0.058±0*	0.059±0*	0.060±0*
0.07	0.063±0.003	0.065±0.001	0.066±0.001	0.067±0.001	0.067±0*	0.068±0*	0.069±0*	0.070±0*
0.08	0.071±0.002	0.073±0.001	0.075±0.001	0.076±0.001	0.076±0.001	0.077±0*	0.079±0.001	0.080±0*
0.09	0.080±0.002	0.082±0.001	0.084±0.001	0.085±0.001	0.086±0.001	0.087±0.001	0.088±0.001	0.090±0*
0.1	0.088±0.002	0.091±0.001	0.093±0.001	0.094±0.001	0.095±0.001	0.096±0.001	0.098±0.001	0.100±0*
0.11	0.096±0.002	0.100±0.001	0.102±0.001	0.103±0.001	0.104±0.001	0.106±0.001	0.108±0.001	0.110±0*
0.12	0.104±0.002	0.109±0.001	0.110±0.001	0.112±0.001	0.113±0.001	0.115±0.001	0.117±0.001	0.120±0.001
0.13	0.113±0.002	0.117±0.001	0.119±0.001	0.121±0.001	0.122±0.001	0.124±0.001	0.127±0.001	0.130±0.001
0.14	0.121±0.002	0.126±0.001	0.128±0.001	0.130±0.001	0.131±0.001	0.133±0.001	0.136±0.001	0.140±0.001
0.15	0.129±0.002	0.134±0.001	0.136±0.001	0.139±0.001	0.139±0.001	0.142±0.001	0.145±0.001	0.150±0.001
0.16	0.137±0.002	0.143±0.001	0.145±0.001	0.147±0.001	0.148±0.001	0.151±0.001	0.155±0.001	0.160±0.001
0.17	0.145±0.002	0.151±0.001	0.153±0.001	0.156±0.001	0.157±0.001	0.160±0.001	0.164±0.001	0.170±0.001
0.18	0.153±0.002	0.159±0.001	0.162±0.001	0.164±0.001	0.165±0.001	0.169±0.001	0.173±0.001	0.180±0.001
0.19	0.161±0.002	0.167±0.001	0.170±0.001	0.173±0.001	0.173±0.001	0.178±0.001	0.183±0.001	0.190±0.001
0.2	0.169±0.002	0.176±0.001	0.178±0.001	0.181±0.001	0.182±0.001	0.186±0.001	0.192±0.001	0.200±0.001

Table S2 The reported numbers are the median ± the standard error of the 'output' death probability (found through the difference equation fitting procedure) for the given 'input' death probability (used to create the network data), and density d with $N = 1000$ across all infection and recovery probabilities ($20 \times 26 = 520$ simulations, see Section 2.5). To save space, all values are rounded to 2 decimal places, and 0^* is used to notate that the standard error, rounded to three decimal places, is zero.

Death ($N = 1000$)								
Input	$d = 0.01$	$d = 0.02$	$d = 0.03$	$d = 0.06$	$d = 0.1$	$d = 0.3$	$d = 0.6$	$d = 1$
0	0.002±0.003	0.000±0.001	0.000±0.001	0.000±0*	0.000±0*	0.000±0*	0.000±0*	0.000±0*
0.01	0.012±0.003	0.010±0.002	0.010±0.001	0.010±0*	0.010±0*	0.010±0*	0.010±0*	0.010±0*
0.02	0.022±0.003	0.020±0.002	0.020±0.001	0.020±0*	0.020±0*	0.020±0*	0.020±0*	0.020±0*
0.03	0.033±0.004	0.030±0.002	0.030±0.001	0.030±0.001	0.030±0*	0.030±0*	0.030±0*	0.030±0*
0.04	0.043±0.004	0.040±0.002	0.040±0.001	0.040±0.001	0.040±0*	0.040±0*	0.040±0*	0.040±0*
0.05	0.054±0.005	0.050±0.003	0.050±0.002	0.050±0.001	0.050±0*	0.050±0*	0.050±0*	0.050±0*
0.06	0.064±0.006	0.060±0.003	0.060±0.002	0.060±0.001	0.060±0*	0.060±0*	0.060±0*	0.060±0*
0.07	0.075±0.006	0.070±0.003	0.070±0.002	0.070±0.001	0.070±0*	0.070±0*	0.070±0*	0.070±0*
0.08	0.085±0.006	0.081±0.003	0.080±0.003	0.080±0.001	0.080±0*	0.080±0*	0.080±0*	0.080±0*
0.09	0.096±0.007	0.091±0.004	0.090±0.003	0.090±0.001	0.090±0*	0.090±0*	0.090±0*	0.090±0*
0.1	0.107±0.008	0.101±0.004	0.100±0.003	0.100±0.002	0.100±0.001	0.100±0*	0.100±0*	0.100±0*
0.11	0.118±0.009	0.111±0.005	0.110±0.004	0.110±0.002	0.110±0.001	0.110±0*	0.110±0.001	0.110±0*
0.12	0.129±0.010	0.121±0.006	0.120±0.004	0.120±0.002	0.120±0.001	0.120±0.001	0.120±0.001	0.120±0.001
0.13	0.139±0.011	0.131±0.006	0.130±0.005	0.130±0.002	0.130±0.001	0.130±0.001	0.130±0.001	0.130±0.001
0.14	0.151±0.011	0.141±0.007	0.140±0.005	0.140±0.003	0.140±0.001	0.140±0.001	0.140±0.001	0.140±0.001
0.15	0.161±0.013	0.151±0.007	0.151±0.006	0.150±0.003	0.150±0.001	0.150±0.001	0.150±0.001	0.150±0.001
0.16	0.172±0.014	0.162±0.008	0.161±0.006	0.160±0.003	0.160±0.001	0.160±0.001	0.160±0.001	0.160±0.001
0.17	0.183±0.015	0.172±0.009	0.171±0.006	0.170±0.004	0.170±0.001	0.170±0.001	0.170±0.001	0.170±0.001
0.18	0.194±0.016	0.182±0.009	0.181±0.007	0.180±0.004	0.180±0.001	0.180±0.001	0.180±0.001	0.180±0.001
0.19	0.205±0.018	0.192±0.010	0.191±0.008	0.190±0.005	0.190±0.001	0.190±0.001	0.190±0.001	0.190±0.001
0.2	0.216±0.019	0.202±0.010	0.201±0.009	0.200±0.005	0.200±0.001	0.200±0.001	0.200±0.001	0.200±0.001
0.21	0.228±0.021	0.212±0.012	0.211±0.009	0.210±0.006	0.210±0.001	0.210±0.001	0.210±0.001	0.210±0.001
0.22	0.238±0.022	0.223±0.012	0.221±0.009	0.220±0.006	0.220±0.001	0.220±0.001	0.220±0.001	0.220±0.001
0.23	0.249±0.024	0.233±0.013	0.231±0.011	0.230±0.007	0.230±0.001	0.230±0.001	0.230±0.001	0.230±0.001
0.24	0.260±0.025	0.243±0.014	0.241±0.011	0.240±0.008	0.240±0.001	0.240±0.001	0.240±0.001	0.240±0.001
0.25	0.271±0.027	0.253±0.016	0.251±0.012	0.250±0.008	0.250±0.001	0.250±0.001	0.250±0.001	0.250±0.001

Table S3 The reported numbers are the median ± the standard error of the 'output' recovery probability (found through the difference equation fitting procedure) for the given 'input' recovery probability (used to create the network data), and density d with $N = 1000$ across all infection and death probabilities ($20 \times 26 = 520$ simulations, see Section 2.5). To save space, all values are rounded to 2 decimal places, and 0^* is used to notate that the standard error, rounded to three decimal places, is zero.

Recovery ($N = 1000$)								
Input	$d = 0.01$	$d = 0.02$	$d = 0.03$	$d = 0.06$	$d = 0.1$	$d = 0.3$	$d = 0.6$	$d = 1$
0	0.000±0*	0.000±0*	0.000±0*	0.000±0*	0.000±0*	0.000±0*	0.000±0*	0.000±0*
0.01	0.010±0*	0.010±0*	0.010±0*	0.010±0*	0.010±0*	0.010±0*	0.010±0*	0.010±0*
0.02	0.020±0.001	0.020±0.001	0.020±0*	0.020±0*	0.020±0*	0.020±0*	0.020±0*	0.020±0*
0.03	0.030±0.001	0.030±0.001	0.030±0.001	0.030±0*	0.030±0*	0.030±0*	0.030±0*	0.030±0*
0.04	0.040±0.002	0.040±0.001	0.040±0.001	0.040±0*	0.040±0*	0.040±0*	0.040±0*	0.040±0*
0.05	0.051±0.002	0.050±0.001	0.050±0.001	0.050±0*	0.050±0*	0.050±0*	0.050±0*	0.050±0*
0.06	0.061±0.003	0.060±0.001	0.060±0.001	0.060±0*	0.060±0*	0.060±0*	0.060±0*	0.060±0*
0.07	0.071±0.003	0.070±0.002	0.070±0.001	0.070±0*	0.070±0*	0.070±0*	0.070±0*	0.070±0*
0.08	0.082±0.004	0.080±0.002	0.080±0.001	0.080±0.001	0.080±0*	0.080±0*	0.080±0*	0.080±0*
0.09	0.092±0.005	0.090±0.002	0.090±0.002	0.090±0.001	0.090±0*	0.090±0*	0.090±0*	0.090±0*
0.1	0.103±0.005	0.100±0.003	0.100±0.002	0.100±0.001	0.100±0*	0.100±0*	0.100±0*	0.100±0*
0.11	0.113±0.006	0.110±0.003	0.110±0.002	0.110±0.001	0.110±0*	0.110±0.001	0.110±0*	0.110±0*
0.12	0.123±0.007	0.121±0.004	0.120±0.003	0.120±0.001	0.120±0.001	0.120±0.001	0.120±0.001	0.120±0.001
0.13	0.134±0.008	0.131±0.004	0.130±0.003	0.130±0.001	0.130±0.001	0.130±0.001	0.130±0.001	0.130±0.001
0.14	0.145±0.009	0.141±0.004	0.140±0.004	0.140±0.001	0.140±0.001	0.140±0.001	0.140±0.001	0.140±0.001
0.15	0.156±0.011	0.151±0.005	0.150±0.004	0.150±0.002	0.150±0.001	0.150±0.001	0.150±0.001	0.150±0.001
0.16	0.167±0.012	0.161±0.006	0.160±0.005	0.160±0.002	0.160±0.001	0.160±0.001	0.160±0.001	0.160±0.001
0.17	0.178±0.013	0.171±0.006	0.170±0.005	0.170±0.002	0.170±0.001	0.170±0.001	0.170±0.001	0.170±0.001
0.18	0.188±0.014	0.181±0.007	0.180±0.006	0.180±0.003	0.180±0.001	0.180±0.001	0.180±0.001	0.180±0.001
0.19	0.199±0.015	0.191±0.008	0.190±0.005	0.190±0.003	0.190±0.001	0.190±0.001	0.190±0.001	0.190±0.001
0.2	0.211±0.017	0.201±0.009	0.201±0.007	0.200±0.004	0.200±0.001	0.200±0.001	0.200±0.001	0.200±0.001
0.21	0.221±0.018	0.211±0.009	0.211±0.008	0.210±0.004	0.210±0.001	0.210±0.001	0.210±0.001	0.210±0.001
0.22	0.232±0.020	0.221±0.011	0.221±0.008	0.220±0.005	0.220±0.001	0.220±0.001	0.220±0.001	0.220±0.001
0.23	0.243±0.021	0.232±0.011	0.231±0.008	0.230±0.006	0.230±0.001	0.230±0.001	0.230±0.001	0.230±0.001
0.24	0.254±0.023	0.241±0.012	0.241±0.010	0.240±0.006	0.240±0.001	0.240±0.001	0.240±0.001	0.240±0.001
0.25	0.265±0.024	0.252±0.014	0.251±0.009	0.250±0.007	0.250±0.001	0.250±0.001	0.250±0.001	0.250±0.001

Table S4 The reported numbers are the median ± the standard error of the 'output' infection probability (found through the difference equation fitting procedure) for the given 'input' infection probability (used to create the network data), and density d with $N = 5000$ across all death and recovery probabilities ($14 \times 14 = 196$ simulations, see Section 2.5). To save space, all values are rounded to 2 decimal places, and 0^* is used to notate that the standard error, rounded to three decimal places, is zero.

Infection ($N = 5000$)								
Input	$d = 0.01$	$d = 0.02$	$d = 0.03$	$d = 0.06$	$d = 0.1$	$d = 0.3$	$d = 0.6$	$d = 1$
0.01	0.010±0*	0.010±0*	0.010±0*	0.010±0*	0.010±0*	0.010±0*	0.010±0*	0.010±0*
0.02	0.019±0*	0.020±0*	0.020±0*	0.020±0*	0.020±0*	0.020±0*	0.020±0*	0.020±0*
0.04	0.039±0*	0.039±0*	0.039±0*	0.039±0*	0.039±0*	0.039±0*	0.040±0*	0.040±0*
0.06	0.057±0*	0.058±0*	0.058±0*	0.058±0*	0.058±0*	0.059±0*	0.059±0*	0.060±0*
0.08	0.076±0*	0.077±0*	0.077±0*	0.077±0*	0.077±0*	0.078±0*	0.079±0*	0.080±0*
0.1	0.094±0*	0.095±0*	0.095±0*	0.095±0*	0.096±0*	0.097±0*	0.098±0*	0.100±0*
0.12	0.112±0*	0.113±0*	0.113±0*	0.113±0*	0.114±0*	0.115±0*	0.117±0*	0.120±0*
0.14	0.129±0*	0.130±0*	0.131±0*	0.131±0*	0.131±0*	0.133±0*	0.136±0*	0.140±0*
0.16	0.147±0*	0.147±0*	0.148±0*	0.148±0*	0.149±0*	0.151±0*	0.155±0*	0.160±0.001
0.18	0.164±0.001	0.164±0*	0.165±0*	0.165±0*	0.166±0*	0.169±0*	0.173±0*	0.180±0.001
0.2	0.180±0.001	0.181±0*	0.182±0*	0.182±0*	0.183±0*	0.186±0*	0.192±0*	0.200±0.001

Table S5 The reported numbers are the median ± the standard error of the 'output' death probability (found through the difference equation fitting procedure) for the given 'input' death probability (used to create the network data), and density d with $N = 5000$ across all infection and recovery probabilities ($11 \times 14 = 154$ simulations, see Section 2.5). To save space, all values are rounded to 2 decimal places, and 0^* is used to notate that the standard error, rounded to three decimal places, is zero.

Death ($N = 5000$)								
Input	$d = 0.01$	$d = 0.02$	$d = 0.03$	$d = 0.06$	$d = 0.1$	$d = 0.3$	$d = 0.6$	$d = 1$
0	0.000±0*	0.000±0*	0.000±0*	0.000±0*	0.000±0*	0.000±0*	0.000±0*	0.000±0*
0.01	0.010±0*	0.010±0*	0.010±0*	0.010±0*	0.010±0*	0.010±0*	0.010±0*	0.010±0*
0.03	0.030±0.001	0.030±0*	0.030±0*	0.030±0*	0.030±0*	0.030±0*	0.030±0*	0.030±0*
0.05	0.050±0.001	0.050±0*	0.050±0*	0.050±0*	0.050±0*	0.050±0*	0.050±0*	0.050±0*
0.07	0.070±0.001	0.070±0*	0.070±0*	0.070±0*	0.070±0*	0.070±0*	0.070±0*	0.070±0*
0.09	0.090±0.001	0.090±0*	0.090±0*	0.090±0*	0.090±0*	0.090±0*	0.090±0*	0.090±0*
0.11	0.110±0.001	0.110±0*	0.110±0*	0.110±0*	0.110±0*	0.110±0*	0.110±0*	0.110±0*
0.13	0.130±0.001	0.130±0*	0.130±0*	0.130±0*	0.130±0*	0.130±0*	0.130±0*	0.130±0*
0.15	0.150±0.002	0.150±0.001	0.150±0*	0.150±0*	0.150±0*	0.150±0*	0.150±0*	0.150±0*
0.17	0.170±0.003	0.170±0*	0.170±0*	0.170±0*	0.170±0*	0.170±0*	0.170±0*	0.170±0*
0.19	0.190±0.002	0.190±0.001	0.190±0*	0.190±0*	0.190±0*	0.190±0*	0.190±0*	0.190±0*
0.21	0.210±0.002	0.210±0.001	0.210±0*	0.210±0*	0.210±0*	0.210±0*	0.210±0*	0.210±0*
0.23	0.230±0.002	0.230±0.001	0.230±0*	0.230±0*	0.230±0*	0.230±0*	0.230±0*	0.230±0*
0.25	0.250±0.003	0.250±0.001	0.250±0.001	0.250±0*	0.250±0*	0.250±0*	0.250±0*	0.250±0*

Table S6 The reported numbers are the median ± the standard error of the 'output' recovery probability (found through the difference equation fitting procedure) for the given 'input' recovery probability (used to create the network data), and density d with $N = 5000$ across all infection and death probabilities ($11 \times 14 = 154$ simulations, see Section 2.5). To save space, all values are rounded to 2 decimal places, and 0^* is used to notate that the standard error, rounded to three decimal places, is zero.

Recovery ($N = 5000$)								
Input	$d = 0.01$	$d = 0.02$	$d = 0.03$	$d = 0.06$	$d = 0.1$	$d = 0.3$	$d = 0.6$	$d = 1$
0	0.000±0*	0.000±0*	0.000±0*	0.000±0*	0.000±0*	0.000±0*	0.000±0*	0.000±0*
0.01	0.010±0*	0.010±0*	0.010±0*	0.010±0*	0.010±0*	0.010±0*	0.010±0*	0.010±0*
0.03	0.030±0*	0.030±0*	0.030±0*	0.030±0*	0.030±0*	0.030±0*	0.030±0*	0.030±0*
0.05	0.050±0*	0.050±0*	0.050±0*	0.050±0*	0.050±0*	0.050±0*	0.050±0*	0.050±0*
0.07	0.070±0*	0.070±0*	0.070±0*	0.070±0*	0.070±0*	0.070±0*	0.070±0*	0.070±0*
0.09	0.090±0*	0.090±0*	0.090±0*	0.090±0*	0.090±0*	0.090±0*	0.090±0*	0.090±0*
0.11	0.110±0.001	0.110±0*	0.110±0*	0.110±0*	0.110±0*	0.110±0*	0.110±0*	0.110±0*
0.13	0.130±0.001	0.130±0*	0.130±0*	0.130±0*	0.130±0*	0.130±0*	0.130±0*	0.130±0*
0.15	0.150±0.001	0.150±0*	0.150±0*	0.150±0*	0.150±0*	0.150±0*	0.150±0*	0.150±0*
0.17	0.170±0.001	0.170±0*	0.170±0*	0.170±0*	0.170±0*	0.170±0*	0.170±0*	0.170±0*
0.19	0.190±0.002	0.190±0*	0.190±0*	0.190±0*	0.190±0*	0.190±0*	0.190±0*	0.190±0*
0.21	0.210±0.002	0.210±0*	0.210±0*	0.210±0*	0.210±0*	0.210±0*	0.210±0*	0.210±0*
0.23	0.230±0.002	0.230±0.001	0.230±0*	0.230±0*	0.230±0*	0.230±0*	0.230±0*	0.230±0*
0.25	0.250±0.003	0.250±0.001	0.250±0*	0.250±0*	0.250±0*	0.250±0*	0.250±0*	0.250±0*

Table S7 The mean and standard deviation of the number of edges m, density d, average degree $<k>$, the clustering coefficient $<cc>$, and the *diameter* for each network of size $N = 1000$ calculated from 100 simulated networks. To save space, all values are rounded to 4 decimal places, and 0^* is used to notate that the standard deviation, rounded to four decimal places, is zero.

$N = 1000$				
m	d	$<k>$	$<cc>$	*diameter*
4995 ± 0	0.0100 ± 0	9.9900 ± 0^*	0.0099 ± 0.0008	5.4300 ± 0.4976
9990 ± 0	0.0200 ± 0	19.9800 ± 0^*	0.0200 ± 0.0005	4 ± 0
14985 ± 0	0.0300 ± 0	29.9700 ± 0	0.0300 ± 0.0005	3.0800 ± 0.2727
29970 ± 0	0.0600 ± 0	59.9400 ± 0	0.0600 ± 0.0003	3 ± 0
49950 ± 0	0.1000 ± 0	99.9000 ± 0	0.1000 ± 0.0002	3 ± 0
149850 ± 0	0.3000 ± 0	299.7000 ± 0	0.3000 ± 0.0001	2 ± 0
299700 ± 0	0.6000 ± 0	599.4000 ± 0	0.6000 ± 0^*	2 ± 0
499500 ± 0	1 ± 0	999 ± 0	1 ± 0	1 ± 0

Table S8 The mean and standard deviation of the number of edges m, density d, average degree $<k>$, the clustering coefficient $<cc>$, and the *diameter* for each network of size $N = 5000$ calculated from 100 simulated networks. To save space, all values are rounded to 4 decimal places, and 0^* is used to notate that the standard deviation, rounded to four decimal places, is zero.

$N = 5000$				
m	d	$<k>$	$<cc>$	*diameter*
124975 ± 0	0.0100 ± 0	49.9900 ± 0	0.0100 ± 0.0001	3.2700 ± 0.4462
249950 ± 0	0.0200 ± 0	99.9800 ± 0^*	0.0200 ± 0^*	3 ± 0
374925 ± 0	0.0300 ± 0	149.9700 ± 0^*	0.0300 ± 0^*	3 ± 0
749850 ± 0	0.0600 ± 0	299.9400 ± 0^*	0.0600 ± 0^*	2.1100 ± 0.3145
1249750 ± 0	0.1000 ± 0	499.9000 ± 0^*	0.1000 ± 0^*	2 ± 0
3749250 ± 0	0.3000 ± 0	1499.7000 ± 0	0.3000 ± 0^*	2 ± 0
7498500 ± 0	0.6000 ± 0	2999.4000 ± 0	0.6000 ± 0^*	2 ± 0
12497500 ± 0	1 ± 0	4999 ± 0	1 ± 0	1 ± 0

Table S9 The basic structural features for each simulated network of size N with m edges and density d.

N	m	d	$<k>$	$<cc>$	*diameter*
10000	49995	0.0010	9.9990	0.0009	7
10000	249975	0.0050	49.9950	0.0050	4
10000	499950	0.0100	99.9900	0.0101	3
10000	1499850	0.0300	299.9700	0.0300	3
50000	1249975	0.0010	49.9990	0.0010	4
50000	6249875	0.0050	249.9950	0.0050	3
50000	12499750	0.0100	499.9900	0.0100	3
50000	37499250	0.0300	1499.9700	0.0300	2

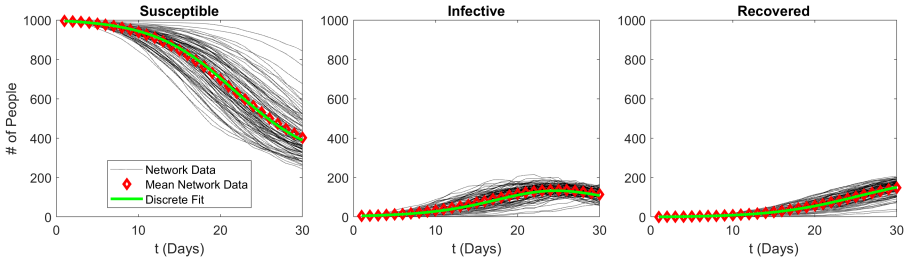

Fig. S1 An example of how the discrete model's predictions align with the simulated network data (100 simulations, one per network). Each ER network contained $N = 1000$ nodes, had density $d = 0.01$, and network epidemic parameters $(\beta, \mu, \rho) = (0.05, 0.16, 0.07)$.

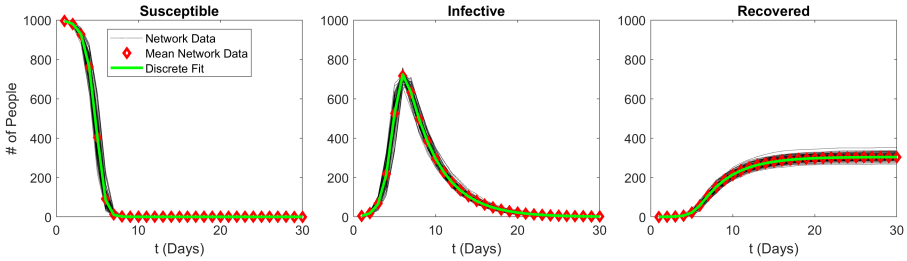

Fig. S2 An example of how the discrete model's predictions align with the simulated network data (100 simulations, one per network). Each ER network contained $N = 1000$ nodes, had density $d = 0.06$, and network epidemic parameters $(\beta, \mu, \rho) = (0.05, 0.16, 0.07)$.