Network Science **7** (4): 556–580, 2019
doi:10.1017/nws.2019.54

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

An SIR epidemic on a weighted network

Kristoffer Spricer* and Tom Britton

Department of Mathematics, Stockholm University, 106 91 Stockholm, Sweden
*Corresponding author. Email: kspricer@hotmail.com

Action Editor: Stanley Wasserman

Abstract

We introduce a weighted configuration model graph, where *edge weights* correspond to the probability of infection in an epidemic on the graph. On these graphs, we study the development of a Susceptible–Infectious–Recovered epidemic using both Reed–Frost and Markovian settings. For the special case of having two different edge types, we determine the *basic reproduction number* R_0, the *probability of a major outbreak*, and the *relative final size of a major outbreak*. Results are compared with those for a calibrated unweighted graph. The degree distributions are based on both theoretical constructs and empirical network data. In addition, bivariate standard normal copulas are used to model the dependence between the degrees of the two edge types, allowing for modeling the correlation between edge types over a wide range. Among the results are that the weighted graph produces much richer results than the unweighted graph. Also, while R_0 always increases with increasing correlation between the two degrees, this is not necessarily true for the probability of a major outbreak nor for the relative final size of a major outbreak. When using copulas we see that these can produce results that are similar to those of the empirical degree distributions, indicating that in some cases a copula is a viable alternative to using the full empirical data.

Keywords: epidemics; basic reproduction number; weighted graph; configuration model; final size; major outbreak; copula

1. Introduction

The configuration model is a well-known graph model where each vertex is assigned a number of half-edges which are then connected uniformly at random (Molloy & Reed, 1995; Bollobás, 2001). This model is often used when a specific degree distribution or degree sequence is desired. The development of epidemics can then be studied on these graphs. Three important quantities derived from such epidemics are the *basic reproduction number* R_0, the *probability of a major outbreak*, and the *relative final size of a major outbreak*, see, e.g., Britton (2010). Often all edges are treated as identical and for instance the transmission risk (of the infection) is assumed to be the same for all edges, which is not true for many real world networks. Therefore, models where vertices or edges are of different types have been extensively studied. In models where vertices are of different types, these models typically assume different transmission rates for edges going to other vertices of the same type, compared to edges going between vertices of different types. In Ball & Neal (2002), various models with two-level mixing (local and global contacts) including household models are explored. Susceptibility sets are used in the analysis (as in our paper), but the mixing is homogeneous (compared to the configuration model in our paper). The configuration model is used in Ball & Sirl (2012), but focus is on individuals of different types. In Deijfen & Fitzner (2017), the population is split into two groups also allowing connections between the two groups (so that nodes are of two different types) with degree distributions that depend on the type, using (among other) a configuration model approach to connect edges within groups and

between groups. Another approach with two groups is Shu et al. (2012) with two communities with weak ties (edges) between the two communities, allowing also for scale-free networks (heavy-tailed degree distributions). From a slightly different field, we have Larson (2017) which studies diffusion of information on weak and strong ties (edges) within a population, including empirical examples. Finally Grassberger (1983) also considers the spread of an epidemic on a network, both the Reed–Frost and the Markovian versions, but there the network is the d-dimensional infinite lattice and focus is on behavior near critical.

In Britton et al. (2011) a degree is assigned to each vertex, and for each half-edge a weight is assigned independently from a weight distribution that is only allowed to depend on the degree of the vertex. Half-edges of identical weight are then connected to each other as in the configuration model. In Kamp et al. (2013), vertices are assigned a degree and a number of *interactions* from a simultaneous distribution. Each interaction is then distributed independently and uniformly among all edges of the vertex. The number of interactions constitute the weight of the edge. Half-edges having the same (or similar) weight are then connected according to the configuration model. Both of these models place restrictions on the allowed degree distributions. Another recent model that discusses allowing edges to have different transmission rates is Miller & Volz (2013), see Section 2.2.4.

One possible generalization is a model with an arbitrary number of edge types, each with its own weight, allowing for an arbitrary dependence between the degrees of the different edge types. Each vertex is assigned a multivariate degree which specifies how many half-edges of each type it has. The degree can be assigned from a given degree sequence (such as from an empirical graph) or from a given degree distribution. Half-edges *of the same type* are then paired just as in the normal configuration model to create a multilayer configuration model, where two layers are only connected at vertices which have edges of both types. In this paper we study this model, but for simplicity restrict it to two types of edges and thus two weights, and the development of Susceptible–Infectious–Recovered (SIR) epidemics (explained in Section 2.4) on it. Even then it is possible to study some interesting configurations—e.g., we can assume that each person has two different types of contacts with other people and that the probability of infection differs on these types. Examples of such situations are family relations vs job relations, or casual vs more permanent sexual relationships. The theoretical foundations for the model, including the graph model and the epidemic model, are presented in Section 2.

Important questions are if the quantities of interest (described above) differ on the weighted graph versus a calibrated unweighted graph, and also how these quantities are affected by the level of correlation between the degrees of the different types of edges. We construct the unweighted graph from the weighted model by ignoring edge type and create a single configuration model network from all edges, but without doing any other changes to the degree distribution. We calibrate the weighted and unweighted graphs such that both have the same *mean infectious activity* (see Section 2.5). We then compare the development of SIR epidemics on both graphs. For the weighted graph we also vary the correlation between the two degrees of an individual and study the effect on the quantities of interest. We see that the weighted graph produces much richer results, in that all the studied parameters typically show much more variation over the allowed range of the parameters that we can vary, compared with the unweighted graph. Results also indicate that a model, where the dependence structure between the two degrees has been defined through a standard normal copula (see Section 4.1 and Appendix B.1), often works equally well as a model where the dependence structure is taken from the empirical degree distribution. The copula model allows for varying the correlation through a wide range that is only limited by the marginal degree distributions (for the different edge types). Some theoretical results are given in Section 3, while numerical results are shown in Section 4. A discussion can be found in Section 5.

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

2. The model

In this section, we first briefly discuss graphs (Section 2.1) and then present the unweighted (Section 2.2) and the weighted (Section 2.3) configuration models. For the weighted configuration model we limit the discussion to graphs with two edge types. While the results for the unweighted configuration model are generally known, they are included here for direct comparison with the weighted configuration model. In Section 2.4, we define SIR epidemics in both Reed–Frost and Markovian settings on configuration model graphs, as an introduction to the theoretical results obtained in Section 3. In order to be able to compare SIR epidemics on weighted and unweighted graphs, we need to calibrate the different graphs properly. We do this by defining the *mean infectious activity* (Section 2.5) and set it to be the same for weighted and unweighted graphs.

2.1 Graphs

In this paper we use the words *graph* and *network* interchangeably. The number of vertices n is given and (typically) the case $n \to \infty$ is studied, although we do not always mention this explicitly. In the context of epidemics on graphs, vertices represent people and edges represent some type of relationship making transmission of an infection possible. We work with undirected graphs, so if two vertices are connected by an edge they can both infect each other. Edges can be of different type and with different properties. We use ξ as an index to indicate the edge type, whenever appropriate. In this paper we limit the analysis to two edge types, so $\xi \in \{1, 2\}$.

The degrees of all vertices in the graph is called a *degree sequence*. Graphs and degree sequences can, e.g., be obtained from empirical data or from more theoretical constructs. Graphs that are created by random processes are called random graphs. One such random graph model is the configuration model which is discussed in Sections 2.2 and 2.3.

2.2 Unweighted configuration model

The configuration model has already been thoroughly investigated (see, e.g., Molloy & Reed, 1995 or Britton et al., 2006), and here we just briefly recapitulate how a configuration model graph is created and some properties of it.

A configuration model graph is always finite, having n vertices, but asymptotic results are obtained by letting $n \to \infty$. Initially each vertex is assigned a number of yet unconnected half-edges that can, e.g., be drawn from some given degree distribution, D. Then half-edges are paired uniformly at random. Parallel edges (several edges going between the same vertices) and self loops (edges with both ends going to the same vertex) can occur, but with suitable restrictions on the degree sequence or the degree distribution the number of such edges is small compared with the total number of edges in the graph and thus (asymptotically) do not affect the properties of the graph. A finite first moment is needed in order for the configuration model to converge in distribution as $n \to \infty$, and a finite second moment is needed in order to obtain a finite first moment for the *size-biased* distribution (see below).

The degree distribution is defined by

$$p_i = \mathbb{P}(D = i)$$

This is the degree of a vertex that is chosen uniformly at random from the graph. The important properties of the graph that we return to later in this paper are the mean and the variance:

$$\mu = \mathrm{E}(D),$$
$$\sigma^2 = \mathrm{Var}(D)$$

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

If a vertex is instead chosen by first selecting an edge uniformly at random and then selecting one of the vertices connected to the edge with the same probability, we obtain the *size-biased* distribution $\widetilde{D}$. The size-biased distribution has the following (asymptotic) properties:

$$\widetilde{p}_i = \mathbb{P}(\widetilde{D} = i) = \frac{ip_i}{\mu}, \tag{1}$$

$$\widetilde{\mu} = \mathrm{E}(\widetilde{D}) = \frac{\mathrm{E}(D^2)}{\mu} = \mu + \frac{\sigma^2}{\mu} \tag{2}$$

2.3 Weighted configuration model

In the weighted graph we have two types of edges (labeled 1 and 2 in this paper). Starting with a given number of vertices, each vertex is assigned a number of half-edges drawn independently for each vertex from a given degree distribution $\mathbf{D} = (D_1, D_2)$. Half-edges *of the same type* are then connected uniformly at random, effectively creating two configuration model graphs that are connected only at vertices that have both types of edges.

The properties of this graph are given by the degree distribution $\mathbf{D} = (D_1, D_2)$. The distribution is defined by the probabilities

$$p_{ij} = \mathbb{P}(D_1 = i, D_2 = j)$$

We only place a minimum set of requirements on this distribution. First, we require that at least one $p_{ij} > 0$ for some $i, j > 0$. Otherwise we effectively have two different vertex categories, one with only type 1 edges and another with only type 2 edges, and these never interact—creating two separate configuration models. Second, we require that the first and second moments are finite, so that $\mathrm{E}(D_\xi) < \infty$ and $\mathrm{Var}(D_\xi) < \infty$, where $\xi \in \{1, 2\}$ indicates the edge type. This ensures that the parallel edges and self loops can be ignored in the resulting configuration model graphs (Britton et al., 2006), and that the first moment of the size-biased distribution is finite (just as for the unweighted configuration model).

When studying this distribution the following definitions are useful:

$$\begin{aligned}
\mu_\xi &= \mathrm{E}(D_\xi), \\
\sigma_\xi^2 &= \mathrm{Var}(D_\xi), \\
\sigma_{12} &= \mathrm{Cov}(D_1, D_2), \\
\rho &= \frac{\sigma_{12}}{\sigma_1 \sigma_2}, \text{ if } \sigma_1, \sigma_2 > 0
\end{aligned}$$

where the last one is the correlation coefficient between D_1 and D_2.

The degree of a vertex selected uniformly at random from the graph is distributed according to $\mathbf{D}$. If we instead select a vertex by following an *edge* of specified type, selected uniformly at random, the resulting degree distribution is different and also depends on the type of the edge that we follow. Given that we follow a uniformly selected edge of type $\xi \in \{1, 2\}$, the *size-biased* degree distribution is $\widetilde{\mathbf{D}}_\xi = (\widetilde{D}_{1|\xi}, \widetilde{D}_{2|\xi})$. The tilde above the symbols indicates quantities obtained from the size-biased distribution. The probability mass function $\widetilde{p}_\xi(i, j) = \mathbb{P}(\widetilde{D}_{1|\xi} = i, \widetilde{D}_{2|\xi} = j)$ distribution is then

$$\widetilde{p}_1(i, j) = \frac{ip_{ij}}{\mu_1}, \tag{3}$$

$$\widetilde{p}_2(i, j) = \frac{jp_{ij}}{\mu_2} \tag{4}$$

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

when following an edge of type 1 and 2, respectively. Equation (3) can be understood intuitively by realizing that when following an edge selected uniformly at random, the probability of connecting to a vertex with degree i is proportional to i (thus the name *size-biased* distribution). This probability must also be proportional to the relative occurrence of vertices with this degree (quantified by p_{ij}). Finally the $1/\mu_1$ is a norming constant needed to make $\widetilde{p}_1(i,j)$ a proper probability mass function. Equation (4) can be derived in the same way.

When following an edge of type 1, we now obtain (using Equations (3) and (4))

$$\widetilde{\mu}_{1|1} = \mathrm{E}(\widetilde{D}_{1|1}) = \frac{\mathrm{E}(D_1^2)}{\mu_1} = \mu_1 + \frac{\sigma_1^2}{\mu_1},$$
$$\widetilde{\mu}_{2|1} = \mathrm{E}(\widetilde{D}_{2|1}) = \frac{\mathrm{E}(D_1 D_2)}{\mu_1} = \mu_2 + \frac{\sigma_{12}}{\mu_1}$$

The corresponding equations are valid when starting with an edge of type 2—just switch 1 and 2 in the equations. We return to these equations in Section 2.4.

In the later sections, we compare epidemics on the weighted graph with epidemics on the corresponding unweighted graph where we simply neglect the weights, so $D = D_1 + D_2$. Expressions for the mean, the variance, and the probability mass function for the unweighted configuration model graph are

$$\mu = \mu_1 + \mu_2,$$
$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + 2\sigma_{12},$$
$$p_i = \mathbb{P}(D = i) = \sum_{k=0}^{i} p_{k,i-k}$$

These quantities can be used directly in the results for the *size-biased* distribution for the unweighted configuration model graph in Section 2.2.

2.4 SIR epidemics

We work with the SIR model in Reed–Frost and Markovian settings, see, e.g., Lefèvre (1990). In this model, vertices in the graph represent people and edges represent paths by which people can infect each other. Initially only one vertex (the *index case*) is infected. Throughout this paper, we assume that the index case is selected uniformly at random among all vertices in the graph. An infected vertex is infectious (can infect susceptible neighbors) until it has recovered. After recovering the vertex is immune forever and cannot ever infect any other vertex. The epidemic stops when there are no more infected vertices. At this time typically some portion of all vertices are recovered and some are still susceptible. Epidemic models of this type are in general known to have the probability distribution for the final size of the epidemic located in two different parts: either only few get infected, meaning that in a large population the fraction getting infected is close to 0. The remaining possibility is that a close to deterministic *fraction* get infected. The former is referred to as a *minor outbreak* and the latter a major outbreak. The size of the fraction is called the final size of the epidemic. In this paper two of the parameters we study are the relative final size of a major outbreak and the probability of a major outbreak. The expected number of vertices that a typical infected vertex infects early on in the epidemic, when the population is almost completely susceptible, is called the *basic reproduction number* (denoted R_0). All quantities are derived in the limit $n \to \infty$, although this is not always mentioned explicitly and the derivations are not formal.

In the Reed–Frost setting (see, e.g., Bailey, 1975, Section 8) the epidemic develops in generations, starting with a single infected individual in the first generation. In each generation each infected vertex tries to infect its susceptible neighbors after which the vertex recovers. Thus in the

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

next generation only the newly infected vertices continue to spread the infection. We assume that for each edge (among the susceptible neighbors) infection occurs independently with probability π_ξ that is allowed to depend only on the edge type $\xi \in \{1, 2\}$. In this model the infectious period can be thought of as being deterministic and the same for all vertices.

In the Markovian setting, an infected vertex has an infectious period that is exponentially distributed with recovery rate γ. Thus the infectious period is random, rather than deterministic as in the Reed–Frost setting. An infectious vertex has infectious contacts with each susceptible neighbor independently according to a Poisson process with intensity β_ξ that depends only on the edge type. The probability of an arbitrary edge of an infected vertex passing on the infection to a susceptible edge before the end of the infectious period is then

$$\pi_\xi = \frac{\beta_\xi}{\beta_\xi + \gamma}$$

When comparing the Reed–Frost and the Markovian settings, we choose β_ξ such that π_ξ are the same in both models. We must, however, keep in mind that the infectious period affects all neighbors of a vertex and thus in the Markovian setting, infectious events along different edges of the same vertex are not independent. This must be taken into account when determining the probability of a major outbreak as pointed out, e.g., by Kenah & Robins (2007).

2.5 The mean infectious activity

One of the main purposes of this paper is to compare how an epidemic on an unweighted network compares with that on a weighted network. We do this by comparing a graph model in which individuals on average have the same transmission probability to all its neighbors, with a graph model where these transmission probabilities are different (specifically studying a network with two different transmission probabilities). In order to calibrate the two networks, before comparing epidemics on them, we define the *mean infectious activity* ($\mathscr{A}_\mathbf{I}$) as the expected number of secondary infections caused by a single infected vertex, *selected uniformly at random*, when all other vertices are susceptible. We calibrate the unweighted and the weighted networks by requiring that the *mean infectious activity* is the same for both networks. Note that the *mean infectious activity* can be estimated without knowing exactly how individuals are connected to each other in the network, and this is the main reason for selecting this quantity to calibrate the unweighted and the weighted networks. We can estimate the infectious activity from a sample of individuals (selected uniformly at random) from the network or we can obtain it directly from the degree distribution for the entire network (if it is known). In our case, for the unweighted and the weighted networks, we have

$$\mathscr{A}_\mathbf{I}^{(unweighted)} = \pi\mu \text{ and} \tag{5}$$

$$\mathscr{A}_\mathbf{I}^{(weighted)} = \pi_1\mu_1 + \pi_2\mu_2 \tag{6}$$

When comparing the two different networks we require that

$$\mathscr{A}_\mathbf{I}^{(unweighted)} = \mathscr{A}_\mathbf{I}^{(weighted)} = \mathscr{A}_\mathbf{I} \tag{7}$$

We study how the *infectious activity*, the degree distribution (for some selected networks), and the amount of transmission heterogeneity (in the weighted graph model) affect the basic reproduction number R_0, the relative final size of a large epidemic, and the probability of a large epidemic.

If we increase the mean infectious activity (without changing anything else) the epidemic spreads more easily, resulting in a larger R_0, an increased relative final size, and an increased probability of a major outbreak.

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

Remembering that in the unweighted graph model $D = D_1 + D_2$, we have

$$\pi = \frac{\pi_1\mu_1 + \pi_2\mu_2}{\mu_1 + \mu_2}$$

since $\mu = \mu_1 + \mu_2$.

If we divide Equation (6) by $\mathscr{A}_{\mathbf{I}}$, we obtain

$$r_1 + r_2 = 1$$

where $r_\xi = \frac{\pi_\xi \mu_\xi}{\mathscr{A}_{\mathbf{I}}}$ determines how the mean infectious activity is distributed between the different edge types—we call it the *relative infectious activity*. By varying (e.g.) r_1, we can study how the epidemic is affected by a change in the balance in the transmission probability between the edge types, independently of the mean infectious activity in the network. We use the mean infectious activity together with the relative infectious activity in the theoretical results (Section 3) and in the numerical results (Section 4) where they allow for a consistent way of comparing different models.

3. Theoretical results

In this section we analyze the development of SIR epidemics, on the weighted and the unweighted configuration model graphs, in both Reed–Frost and Markovian settings, to obtain explicit expressions or algorithms for calculating R_0 (Section 3.1), the probability of a major outbreak (Section 3.2), and the relative final size of a major outbreak (Section 3.3). Each subsection contains results for both the unweighted and the weighted configuration model. While the results for the unweighted configuration model are generally known, they are included here for direct comparison with the results for the weighted configuration model.

3.1 Basic reproduction number

In the configuration model, the expected degree of a vertex that is reached by following an edge from an infected vertex is determined by the size-biased distribution (see Section 2.3 and Equation (2)). When looking at the number of edges that can spread an infection in a mostly susceptible population, we must deduct one edge, since the infection cannot spread back along the infecting edge, and must also multiply by the probability that an edge infects a susceptible vertex.

3.1.1 Unweighted graph model

For the unweighted graph, we obtain

$$R_0 = \pi\,(\widetilde{\mu} - 1) = \pi\left(\mu + \frac{\sigma^2}{\mu} - 1\right) = \mathscr{A}_{\mathbf{I}}\left(1 + \left(\frac{\sigma}{\mu}\right)^2 - \frac{1}{\mu}\right)$$

remembering that the mean infectious activity $\mathscr{A}_{\mathbf{I}} = \pi\mu$ in the final step. We want to compare the unweighted model with the weighted model, where simultaneous degree distribution $\mathbf{D} = (D_1, D_2)$ is given, and set $D = D_1 + D_2$, counting only the total number of edges, regardless of type. Using results from Section 2.5, we obtain:

$$R_0 = \mathscr{A}_{\mathbf{I}}\left(1 + \frac{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}{(\mu_1 + \mu_2)^2} - \frac{1}{\mu_1 + \mu_2}\right) \tag{8}$$

This is strictly increasing in ρ (all other parameters being fixed) and obtains its minimum and maximum values when ρ obtains its minimum and maximum values, respectively. The allowed range of ρ is obtained from the weighted model, see below.

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

A special case is when $\mu_\xi = \sigma_\xi^2$ (such as for the Poisson distribution), and then

$$R_0 = \mathscr{A}_\mathbf{I}\left(1 + 2\rho\frac{\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2}\right)$$

If, in addition, $\rho = 0$, we obtain $R_0 = \mathscr{A}_\mathbf{I}$.

3.1.2 Weighted graph model

In the weighted graph model, we need to take into account which type of edge that infected the vertex since the size-biased distribution depends on this. We must remove one edge *of the type that infected the vertex*, since a vertex cannot reinfect its infector. This results in the *next generation matrix*

$$\mathbf{K} = \begin{pmatrix} \pi_1\left(\widetilde{\mu}_{1|1} - 1\right) & \pi_2\,\widetilde{\mu}_{2|1} \\ \pi_1\,\widetilde{\mu}_{1|2} & \pi_2\left(\widetilde{\mu}_{2|2} - 1\right) \end{pmatrix}$$

(see also Section 2.3 for definitions). In a model where there are different types of vertices, this matrix gives the number of secondary infections of type i caused by a single individual of type j, early on in the epidemic (while essentially the entire population is susceptible (see Diekmann et al., 2012, Chapter 7). In our paper we instead keep track of infecting edges and, with a slight change to the definition, the matrix consists of the expected number of *infecting edges* of each type (column 1 and column 2) in the next generation when a vertex is infected through *an edge* of type 1 (row 1) and type 2 (row 2). R_0 is given by the largest eigenvalue of this matrix (see, e.g., Diekmann et al., 2012, Chapter 7).

The eigenvalue λ is obtained as a solution to the characteristic equation

$$\det\left(\mathbf{K} - \lambda\mathbf{I}\right) = 0$$

The solution can be written in relatively compact form using some definitions from Section 2.5 and some additional definitions, including the coefficient of variation $\mathscr{CV}_\xi$:

$$\begin{aligned} \mathscr{A}_\mathbf{I} &= \pi_1\mu_1 + \pi_2\mu_2, \\ r_\xi &= \frac{\pi_\xi\mu_\xi}{\mathscr{A}_\mathbf{I}}, \\ \mathscr{CV}_\xi &= \frac{\sigma_\xi}{\mu_\xi}, \\ \nu_\xi &= \mathscr{CV}_\xi{}^2 - \frac{1}{\mu_\xi} \end{aligned}$$

remembering that $\xi \in \{1, 2\}$ indicates the edge type.

The full solution is

$$R_0 = \frac{\mathscr{A}_\mathbf{I}}{2}\left(1 + r_1\nu_1 + r_2\nu_2 + \sqrt{\left(1 + r_1\nu_1 + r_2\nu_2\right)^2 + 4r_1r_2\left(\left(\rho\mathscr{CV}_1\mathscr{CV}_2 + 1\right)^2 - \left(\nu_1 + 1\right)\left(\nu_2 + 1\right)\right)}\right) \tag{9}$$

The allowed range of ρ depends on the marginal distributions of the degrees of the two edge types as well as on the correlation structure between them. An important observation is that R_0 increases when ρ increases. Thus the minimum and maximum must be obtained for the minimum and maximum value of ρ, respectively.

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

The full solution simplifies in some cases, e.g., if $\mu_\xi = \sigma_\xi^2$ (such as for the Poisson distribution). Then $\nu_\xi = 0$, and we have

$$R_0 = \frac{\mathscr{A}_\mathbf{I}}{2}\left(1 + \sqrt{1 + 4r_1 r_2\left((\rho\mathscr{C}\mathscr{V}_1\mathscr{C}\mathscr{V}_2 + 1)^2 - 1\right)},\right)$$

If we, in addition, require that $\rho = 0$, this gives that $R_0 = \mathscr{A}_\mathbf{I}$, just as for the unweighted model. An *example* is when $\{D_\xi\}$ are *independent* Poisson distributed variables.

3.2 Probability of a major outbreak

In the configuration model (in the limit of an infinite population) the probability of a major outbreak can be calculated as the probability of survival of a Galton–Watson branching process where the offspring is the number of new infected vertices in each generation, see, e.g., Britton et al. (2007) and also Appendix C.1. In our application, infection starts with the index case and then spreads through one type of edge (the unweighted configuration model) or two types of edges (the weighted configuration model). In the weighted model, the degree distribution of a newly infected vertex depends on the type of edge through which the vertex was infected. The distribution of the number of vertices that are actually infected is different in the Reed–Frost and the Markovian settings. In the Reed–Frost setting, the infectious period is deterministic and the same for all vertices, so infectious events on edges from an infected vertex are *independent*. In the Markovian setting, the infectious period is random (exponential), and since infectious events on all edges for a specific vertex are simultaneously affected by the infectious period, these events become *dependent* (see also Kenah & Robins, 2007).

We thus need to treat the Reed–Frost and Markovian settings slightly differently.

3.2.1 Unweighted model

Let $\{p_i^*\}$ define the distribution of the number of *infecting* edges (edges that do spread the infection) of the index case. Let $\{\widetilde{p}_i^*\}$ define the distribution of the number of infecting edges of a vertex infected after the index case (corresponding to following an edge in the configuration model, see Section 2.2). Define the probability generating functions

$$f^*(s) = \sum_{i=0}^{\infty} s^i p_i^*$$

$$\widetilde{f}^*(s) = \sum_{i=0}^{\infty} s^i \widetilde{p}_i^*$$

The probability q that the branching process dies out, given that we start with an infecting *edge*, is the solution to the fixed-point equation

$$q = \widetilde{f}^*(q)$$

as discussed in Appendix C.1. We then apply this solution to each edge of the index case, giving the probability τ of a major outbreak (the probability that the process does not die out)

$$\tau = 1 - f^*(q)$$

What remains is to obtain expressions for the probability mass functions p_i^* and $\widetilde{p}_i^*$. These can be derived from the distributions p_i and $\widetilde{p}_i$ (see Sections 2.2 and 2.3). Study a given infected vertex that has exactly k edges that could spread the infection. Then the actual number of edges that *do* spread the infection is between 0 and k. Let $\phi(i \mid k)$ denote the probability that exactly i edges out of the k available spread the infection. Clearly $\phi(i \mid k)$ depends on the model for spreading

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

the infection—e.g., we can expect it to be different in the Reed–Frost and the Markovian settings. However, if $\phi(i \mid k)$ is known, we can easily obtain p_i^* and $\widetilde{p}_i^*$ through

$$p_i^* = \sum_{k=i}^{\infty} \phi(i \mid k) p_k \text{ and} \tag{10}$$

$$\widetilde{p}_i^* = \sum_{k=i}^{\infty} \phi(i \mid k) \widetilde{p}_{k+1} \tag{11}$$

remembering that $k+1$ is required for the *size-biased* distribution since the infection cannot spread back on the edge that infected the current vertex.

In the Reed–Frost model, infectious events on edges are independent and, given the probability π that an edge spreads an infection, the number of edges that spread the infection is distributed as $\text{Bin}(k, \pi)$ so that

$$\phi(i \mid k) = \binom{k}{i} \pi^i (1-\pi)^{k-i}$$

When this result is inserted into Equations (10) and (11) typically further simplification is not possible except in some special cases, like when D is Poisson or binomially distributed.

In the Markovian setting, the derivation of $\widetilde{p}_i^*$ is slightly more complicated. If we fix the duration of the infectious period $T = t$, the infectious events on the edges become independent and each edge infects another vertex with probability $\pi(t) = 1 - e^{-\beta t}$, independently of the other edges. Let D^* denote the number of edges that pass on the infection. Then (when $k \geq i$)

$$\mathbb{P}\left(D^* = i \mid k, T = t\right) = \binom{k}{i} \pi(t)^i (1-\pi(t))^{k-i}, \text{ and so}$$

$$\begin{aligned} \phi(i \mid k) &= \int_0^{\infty} \binom{k}{i} \pi(t)^i (1-\pi(t))^{k-i} f_T(t) \mathrm{d}t \\ &= \int_0^{\infty} \binom{k}{i} (1-e^{-\beta t})^i (e^{-\beta t})^{k-i} \gamma e^{-\gamma t} \mathrm{d}t \end{aligned}$$

When this result is inserted into Equations (10) and (11) typically further simplification is not possible.

3.2.2 Weighted model

Results for the weighted model follow the same method as for the unweighted model, taking into account the two different edge types. Let $p^*(i, j)$ define the distribution of *infecting* edges of the index case. Let $\widetilde{p}_\xi^*(i, j)$ define the distribution of the number of infecting edges of a vertex that were infected after the index case (through a type $\xi \in \{1, 2\}$ edge, see Section 2.3). Let $\mathbf{s} = (s_1, s_2)$ and define the probability generating functions

$$f^*(\mathbf{s}) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} s_1^i s_2^j p^*(i, j)$$

$$\widetilde{f}_\xi^*(\mathbf{s}) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} s_1^i s_2^j \widetilde{p}_\xi^*(i, j)$$

The probabilities $\{q_\xi\}$ that the branching process dies out, *given that we start with an infecting edge* of type $\xi \in \{1, 2\}$, are given by the solution to the fixed-point equation

$$\mathbf{q} = (\widetilde{f}_1^*(\mathbf{q}), \widetilde{f}_2^*(\mathbf{q}))$$

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

where $\mathbf{q} = (q_1, q_2)$, as discussed in Appendix C.1. Just as in the unweighted case, we then apply this solution to each edge of the index case, giving the probability τ of a major outbreak (the probability that the process does not die out)

$$\tau = 1 - f^*(\mathbf{q})$$

The probability mass functions $p^*(i,j)$ and $\widetilde{p}^*_\xi(i,j)$ can be derived from the distributions p_{ij} and $\widetilde{p}_{ij}$ (see Sections 2.2 and 2.3). Study a given infected vertex that has k edges of type 1 and l edges of type 2 that could spread the infection. Let $\phi(i,j \mid k,l)$ denote the probability that (i,j) edges out of the (k,l) available spread the infection.

Then

$$p^*(i,j) = \sum_{k=i}^{\infty}\sum_{l=j}^{\infty} \phi(i,j \mid k,l) p_{k,l}, \tag{12}$$

$$\widetilde{p}^*_1(i,j) = \sum_{k=i}^{\infty}\sum_{l=j}^{\infty} \phi(i,j \mid k,l) \widetilde{p}_{k+1,l} \text{ and} \tag{13}$$

$$\widetilde{p}^*_2(i,j) = \sum_{k=i}^{\infty}\sum_{l=j}^{\infty} \phi(i,j \mid k,l) \widetilde{p}_{k,l+1} \tag{14}$$

again remembering that the infection cannot spread back on the edge that infected the current vertex.

In the Reed–Frost setting infectious events on edges are independent and if the probability that an edge spreads an infection is π_ξ, then the number of edges that spread the infection are distributed independently as $\text{Bin}(k,\pi_1)$ and $\text{Bin}(l,\pi_2)$ so that

$$\phi(i,j \mid k,l) = \binom{k}{i}\pi_1^i(1-\pi_1)^{k-i}\binom{l}{j}\pi_2^j(1-\pi_2)^{l-j}$$

When this result is inserted into Equations (12), (13), and (14), then typically further simplification is not possible except in some special cases, e.g., when D_ξ are independently Poisson distributed.

In the Markovian setting, given that the duration of the infectious period $T = t$, each edge infects another vertex with probability $\pi_\xi(t) = 1 - e^{-\beta_\xi t}$, independently of the other edges. Let D^*_ξ denote the number of edges of type ξ that pass the infection on. Then (when $k \geq i$ and $l \geq j$)

$$\mathbb{P}\left(D^*_1 = i, D^*_2 = j \mid k,l,T = t\right) = \binom{k}{i}\pi_1(t)^i(1-\pi_1(t))^{k-i}\binom{l}{j}\pi_2(t)^j(1-\pi_2(t))^{l-j}$$

and so

$$\begin{aligned}\phi(i,j \mid k,l) &= \int_0^\infty \binom{k}{i}\pi_1(t)^i(1-\pi_1(t))^{k-i}\binom{l}{j}\pi_2(t)^j(1-\pi_2(t))^{l-j} f_T(t)\mathrm{d}t \\ &= \int_0^\infty \binom{k}{i}\left(1-e^{-\beta_1 t}\right)^i\left(e^{-\beta_1 t}\right)^{k-i}\binom{l}{j}\left(1-e^{-\beta_2 t}\right)^j\left(e^{-\beta_2 t}\right)^{l-j}\gamma e^{-\gamma t}\mathrm{d}t\end{aligned}$$

When this is inserted into Equations (12), (13), and (14) typically further simplification is not possible.

3.3 Final size of a major outbreak

In the Reed–Frost setting, the relative final size of a major outbreak is equal to the probability of a major outbreak (Britton et al., 2007), and this has already been calculated in Section 3.2.

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

The setting in this paper is not identical to that of Britton et al. (2007), so some justification for the case where there are different types of edges is needed. Consider the case where the infection probabilities between each pair of individuals are symmetric (but may vary between different pairs), and the Reed–Frost epidemic applies making infection events along different edges independent. A transmission between one pair of individuals can at most go in one direction, so when the two events have the same probability one random variable for "potential transmission event" is sufficient to model possible transmission between the two. Further, since all transmission events occur independently, the epidemic can be constructed by first "thinning" the original network, by which we mean that an original edge should be removed if the random variable of that potential transmission event indicates no transmission event. The remaining edges then signify pairs of individuals that are neighbors that both get infected should one of them get infected. The final size is then the connected component of those connected to the index case. Such a graph will have one giant connected component and all other components being of order log n or smaller. The probability of a major outbreak is identical to the probability that the index case belongs to the giant, and this probability is simply the relative size of the giant, i.e., the relative final size of a major outbreak.

In the Markovian setting, infectious events on edges are dependent and so further analysis is needed. Instead of studying how the epidemic develops forward in time (starting with the index case), we instead select a vertex uniformly at random and study which vertices that would infect it if they were infected. We continue this process (in the limit of an infinite population) by following edges backwards in time to create the *susceptibility set*, see Ball & Neal (2008). The susceptibility set for vertex j consists of all vertices that would have infected it, had they become infected. Intuitively, if this set consists of a significant number of vertices in the graph, then vertex j cannot escape a major outbreak if one should occur (since it is unlikely that all of the many vertices in the susceptibility set will escape infection). We can thus investigate all vertices in the graph and see what proportion of the vertices that have susceptibility sets that are of significant size.

More formally, in the configuration model this once again corresponds to studying a Galton–Watson branching process and determining the probability that the process does not die out. Following the branching process in the reverse direction (as for susceptibility sets) also ensures that infectious events on edges are independent since the incoming edges are attached to different vertices which are independent. The probability that the branching process (that creates the susceptibility set) *survives* is equal to the relative final size of a major outbreak. The probability of infection is the same as when calculating the probability of a major outbreak in the Reed–Frost setting in Section 3.2, and so the size of the outbreak in the Markovian setting is equal to the probability of a major outbreak in the Reed–Frost setting. Thus the relative final size of a major outbreak in the Reed–Frost and in the Markovian settings are equal. This same argument is also a justification for the probability of a major outbreak and the final size of a major outbreak being equal in the Reed–Frost setting. Note that, in the Markovian setting, the *relative final size of a major outbreak* and the *probability of a major outbreak* are in general not equal. Useful reading is Newman (2002) and the important note in Kenah & Robins (2007).

4. Numerical results for some theoretical and empirical degree distributions

To illustrate the weighted configuration model and to compare it with the unweighted model, we numerically analyze some theoretical and empirical degree distributions (see Section 4.2) with respect to R_0, the relative final size of a major outbreak, and the probability of a major outbreak. Whenever possible we plot the results as a function of the balance between the two edge types (the relative infectious activity) and the correlation coefficient between the edge types in Section 4.3. In order to be able to vary the correlation coefficient for otherwise fixed bivariate degree distribution, we make use of copulas (see Section 4.1 and Appendix B.1).

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

4.1 Network models using copulas

Theoretical and empirical bivariate degree distributions often have a fixed correlation coefficient between the two edge types. In other cases a bivariate distribution does not exist (or we do not have access to it). In such cases modeling a bivariate distribution using a copula can be useful, see, e.g., Nelsen (2007). We must remember that within this framework, different copulas are possible and these will result in different properties for the resulting bivariate distribution. For this paper we choose to use a copula based on the bivariate standard normal distribution. It is simple to simulate and it allows for modeling the correlation between the degrees of the two edge types through a large range. We obtain the bivariate distribution function

$$F(i,j) = C_\rho(F_1(i), F_2(j))$$

where $C_\rho(\cdot,\cdot)$ denotes the bivariate standard normal copula with correlation ρ and $F_\xi(\cdot)$ represents the marginal distribution function for edge type ξ. The marginal distributions for each edge type can be taken from empirical networks or from theoretical distributions. In this paper we do both. Note that the correlation derived from $F(i,j)$ will depend not only on the copula but also on the marginal distributions, and will thus typically be different from the ρ that is used in the equation above. More information on copulas can be found in Appendix B.1.

4.2 Theoretical and empirical distributions

- *A theoretical binomial network*
  This is a degree distribution where the marginal distributions are both binomial regardless of the correlation between D_1 and D_2 (Biswas & Hwang, 2002). For a more complete description see Appendix A.1. Here we only mention that it has five parameters, n_1, p_1, n_2, p_2, and ρ (the correlation). The mean and the variance of the marginal distributions are

$$\mathrm{E}(D_\xi) = n_\xi p_\xi, \tag{15}$$

$$\mathrm{Var}(D_\xi) = n_\xi p_\xi \left(1 - p_\xi\right) \tag{16}$$

  for $\xi \in \{1, 2\}$. These can thus be varied separately, although $\mathrm{Var}(D_\xi)$ is restricted to the range $\left[0, \mathrm{E}(D_\xi)\right]$. The correlation can be varied within the full range $[-1, 1]$ only when $n_1 = n_2$ and $p_1 = p_2$. This distribution can also be used to approximate a bivariate Poisson distribution if $\{n_\xi\}$ are chosen to be large and $\{p_\xi\}$ are chosen to be small.
- *A theoretical heavy-tailed network*
  This is a heavy-tailed distribution that is based on typical empirical distributions, e.g., distributions that appear for various *preferential attachment* models. Such distributions often have a tail that goes as $k^{-\alpha}$, where α is a parameter that is often in the range 2−4 for empirical networks, although it can go outside this range also. An overview of such networks and models (often denoted *scale-free*) can be found in Barabási & Bonabeau (2003). For low degrees the empirical distributions often do not decay as rapidly. See Liljeros et al. (2001) and Adamic & Huberman (2000) for some empirical examples.
  We choose a distribution that approximates the described properties. Let

$$p_k = \frac{1}{c}(k + k_0)^{-\alpha},\ k = 0, 1, \ldots$$

  where c is a norming constant (the Hurwitz zeta function) that is finite whenever $k_0 > 0$ and $\alpha > 1$. Note that k_0 essentially determines the shape of the probability mass function for *lower degrees*. In this paper we model bivariate heavy-tailed distributions by using standard normal copulas (see also Section 4.1).
- *An empirical sexual network*
  This is a degree distribution for sexual relationships for a heterosexual population.[1] People have stated how many casual and how many stable sexual relationships they have had during

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

the last year, see Hansson et al. (2018) for details. The dataset analyzed here consists of 645 individuals. This information is treated as a bivariate edge distribution where the transmission probability of the two types of edges is allowed to differ. It is important to note that this is not a valid model for the actual sexual interactions within this group of people since the two sexes interact (mainly) with someone of the other sex, while in our model we treat all individuals as identical (apart from the bivariate degree). Also, this is not a random subset of the general population. Here we use the dataset only as an example of an empirical dataset.

Letting type 1 edges represent casual relationships and type 2 edges represent stable relationships, we have $\mu_1 \approx 1.42$, $\mu_2 \approx 1.70$, $\sigma_1 \approx 0.99$, $\sigma_2 \approx 1.73$, and $\rho \approx -0.0652$.

- *An empirical Swedish population network*
  This is a large empirical degree distribution for the Swedish population[2] that is based on data containing only the workplace and family affiliation of people in Sweden, see Holm et al. (2006) for details. Edges to family members and within workplaces are treated as different types. Some workplaces are very large, and to reduce computational workload and make the model somewhat more realistic people are randomly assigned to work groups within each workplace. Letting type 1 edges represent family edges and type 2 edges represent company edges, we have $\mu_1 \approx 2.00$, $\mu_2 \approx 20.98$, $\sigma_1 \approx 1.44$, $\sigma_2 \approx 27.8$, and $\rho \approx -0.241$.

4.3 Numerical results

Here we present numerical results for the theoretical and the empirical distributions, together with some comparisons with the copula model. The correlation coefficients that are presented for the copula model always relate to the resulting bivariate degree distribution.

We do not always explicitly comment on which setting, Reed–Frost or Markovian, that is used for each figure. However, we remind the reader that in our model, the relative final size of a major outbreak is the same for both settings and this is also equal to the probability of a major outbreak in the Reed–Frost setting.

4.3.1 Binomial network

Figure 1 shows R_0, the relative final size of a major outbreak and the probability of a major outbreak (Markovian) for a bivariate binomial distribution as described in Section 4.2 and also compared with a copula model. In the copula model, the two marginal distributions have been modeled using a bivariate standard normal copula. In the figure the correlation is plotted on the horizontal axis and r_2 is plotted on the vertical axis. Low or high values of r_2 indicate that the infection is spread mainly by type 1 or type 2 edges, respectively. The horizontal red line indicates when the probability of infection is the same for type 1 and type 2 edges.

The most important observations are that the binomial model and the copula model produce results that are almost identical, within the range where both models are defined, but that the copula model allows for modeling the correlation through a much wider range than does the binomial model. The binomial model is strongly limited in the allowed range for the correlation between the two edge types, because of the way the model is defined. In the case of the copula model, only correlations close to -1 or $+1$ are excluded. This means that using the marginal distributions together with the copula model gives a more useful model with respect to investigating the effect of correlation.

In addition, we note that the unweighted model produces much more narrow ranges of the investigated parameters than the weighted model. For example, relative final size of the unweighted model spans the range 0.37–0.42 (approximately), while in the weighted model it spans the range 0.15–0.43. Values in the lower range occur when more of the epidemic is transmitted on type 1 edges. This is an effect of the marginal degree distribution for the type 1 edges having a smaller variance, which is an important parameter in the configuration model.

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

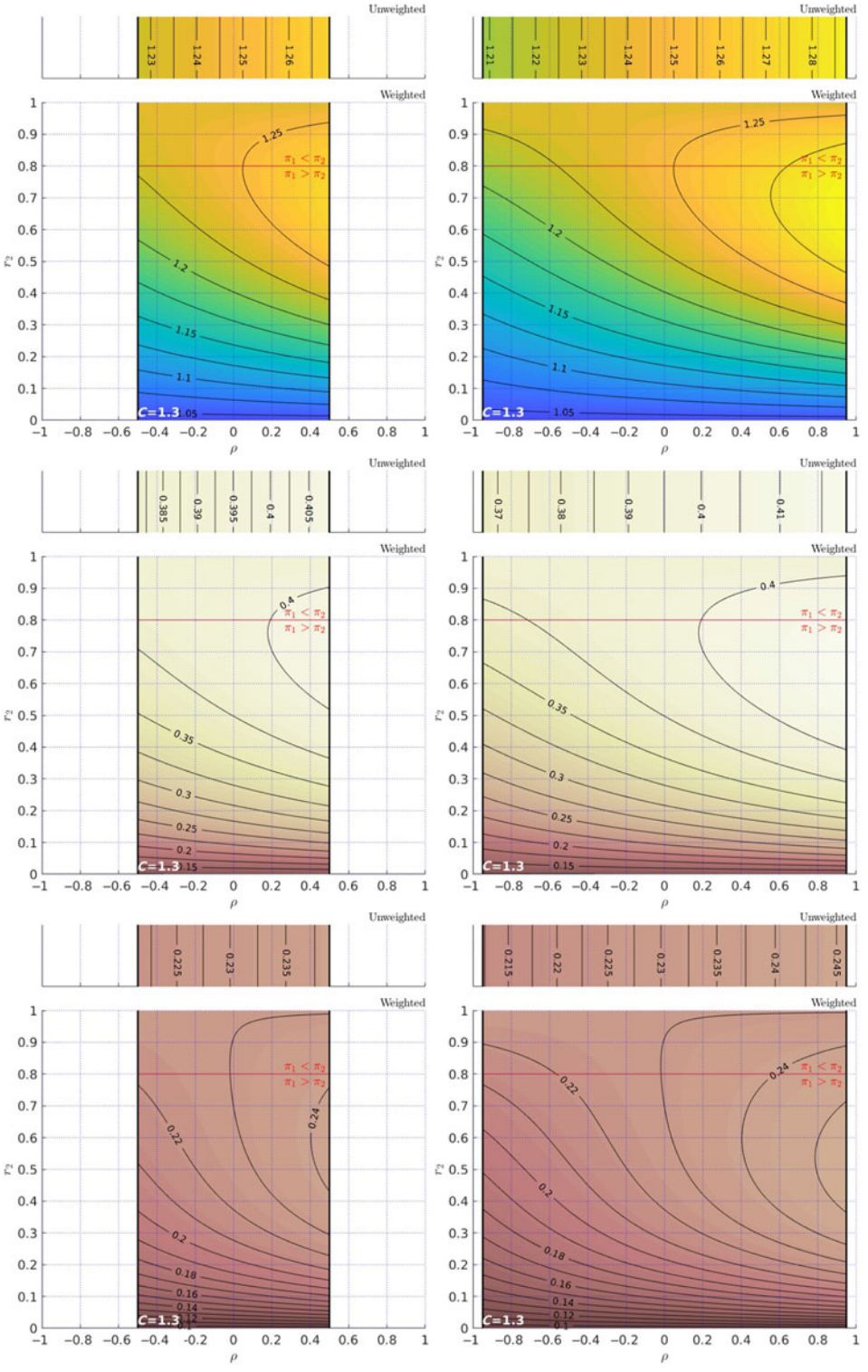

Figure 1. The figure shows contour plots of R_0 (upper), the relative final size of a major outbreak (middle), and the probability of a major outbreak in the Markovian setting (lower) for the bivariate binomial model (left) and for the same marginal distributions modeled through the bivariate normal copula (right). The parameters are $n_1 = 5$, $p_1 = 0.5$, $n_1 = 20$, and $p_1 = 0.5$. The mean infectious activity was set to 1.3 (denoted $C = 1.3$ in the figure).

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

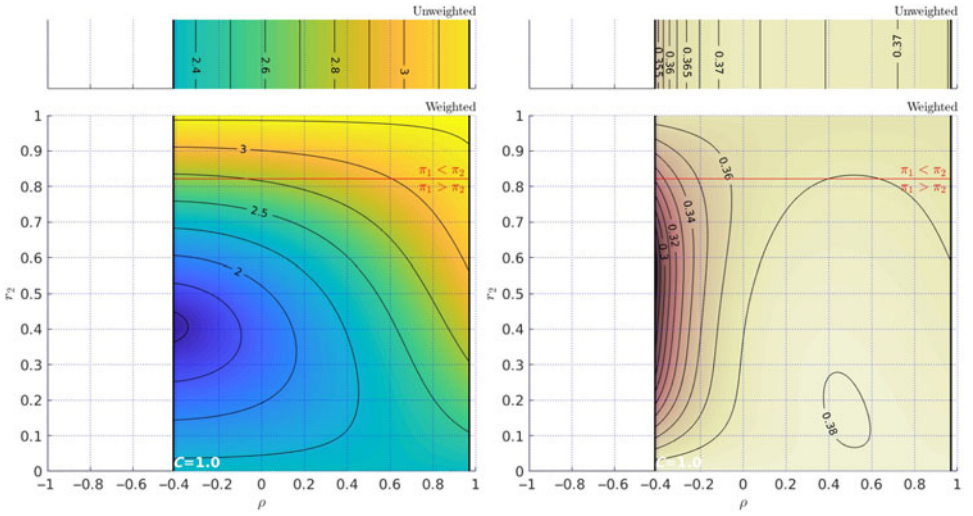

Figure 2. The figure shows contour plots of R_0 (upper left) and relative final size of a major outbreak for the heavy-tailed network modeled using the bivariate standard normal copula. For type 1 edges the parameters for the marginal degree distributions are $\alpha = 10$ and $k_0 = 20$, but the distribution is truncated at 25 edges (allowing only degrees between 0 and 25 to have positive probability). For type 2 edges the parameters are $\alpha = 4$ and $k_0 = 20$, and the distribution is truncated at 200 edges. Truncating the distributions is necessary to be able to perform the calculation within reasonable time. The mean infectious activity was set to 1 (denoted by $C = 1.0$ in the figure).

4.3.2 Heavy tail network

Figure 2 shows R_0 and the relative final size of a major outbreak for a heavy-tailed degree distribution (see Section 4.2) modeled using the bivariate standard normal copula so that ρ can be varied. Note that the degree distributions were truncated (see the caption). Truncation can severely affect the epidemic calculations for heavy-tailed distributions, especially when these have a parameter α close to 2 (from above), since the second moment which is important in the configuration model becomes infinite at $\alpha \leq 2$. In this paper we have chosen heavy-tailed distributions with relatively fast fall-off (high values of α), making the distributions less susceptible to truncation. Also, we use the distribution mainly as an illustration of results that can be obtained, and we have therefore not explored the effect of the truncation further.

One observation from the plot is that the maximum possible negative correlation is approximately -0.4. This limitation is related to the shape of the marginal distributions which have high probability mass for low degrees and low probability mass for high degrees. A negative correlation requires matching high type 1 degrees with low type 2 degrees (and vice versa), and this is limited by the mismatch in probability mass. In the plot we did not explicitly calculate the available range of the correlation coefficient, instead the ρ in the copula model (see Appendix B.1) was varied between -1 and $+1$ and the calculated epidemic parameters were plotted at the correlation coefficient of the resulting bivariate degree distribution.

The most important observations for the heavy-tailed distribution is the relatively wide range of R_0 (from about 1.5 to 3.3), while the relative final size of a major outbreak does not vary much. We note that, in the weighted model, the relative final size does not vary monotonically with the correlation. However, the relative final size does not vary much, so in this case this property may not be of practical importance.

4.3.3 Sexual network

Figure 3 shows R_0 and the relative final size of a major outbreak (for fixed $\mathscr{A}_1$) for the sexual relationship network modeled using the bivariate standard normal copula so that ρ can be varied.

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

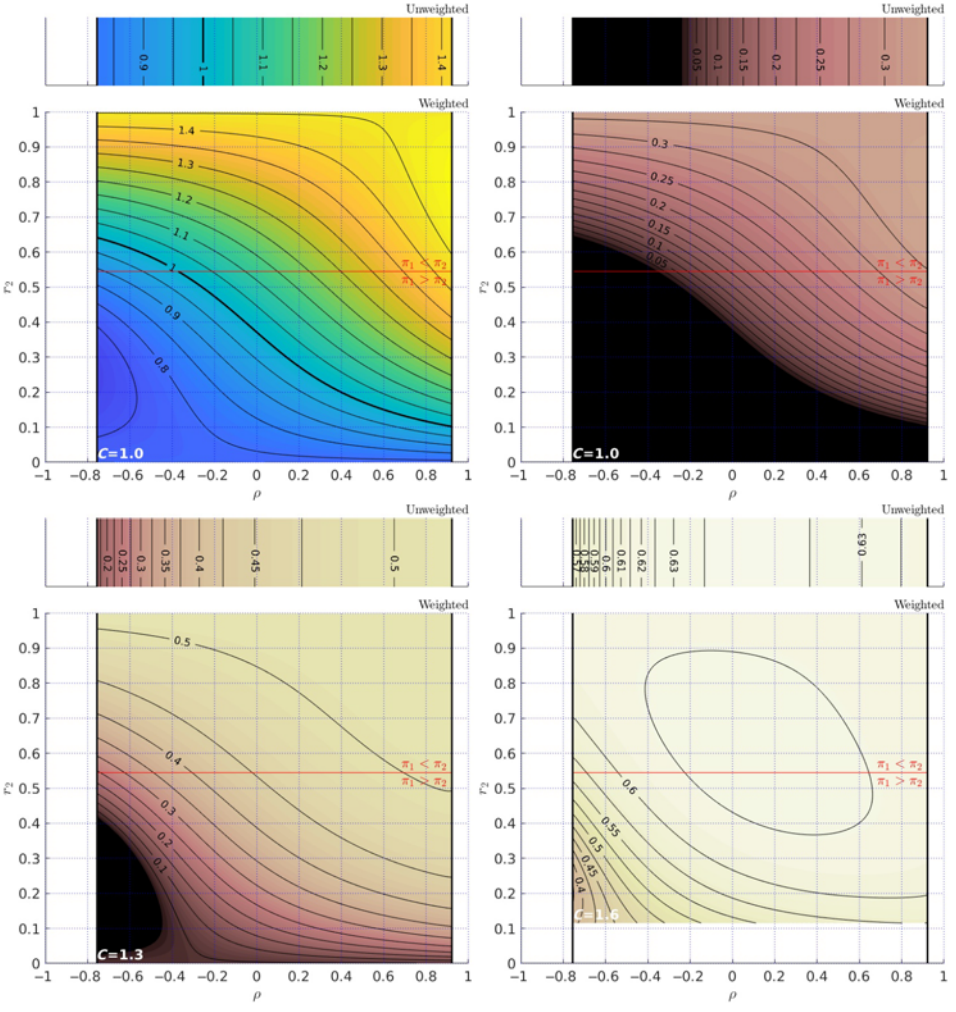

Figure 3. The figure shows contour plots of R_0 (upper left) and the relative final size of a major outbreak (the remaining three) for different values of $\mathscr{A}_\mathbf{I}$ for the sexual relationship network modeled using the bivariate standard normal copula. Each plot consists of a larger (lower) area corresponding to the weighted model and a smaller (upper) area corresponding to the unweighted model. The horizontal axis represents the correlation and the vertical axis represents the balance between the degrees of type 1 and type 2 edges (only in the weighted model). Note that R_0-results for other values of $\mathscr{A}_\mathbf{I}$ can be obtained simply by multiplying the values in the R_0-plot with $\mathscr{A}_\mathbf{I}$, since this plot was based on $\mathscr{A}_\mathbf{I} = 1$ and R_0 is directly proportional to $\mathscr{A}_\mathbf{I}$. The value of the mean infectious activity is denoted C in the figures.

From the plot we see that both R_0 and the final size depend greatly on the correlation between the degrees *and* on the balance between the probability of infection on type 1 and type 2 types. We remind the reader that for the empirical degree distribution, letting type 1 edges represent casual relationships and type 2 edges represent stable relationships, we have $\mu_1 \approx 1.42, \mu_2 \approx 1.70, \sigma_1 \approx 0.99, \sigma_2 \approx 1.73$, and $\rho \approx -0.0652$. The gap at the bottom of the bottom right plot in Figure 3 is caused by there being on average 1.42 edges of type 1 for each vertex while $\mathscr{A}_\mathbf{I} = 1.6$. Since infection probabilities cannot be larger than 1, edges of type 2 need to contribute to the epidemic through some infection probability that is larger than zero. Thus there will be a region of r_2 values (close to 0) that are not allowed.

While R_0 must always increase with ρ (as can be seen from Equations (8) and (9)), this is not the case for the relative final size of a major outbreak. When $\mathscr{A}_\mathbf{I} = 1.6$ we see that the relative final

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

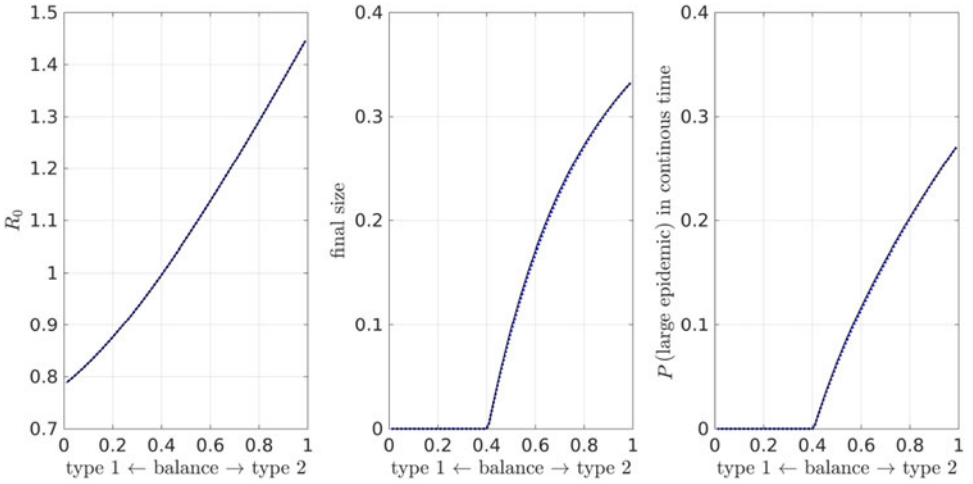

Figure 4. For the sexual network the figure shows a comparison between results from the empirical probability mass function (blue dotted lines) and results from the standard normal copula applied to the marginal distributions (solid black lines). The epidemic parameters are plotted vs r_2 on the horizontal axis. We see R_0 in the leftmost plot, the relative final size of a major outbreak in the middle plot, and the probability of major outbreak (in the Markovian setting) in the rightmost plot. All curves use the same correlation coefficient $\rho \approx -0.0652$ (from the empirical distribution) and $\mathscr{A}_1 = 1$. Note that all three curves depict two curves, but because they are so similar they appear almost as one curve. In the left figure the curves do coincide exactly.

size of a major outbreak achieves a local maximum inside the plotted region for both the weighted and the unweighted models.

To the right of the maximum, an increase in the correlation results in a decrease in the relative final size of a major outbreak. From the plot we note that changing the balance between type 1 and type 2 edges clearly affects the epidemics that develop on the graph. When the balance is strongly shifted towards type 1 edges, major outbreaks tend to be small (or not occur at all). When the balance is strongly shifted towards type 2 edges epidemics are typically large. Results for the probability of a major outbreak are similar (not shown here). Such plots thus give information on which type of edge (which type of relationship) should be targeted in order to best reduce the probability and the size of major outbreaks. Clearly, the unweighted model does not provide this information.

We also compare how the epidemic parameters are affected when using the standard normal copula to generate the bivariate degree distribution instead of using the empirical degree distribution. We do this for $\mathscr{A}_1 = 1$ and fixed correlation (the same as for the empirical distribution), and vary only the balance between the degrees of type 1 and type 2 edges.

Figure 4 is an illustration of how the balance between the two edge types affects the epidemic parameters, but the figure also shows a very close correspondence between empirical results and copula generated results. This is somewhat surprising for the relative final size of a major outbreak (center) and the probability of a major outbreak in the Markovian setting (right).

4.3.4 Sweden network

Figure 5 shows R_0 and the relative final size of a major outbreak for the Swedish population network modeled using the bivariate standard normal copula so that ρ can be varied. The result depends greatly on both the correlation and the balance between the degrees of type 1 and type 2 edges. We do not see the same variation in the epidemic parameters when looking at the unweighted plots. As mentioned before, R_0 always increases with increasing ρ (see Section 3.1).

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

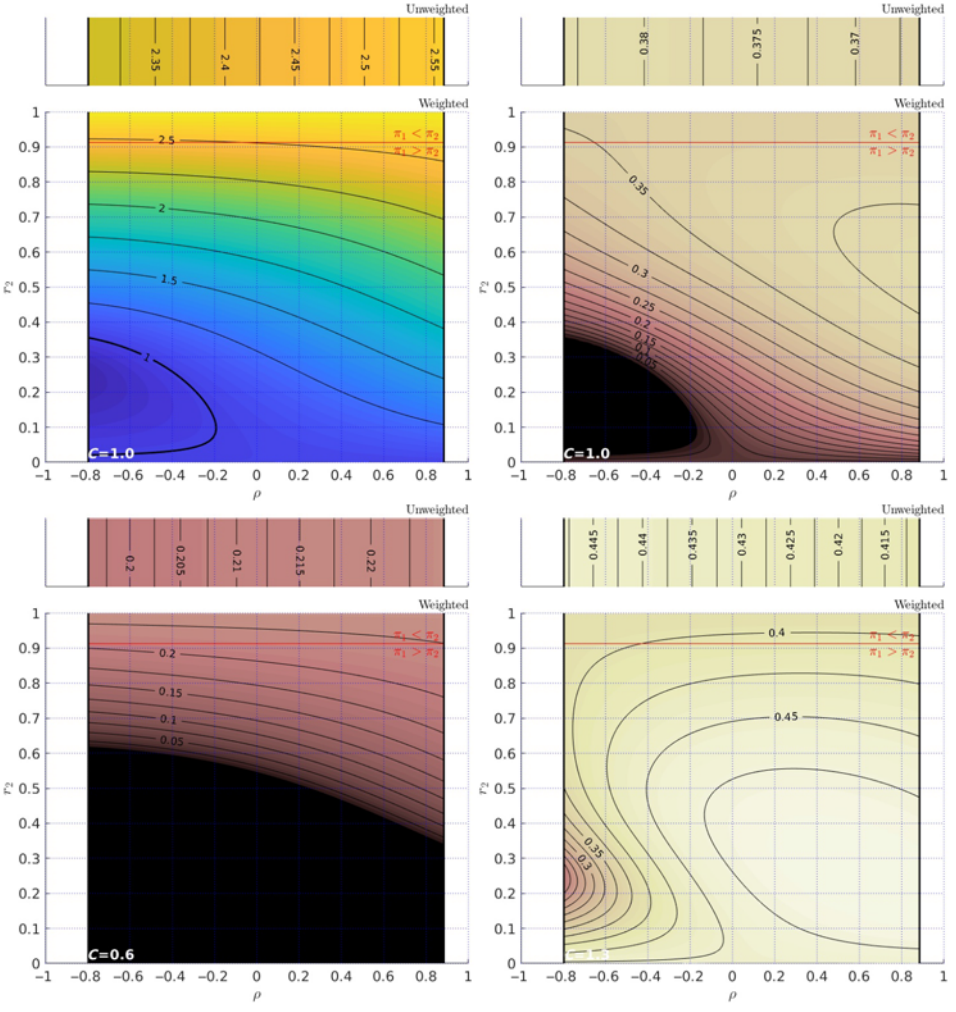

Figure 5. The figure shows contour plots of R_0 (upper left) and the relative final size of a major outbreak (the remaining three) for different values of $\mathscr{A}_\mathbf{I}$ for the Swedish population network modeled using the bivariate standard normal copula. For additional information see Figure 3.

However, from the plot where $\mathscr{A}_\mathbf{I} = 1.3$, we observe that as ρ increases the relative final size of a major outbreak decreases in the weighted model for some values of the balance parameter (e.g., follow the 0.5 balance line towards the right side of the figure). The unweighted model shows a small but steady decrease through the range of ρ. We remind the reader that the parameters for the empirical degree distribution, letting type 1 edges represent family edges and type 2 edges represent company edges, are $\mu_1 \approx 2.00$, $\mu_2 \approx 20.98$, $\sigma_1 \approx 1.44$, $\sigma_2 \approx 27.8$, and $\rho \approx -0.241$.

We again compare the results for the copula generated probability mass function with the empirical probability mass function. We do this for $\mathscr{A}_\mathbf{I} = 1$ and fixed correlation, and vary only the balance between the degrees of type 1 and type 2 edges. Results in Figure 6 show good correspondence between empirical results and copula generated results, just as for the sexual contact network (Figure 3). We also expect that the difference may be larger if $\mathscr{A}_\mathbf{I}$ is higher, since then more of the original probability mass function is preserved. We test this by setting $\mathscr{A}_\mathbf{I} = 1.3$. Results in Figure 7 indeed show a slight difference in the relative final size of a major outbreak and in the probability of a major outbreak.

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

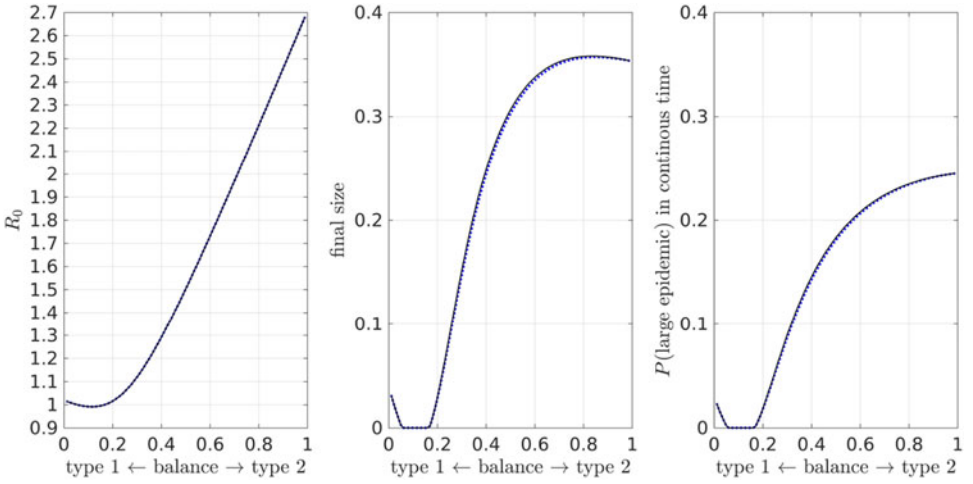

Figure 6. For the Swedish population network, the figure shows a comparison between results from the empirical probability mass function (blue dotted lines) and results from the standard normal copula applied to the marginal distributions (solid black lines). The epidemic parameters are plotted vs r_2 on the horizontal axis. All curves use the same correlation coefficient $\rho \approx -0.241$ (from the empirical distribution) and $\mathscr{A}_\mathbf{I} = 1$.

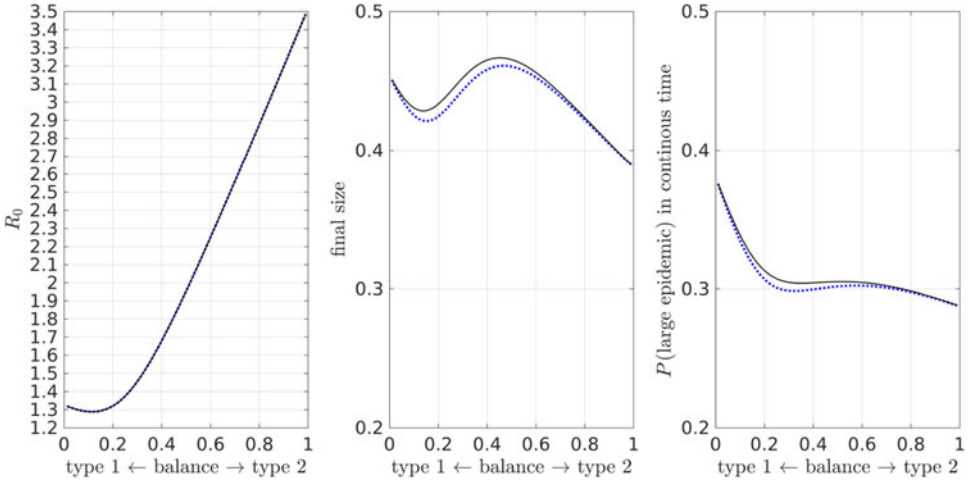

Figure 7. The same plot as in Figure 6, except for setting $\mathscr{A}_\mathbf{I} = 1.3$. The epidemic parameters are plotted vs r_2 on the horizontal axis. Some differences can be seen for this higher value of $\mathscr{A}_\mathbf{I}$.

5. Discussion

We have modeled weighted and unweighted configuration model networks based on different empirical and theoretical distributions, and have studied epidemics taking place on these networks. We show that the weighted network model produces much richer results in terms of the variation of R_0, the probability of a major outbreak, and the relative final size of a major outbreak as functions of ρ and the balance between the edge types.

We have used a parametrization that separates

a. the mean infectious activity in the network,
b. how the activity is distributed between the different edge types, and
c. the correlation between the degrees of the edge types.

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

This parametrization simplifies visualization of results. This is especially evident for R_0 plotted for $\mathscr{A}_{\mathbf{I}} = 1$ as R_0 can be obtained for other values of $\mathscr{A}_{\mathbf{I}}$ simply by multiplying the values in the plot by $\mathscr{A}_{\mathbf{I}}$. The plot gives immediate information as to which combination of parameters that can produce major outbreaks. The data needed can be obtained by only studying the so-called egocentric degree distribution of individuals sampled from the population together with estimates of the activity on different types of edges. Modeling of the network is then done through the configuration model.

The introduction of copulas allows for modeling situations outside the range of the empirical data. As an example we can use the sexual network analyzed in Section 4.3.3 for which we have a single sample with fixed correlation. By applying a bivariate standard normal copula to the marginal edge distributions, we are able to model the network for other values of the correlation ρ.

Results indicate that the standard normal copula model can produce results that are almost identical to those generated by the original network data model (when setting the same correlation in both models). Thus, if the correlation and the marginal distributions are known, the exact dependence between the degrees of the different edge types may not always be so important. The copula approach also gives some insight to what is possible for a specific set of marginal distributions. For instance, the minimum and the maximum copulas can be used to calculate which minimum and maximum correlation that is possible for the given marginal distributions.

The disadvantage with the described modeling approach is that it is fairly computer intensive when calculating the relative final size and the probability of a major outbreak. In the calculations done for this paper, the maximum degree of the degree distributions was purposely limited to reduce the computational time to a few hours (rather than days) for each figure.[3] Expanding the model to more than two edge types may require some simplifications of the model. Such simplifications may indeed be possible when the infection probabilities π_ξ are low, so that the exact dependence structure (exactly which $p_{ij} > 0$) in the empirical data is less important. Note that major outbreaks may still be possible even for low values of π_ξ. This is an effect of the configuration model which puts much emphasis on the second moment of the degree distribution and thus can result in a spread of the epidemic on high degree vertices, even when infection probabilities are low. In many such cases, the major outbreak will however be relatively small in size, since it will mainly be transmitted by nodes with high degree.

Future work could include an investigation of what simplifications of the model are possible, without affecting results significantly, and also how well the model matches real world networks. For example, if the relative final size of a major outbreak on a *finite* empirical network is comparable to the *asymptotic* relative final size of a major outbreak in the configuration model.

Funding. T.B. was supported by Vetenskapsrådet (Swedish Research Council), project 2015-05015.

Conflict of interest. The authors Kristoffer Spricer and Tom Britton have nothing to disclose.

Notes

1 Data kindly supplied by Veronika Fridlund, Department of Sociology, Stockholm University.
2 Data kindly supplied by Fredrik Liljeros, Department of Sociology, Stockholm University.
3 The calculations were performed on a Linux workstation with an Intel Core i7, 2400 MHz processor using 32 GB memory.

References

Adamic, L. A., & Huberman, B. A. (2000). Power-law distribution of the world wide web. *Science*, *287*(5461), 2115.
Bailey, N. T. J. (1975). *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE.
Ball, F., & Neal, P. (2002). A general model for stochastic sir epidemics with two levels of mixing. *Mathematical Biosciences*, *180*(1–2), 73–102.
Ball, F., & Neal, P. (2008). Network epidemic models with two levels of mixing. *Mathematical Biosciences*, *212*(1), 69–87.

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

Ball, F., & Sirl, D. (2012). An sir epidemic model on a population with random network and household structure, and several types of individuals. *Advances in Applied Probability*, *44*(1), 63–86.
Barabási, A.-L., & Bonabeau, E. (2003). Scale-free networks. *Scientific American*, *288*(5), 60–69.
Biswas, A., & Hwang, J.-S. (2002). A new bivariate binomial distribution. *Statistics & Probability Letters*, *60*(2), 231–240.
Bollobás, B. (2001). *Random graphs* (2nd ed.), Cambridge Studies in Advanced Mathematics, vol. 73. Cambridge: Cambridge University Press.
Britton, T. (2010). Stochastic epidemic models: A survey. *Mathematical Biosciences*, *225*(1), 24–35.
Britton, T., Deijfen, M., & Liljeros, F. (2011). A weighted configuration model and inhomogeneous epidemics. *Journal of Statistical Physics*, *145*(5), 1368–1384.
Britton, T., Deijfen, M., & Martin-Löf, A. (2006). Generating simple random graphs with prescribed degree distribution. *Journal of Statistical Physics*, *124*(6), 1377–1397.
Britton, T., Janson, S., & Martin-Löf, A. (2007). Graphs with specified degree distributions, simple epidemics, and local vaccination strategies. *Advances in Applied Probability*, *39*(4), 922–948.
Deijfen, M., & Fitzner, R. (2017). Birds of a feather or opposites attract - effects in network modelling. *Internet Mathematics*, *1*(1).
Diekmann, O., Heesterbeek, H., & Britton, T. (2012). *Mathematical tools for understanding infectious disease dynamics*. Princeton, NJ: Princeton University Press.
Grassberger, P. (1983). On the critical behavior of the general epidemic process and dynamical percolation. *Mathematical Biosciences*, *63*(2), 157–172.
Hansson, D., Fridlund, V., Stenqvist, K., Britton, T., & Liljeros, F. (2018). Inferring individual sexual action dispositions from egocentric network data on dyadic sexual outcomes. *Plos One*, *13*(11), e0207116.
Harris, T. E. (2002). *The theory of branching processes*. Dover Phoenix Editions. Mineola, NY: Dover Publications Inc. Corrected reprint of the 1963 original [Springer, Berlin; MR0163361 (29 #664)].
Holm, E., Lindgren, U., Lundevaller, E., & Strömgren, M. (2006). The SVERIGE spatial microsimulation model. Paper presented at the 8th Nordic Seminar on Microsimulation Models, Oslo (pp. 8–9).
Kamp, C., Moslonka-Lefebvre, M., & Alizon, S. (2013). Epidemic spread on weighted networks. *Plos Computational Biology*, *9*(12), e1003352.
Kenah, E., & Robins, J. M. (2007). Second look at the spread of epidemics on networks. *Physical Review E*, *76*(3), 036113.
Larson, J. M. (2017). The weakness of weak ties for novel information diffusion. *Applied Network Science*, **2**(1), 14.
Lefèvre, C. (1990). Stochastic epidemic models for sir infectious diseases: A brief survey of the recent general theory. In J.-P. Gabriel, C. Lefevre, & P. Picard (Eds.), *Stochastic processes in epidemic theory* (pp. 1–12). Springer-Verlag Berlin Heidelberg.
Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E., & Åberg, Y. (2001). The web of human sexual contacts. *Nature*, *411*(6840), 907.
Miller, J. C., & Volz, E. M. (2013). Incorporating disease and population structure into models of sir disease in contact networks. *Plos One*, *8*(8), e69162.
Molloy, M., & Reed, B. (1995). A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, *6*(2–3), 161–180.
Nelsen, R. B. (2007). *An introduction to copulas*. New York: Springer Science+Business Media, Inc.
Newman, M. E. J. (2002). Spread of epidemic disease on networks. *Physical Review E*, *66*(1), 016128.
Shu, P., Tang, M., Gong, K., & Liu, Y. (2012). Effects of weak ties on epidemic predictability on community networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *22*(4), 043124.

Appendix A

A.1 The bivariate binomial distribution

This bivariate binomial distribution was defined in Biswas & Hwang (2002). The distribution has five parameters, where the first four (n_1, p_1, n_2, and p_2) define the two marginal binomial distributions and the last parameter α is directly related to the correlation coefficient ρ.

The distribution is defined so that $D_\xi = \sum_{i=1}^{n_\xi} X_{\xi,i}$, for $\xi \in \{1, 2\}$. $X_{1,i}$ are independent Bernoulli distributed variables with parameter p_1. $X_{2,i}$ are independent Bernoulli distributed variables, each with parameter

$$p_{2,i} = \begin{cases} \frac{p_2+\alpha(p_2-p_1)+\alpha X_{1,i}}{1+\alpha}, & \text{if } i \le n_1, \\ p_2 & \text{if } i > n_1 \end{cases}$$

Thus, D_2 depends on D_1, but because of the specific choice of parameters, D_2 is still distributed as $\text{Bin}(n_2, p_2)$. Note that α is real valued, but with bounds that depend on the other parameters since $p_{2,i}$ is a probability and must be in the range $[0, 1]$.

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

For the analytical form of the simultaneous probability mass function, we refer the reader to Biswas & Hwang (2002). Because the marginals are binomially distributed

$$\mu_\xi = \mathrm{E}(D_\xi) = n_\xi p_\xi,$$
$$\sigma_\xi^2 = \mathrm{Var}(D_\xi) = n_\xi p_\xi (1 - p_\xi)$$

The correlation ρ can be both positive and negative, and α is closely related to it through

$$\rho = \sqrt{\frac{\min(n_1, n_2)}{\max(n_1, n_2)}} \left(\frac{\alpha}{1+\alpha}\right) \sqrt{\frac{p_1(1-p_1)}{p_2(1-p_2)}}$$

Because of the restrictions that apply to α, the range of the correlation coefficient is also restricted. This range is not correctly given in Biswas & Hwang (2002), so we give it here (there are several cases). We assume that $p_\xi > 0$ and $n_\xi > 0$ for $\xi \in \{1, 2\}$ to avoid a degenerate distribution.
The lower limit:

$$p_1 + p_2 < 1 \implies \begin{cases} \alpha \geq -\frac{p_2}{1+p_2-p1} \\ \rho \geq -\sqrt{\frac{\min(n_1,n_2)}{\max(n_1,n_2)}} \sqrt{\frac{p_1}{1-p_1} \cdot \frac{p_2}{1-p_2}} \end{cases},$$

$$p_1 + p_2 \geq 1 \implies \begin{cases} \alpha \geq -\frac{1-p_2}{1+p_1-p_2} \\ \rho \geq -\sqrt{\frac{\min(n_1,n_2)}{\max(n_1,n_2)}} \sqrt{\frac{1-p_1}{p_1} \cdot \frac{1-p_2}{p_2}} \end{cases}$$

The upper limit:

$$p_1 < p_2 \implies \begin{cases} \alpha \leq \frac{1-p_2}{p_2-p_1} \\ \rho \leq \sqrt{\frac{\min(n_1,n_2)}{\max(n_1,n_2)}} \sqrt{\frac{p_1}{1-p_1} \cdot \frac{1-p_2}{p_2}} \end{cases},$$

$$p_1 > p_2 \implies \begin{cases} \alpha \leq \frac{p_2}{p_1-p_2} \\ \rho \leq \sqrt{\frac{\min(n_1,n_2)}{\max(n_1,n_2)}} \sqrt{\frac{1-p_1}{p_1} \cdot \frac{p_2}{1-p_2}} \end{cases},$$

$$p_1 = p_2 \implies \begin{cases} \alpha \text{ no upper limit} \\ \rho \leq \sqrt{\frac{\min(n_1,n_2)}{\max(n_1,n_2)}} \end{cases}$$

It is only when $n_1 = n_2$ and $p_1 = p_2$ that ρ can take on any value in the range $[-1, 1]$.

Appendix B

B.1 Modeling distributions through copulas

Copulas define the correlation structure between random variables with given marginal distributions. They can be defined for any number of variables, but we limit the scope to bivariate distributions. Let X and Y be two random variables with given (marginal) distributions $F_X(x)$ and $F_Y(y)$. We define the simultaneous distribution $F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$ through a copula $C(u, v)$ as

$$F_{X,Y}(x, y) = C(F_X(x), F_Y(y))$$

The function $C(u, v)$ can be viewed as the simultaneous distribution function for two uniformly distributed random variables on [0,1]. The *Fréchet-Hoeffding Bounds* (see Nelsen, 2007, p. 9) are

$$M(u, v) \leq C(u, v) \leq W(u, v) \; \forall \, u, v \tag{B1}$$

where

$$M(u, v) = \min(u, v),$$
$$W(u, v) = \max(u+v-1, 0)$$

In our application we are mainly interested in using the copula to vary the correlation between the variables, while maintaining the marginal distributions. We thus limit our study to copulas that are uniquely defined by the correlation ρ and indicate this by writing $C_\rho(u, v)$.

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

Noting that

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

using *Hoeffding's Identity* (e.g., see Nelsen, 2007, p. 154)

$$\mathrm{Cov}(X, Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left(F_{X,Y}(x, y) - F_X(x)F_Y(y)\right)\, \mathrm{d}x\, \mathrm{d}y$$

and using Equation (B1) we see that $\mathrm{Cov}(X, Y)$ has bounds that can be found by applying $W(u, v)$ (for the lower bound) and $M(u, v)$ (for the upper bound). We use these bounds when calculating and presenting results in Section 4. At the bounds the copulas are unique and results obtained by using the copulas are also unique. However, between the two extremes the copula that results in a specific correlation coefficient is not unique, and thus results obtained from the simultaneous distribution function are also not necessarily unique.

In our application we want ρ to span as large a range as possible. Not all copulas are able to span the maximum range, but one that is able is the bivariate normal copula. It is defined as

$$C_\rho(u, v) = N_\rho(\Phi^{-1}(u), \Phi^{-1}(v))$$

where $\Phi(x)$ and $N_\rho(x, y)$ are the univariate and the bivariate standard normal distributions, respectively.

In our application we work with empirical bivariate degree distributions. From these we can extract all parameters for the distribution, such as means, variances, and the correlation, but we are not able to vary the correlation coefficient. However, taking the marginal distributions and applying a copula, such as the standard normal copula, enable us to vary the correlation coefficient.

As is mentioned above the choice of copula will affect the results (except at the boundaries of the maximum allowed range for ρ), so care is needed when drawing conclusions based on the results from the use of (arbitrary) copulas.

It is important to note that the ρ that is used in the definition of the copula itself is not the same as the correlation coefficient for the resulting bivariate degree distribution. The two marginal degree distributions also affect the amount of correlation in the resulting distribution. In the comparisons in this paper between a given bivariate degree distribution and the corresponding one obtained using the standard normal copula, care has been taken to closely match the correlation coefficients of the final bivariate distribution by properly adjusting ρ in the definition of the copula.

Appendix C

C.1 Branching processes

Here we only give a brief overview of the part of the theory that we need for this paper. For a more thorough treatment of the theory of branching processes subject, we refer the reader to Harris (2002).

Early on in the epidemic the growth of the number of infected vertices can be modeled through a branching process. In our application the branching process starts with a single infected individual, the index case, and we study how the number of new infected vertices develops in each generation. Unfortunately, the degree distribution is different for the index case and for the subsequent generations. In this section we deal with the most simple form, where we assume that the same degree distribution is valid through the entire branching process. This is still applicable in our model for all epidemic generations after the index case. Given that we understand how the branching process develops for all future generations, it is then an easy task to include the index case in the model. In our application we only need to obtain the probability that the branching process goes extinct. This same quantity can be used to calculate both the asymptotic probability of a major outbreak and the relative final size of it.

Below we will talk of the *offspring* of an individual and in our application this corresponds to the number of people that the individual infects. We start with the unweighted model where there is only one edge type. Then we continue with the weighted model where we have two different types of edges that the epidemic can spread through. Then we need to take into account through which type of edge an individual was infected and also through how many edges of each type the epidemic continues.

We begin with a model with only one type of individual (corresponding to the unweighted configuration model). We study the number of individuals Z_i in each generation $i = 0, 1, \ldots$. We start with a single individual, so $Z_0 = 1$. Each new generation then consists of the offspring of the individuals in the previous generation so that $Z_i = \sum_{j=1}^{Z_{i-1}} X_{i,j}$, where $X_{i,j}$ (the offspring of individual j in generation i) are all independent and identically distributed $X_{i,j} \sim X$. The offspring distribution is defined by $p_k = \mathbb{P}(X = k)$. In the following we will assume that $\mathrm{E}(X) < \infty$ and that $p_k < 1$ for $k = 1, 2$. One important property of a branching process is the probability that it dies out, i.e., that $Z_i = 0$ for some i. Through the probability generating function

$$f(s) = \sum_{i=0}^{\infty} s^i p_k$$

https://doi.org/10.1017/nws.2019.54 Published online by Cambridge University Press

we obtain the probability of extinction q as the smallest nonnegative solution to the fixed point equation

$$q = f(q)$$

Note that there can be at most one solution $q \in [0, 1)$ and that there is always a solution $q = 1$.

We now continue with a multi-type model which corresponds to the weighted configuration model, where there are different types of edges. We restrict the analysis to two types of edges and in effect study how the number of infections via each type of *edge* develops. Let $p_\xi(i, j)$ be the probability that an individual of type $\xi \in \{1, 2\}$ has offspring i individuals of type 1 and j individuals of type 2. Some assumptions are needed and these correspond to the requirements for the single type branching process described above. We assume that the expected number of offspring is finite. In addition, we assume that the process is not *singular* and that it is *positively regular*, for definitions see Harris (2002). In our application positively regular means that, regardless of which edge type we start with, at some time in the future the process is able to produce the other edge type. In turn, this means that at least one $p_\xi(i, j) > 0$ for some $i, j > 0$. If this is not true, then the graph can be separated into subgraphs, each one consisting only of vertices connected by a single edge type. Such separate configuration model graphs are not within the scope of this paper.

Further, let $\mathbf{s} = (s_1, s_2)$ and define the probability generating functions

$$f_\xi(\mathbf{s}) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} s_1^i s_2^j p_\xi(i, j)$$

The probability q_ξ that the branching process dies out, *given that we start with an individual of type* ξ, is a solution to the fixed-point equation

$$\mathbf{q} = (f_1(\mathbf{q}), f_2(\mathbf{q})) \tag{C1}$$

where $\mathbf{q} = (q_1, q_2)$. If there exists a solution $\mathbf{q} \in [0, 1)^2$ (and there can be at most one such solution), then this is the correct solution. Otherwise the solution is $\mathbf{q} = (1, 1)$. As above, Equation (C1) can be solved iteratively

(1) Let $\mathbf{q}^{(0)} = (0, 0)$.
(2) Let $\mathbf{q}^{(k)} = (f_1(\mathbf{q}^{(k-1)}), f_2(\mathbf{q}^{(k-1)}))$, for $k = 1, 2, \ldots$
(3) The probability that the process dies out is $\mathbf{q} = \lim_{k\to\infty} \mathbf{q}^{(k)}$

Note that $\mathbf{q}^{(k)}$ gives the probability that the branching process dies out within k generations.

The quantities q and q_ξ are used in Sections 3.2 and 3.3.

Cite this article: Spricer K., and Britton T. (2019). An SIR epidemic on a weighted network. *Network Science* **7**, 556–580. https://doi.org/10.1017/nws.2019.54