ANZIAM J. 57(2016), 429–444
doi:10.1017/S1446181116000043

EXACT ANALYTICAL EXPRESSIONS FOR THE FINAL EPIDEMIC SIZE OF AN SIR MODEL ON SMALL NETWORKS

K. MCCULLOCH✉[1], M. G. ROBERTS[1] and C. R. LAING[1]

(Received 31 July, 2015; accepted 27 October, 2015; first published online 23 May 2016)

Abstract

We investigate the dynamics of a susceptible infected recovered (SIR) epidemic model on small networks with different topologies, as a stepping stone to determining how the structure of a contact network impacts the transmission of infection through a population. For an SIR model on a network of N nodes, there are 3^N configurations that the network can be in. To simplify the analysis, we group the states together based on the number of nodes in each infection state and the symmetries of the network. We derive analytical expressions for the final epidemic size of an SIR model on small networks composed of three or four nodes with different topological structures. Differential equations which describe the transition of the network between states are also derived and solved numerically to confirm our analysis. A stochastic SIR model is numerically simulated on each of the small networks with the same initial conditions and infection parameters to confirm our results independently. We show that the structure of the network, degree of the initial infectious node, number of initial infectious nodes and the transmission rate all significantly impact the final epidemic size of an SIR model on small networks.

2010 *Mathematics subject classification*: primary 60J28; secondary 34A30, 34F05.

Keywords and phrases: spread of infection, contact network, probability mass functions, network topology, clustering coefficient.

1. Introduction

The use of contact networks to model the spread of an infection through a population has become increasingly popular in recent years [2, 4, 12, 16, 21]. It is well known that models for the spread of an infection which assume that the population is homogeneously mixed are only appropriate for some situations. Such models often overestimate the proportion of the population which becomes infected during an epidemic; this overestimation occurs when there is heterogeneity in the number of

[1]Institute of Natural and Mathematical Sciences, Massey University, Albany, 0745, New Zealand; e-mail: K.McCulloch@latrobe.edu.au.

contacts each individual has within the population. Pair approximation models [6, 7] improve on the accuracy of the homogeneous models by focusing on the interaction between pairs of individuals and assume that a change in the state of an individual depends on the state of its neighbours. However, contact networks provide more detailed model populations; connections between individuals can be permanent (static networks) or the number and duration of connections each individual has may change over time (dynamic networks). In a contact network, each node represents an individual (or group of individuals), and an edge between two nodes represents a connection between individuals. A connection is defined depending on the context; for example, an individual within a shared office would have connections between all other individuals within the same office for the transmission of influenza, or a connection could represent an individual's partners in a sexual contact network for HIV. There has been considerable effort put into understanding how an infection spreads through different network structures such as scale-free, small-world, lattice and random networks [3, 4, 8, 12, 20, 25–27]. However, there is still a lack of models which capture the population structure in enough detail to provide more accurate predictions, while ensuring that the mathematical methods are tractable, so that the derivation of analytical results is possible.

There have been a number of different models for the spread of infection developed with the aim of addressing this issue of incorporating enough detailed information, yet keeping the mathematical methods tractable and in a form which is able to be understood and analysed. There is often a trade off between model complexity and mathematical tractability; thus, most of the models developed so far are tailored to specific assumptions about the population structure or infection dynamics. For example, consider the effective degree network models, where nodes are classified by their infection state and by the infection state of their neighbours. These models produce accurate predictions for SIR (susceptible infected recovered) and SIS (susceptible infected susceptible) type models on large random static [1, 5, 11] and dynamic [14, 24] networks without clustering. Analytical expressions for the basic reproduction ratio, R_0, were found for effective degree models on static networks, and approximations to the final proportion of the population infected were found to be in agreement with stochastic simulations.

Other research has been focused on finding results for the spread of infection through networks by breaking up the networks into smaller specific motifs [10, 22, 23]. These models demonstrate the importance of understanding higher-order structure (such as larger network motifs), by comparing the spread of an infection between networks with the same degree distribution and clustering coefficient. Because these methods are mathematically difficult to analyse, results are obtained via numerical simulation. Edge-based compartmental models (EBCMs) [15, 17, 18, 25] begin to incorporate model complexity while keeping the mathematical concept and derivation tractable. EBCMs are based on the idea of determining the probability that a given node's neighbour is susceptible, infectious or recovered at time t. These models incorporate social heterogeneity and variation in partnership duration by investigating

https://doi.org/10.1017/S1446181116000043 Published online by Cambridge University Press

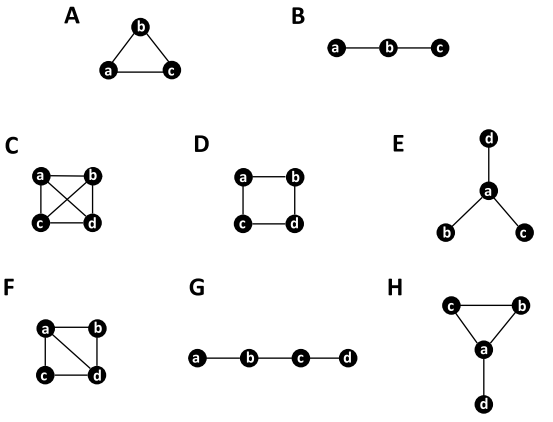

FIGURE 1. Network diagrams in order of increasing complexity. (**A**) Triangle network. (**B**) Line network, $N = 3$. (**C**) Complete network, $N = 4$. (**D**) Square network. (**E**) Star network. (**F**) Toast network. (**G**) Line network, $N = 4$. (**H**) Lollipop network.

an SIR model on static and dynamic networks. EBCMs accurately predict the time evolution of the SIR model and provide a means for finding an analytical expression for the expected final epidemic size (in most cases) and the basic reproduction ratio, R_0. Miller and Kiss [16] and Pellis et al. [21] have given detailed overviews of what has previously been done to analyse epidemic spread in networks, and have outlined some of the future challenges. It is well known that considering only the degree distribution of a network is not enough to accurately predict the outcome of an epidemic [13], since two networks which have the same degree distribution can exhibit different epidemic outcomes due to topological differences. It has also been shown that networks with the same degree distribution and clustering coefficient can exhibit very different epidemic behaviour [22]. There are many properties which describe the structure of a network such as the above-mentioned degree distribution and clustering coefficient [19]. The purpose of this paper is to start addressing the question of which network properties matter in terms of the spread of infection and whether these properties differ between network structures. We focus on an SIR-type compartmental model to describe the spread of an infection through a population. We consider eight small networks of varying topological structure and investigate the dynamics of the model on them (Figure 1).

This paper is structured as follows. In Section 2, we outline the methods used to find the analytical probability mass function (PMF) of the final epidemic size for each network. In Section 3, we illustrate our methods with a line network of three nodes. In Section 4, we briefly describe how we investigate the SIR model on the remaining networks of three and four nodes, followed by the presentation of our results. In Section 5, we discuss our results. In the supplementary material (Appendix A), we explain and illustrate the output of the stochastic simulations which were

https://doi.org/10.1017/S1446181116000043 Published online by Cambridge University Press

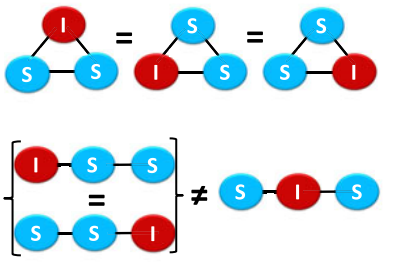

FIGURE 2. Example of grouping together topologically equivalent states for networks of size $N = 3$. The three triangle network states (top) are equivalent and thus can be grouped together into one state. However, for the line network (bottom) the two states on the left, where the end node is infectious, are equivalent and thus are grouped together to form one state. The state on the bottom right, where the central node of the line network is infectious, is not topologically equivalent and so it forms another state of its own.

used to confirm our analytical results. The supplementary material (Appendix B) also contains the detailed methods for each of the remaining seven small networks shown in Figure 1.

2. Methods

In this section, we define the notation used throughout the paper and outline the methods used. Susceptible nodes acquire infection at rate β per S–I edge and infectious nodes recover at rate γ. As all of our results depend on the ratio β/γ, we define the variable $\mathcal{R} = \beta/\gamma$ to simplify our expressions. We denote the state that a network of three nodes is in by XYZ, where X, Y and Z denote the infection state (S, I or R) that nodes a, b and c are in, respectively. Similarly, we denote the state that a network of four nodes is in by $WXYZ$. For an SIR model on a network of N nodes, there are 3^N possible states which the network can be in. Thus, for an SIR model on networks of sizes $N = 3$ and $N = 4$ nodes, there are 27 and 81 possible states which the networks can be in, respectively. To simplify our analysis, we group states together based on symmetries of the network [23]. In the most simple case of a complete network, we group states together that have the same number of susceptible (S), infected (I) and recovered (R) nodes (a complete network is a network in which each node is connected to every other node within the network). For networks which are not complete, we must look at the topology and the number of nodes in each infection state before determining which network states can be grouped together. Thus, for each network we group together the states that are topologically equivalent and have the same number of nodes in each infection state. See Figure 2 for illustrations of topologically equivalent states. Once we have reduced the number of states, we draw a transition diagram which shows how the infection moves through the network.

From the transition diagrams, we can calculate the probability of obtaining different final sizes, which depends only on the initial state and the infection parameters. The final size is the total number of nodes which acquire infection at some point

https://doi.org/10.1017/S1446181116000043 Published online by Cambridge University Press

during the epidemic. For an SIR-type model, this is simply the number of nodes which end up in the recovered state. We obtain these final sizes by summing up all possible infection paths to the appropriate absorbing states, given the infection starts in one of the initial states. Once the network reaches an absorbing state, the epidemic is over and the infection has died out. The analytical expression for the expected final size of the epidemic is also determined. The expected final size is defined by $\mathbb{E}[\text{Final size}] = \sum_{i=1}^{N} i\mathbb{P}(\text{Final size} = i)$. We can also show how the infection progresses through the network by writing down a set of differential equations; each equation describes the evolution of the probability that the network is in a given state at time t. We then numerically solve the equations with the same initial conditions and infection parameters which we used to find the analytical expressions of the probability mass function for the final epidemic size. To independently verify our results, we also ran stochastic simulations on each of the small networks for the same initial conditions and infection parameter ($\mathcal{R}$). The Gillepsie algorithm was used for the stochastic SIR model (see [9]), which was implemented using MATLAB software. To calculate the final size probabilities, we ran 2×10^5 stochastic simulations on each network with the initial conditions and parameters as specified, and counted the frequency of each final size occurring. We also calculated the clustering coefficient, ϕ, for each network to see if any patterns emerge between the final epidemic size and clustering coefficient of the networks.

3. Line network

The triangle network is the simplest of the eight small networks we investigated, being the smallest complete network. However, in order to demonstrate how the network structure impacts the spread of an infection, we illustrate our methods with the line network of three nodes. Details of the triangle network and networks of size $N = 4$ are given in Appendix B. For a line network with three nodes, we have two nodes of degree 1 (a and c) and one node of degree 2 (b) (see Figures 1(B) and 2). For an SIR model on a line network of three nodes, there are 27 possible states in which the network can be. By grouping the appropriate states together, the line network can be reduced to 18 sets of states shown by the transition diagram in Figure 3. In the line network we have three different initial states; two initial states have one infectious node, SSI (where nodes a, b and c are in the susceptible, susceptible and infectious classes, respectively) and SIS (where nodes a, b and c are in the susceptible, infectious and susceptible classes, respectively), and one initial state has two infectious nodes, ISI (where nodes a, b and c are in the infectious, susceptible and infectious classes, respectively). Initial states cannot be reached from any other state in the transition diagram. In states SSI and SIS, the infectious node has degrees 1 and 2, respectively.

Each box in the transition diagram represents a different state of the network, and each arrow represents either recovery or infection of a node. Transition between states is only possible if there is a directed arrow from one to the other. Movement between states in the transition diagram shows how the infection spreads through the network.

https://doi.org/10.1017/S1446181116000043 Published online by Cambridge University Press

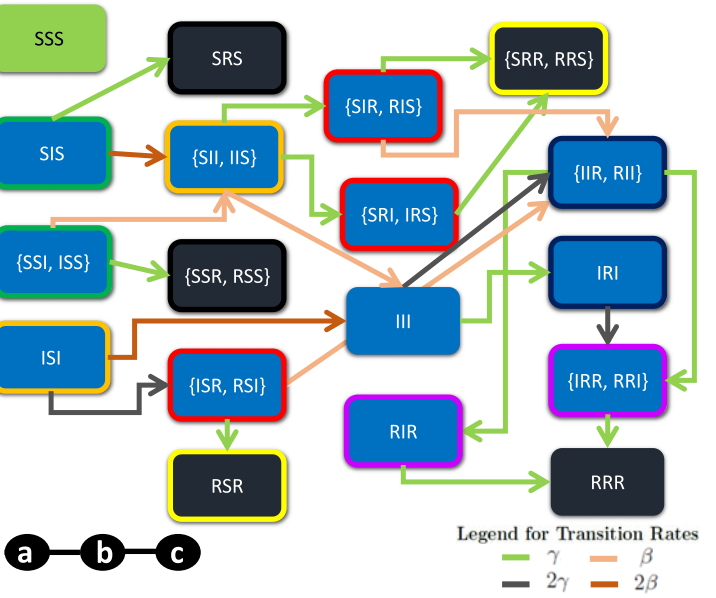

FIGURE 3. Transition diagram showing how an infection can spread through the line network. Possible initial states SIS, SSI and ISI are shown on the left-hand side. The absorbing states for this process are coloured in black. Coloured outlines identify states with the same number of nodes in each infection state but different network configurations (colour available online).

Once the network reaches an absorbing state, the epidemic is over and the infection has died out. Refer to Appendix B for the transition diagrams of the other small networks shown in Figure 1.

3.1. Catalogue of transition probabilities From the transition diagram of the SIR model on the line network, we derive the individual transition probabilities between network states. In the following, $\mathcal{P}_{XYZ}$ denotes the probability that the network is in state XYZ. These probabilities are independent of time and depend only on the infection parameters. From the absorbing state probabilities, we find the final epidemic size probabilities, $\mathbb{P}$.

Possible initial state indicator variables:

$$\mathrm{E}_{\mathrm{SIS}} = \begin{cases} 1 & \text{if initial state is SIS,} \\ 0 & \text{otherwise,} \end{cases}$$

$$\mathrm{E}_{\mathrm{SSI}} = \begin{cases} 1 & \text{if initial state is SSI,} \\ 0 & \text{otherwise,} \end{cases}$$

$$\mathrm{E}_{\mathrm{ISI}} = \begin{cases} 1 & \text{if initial state is ISI,} \\ 0 & \text{otherwise.} \end{cases}$$

https://doi.org/10.1017/S1446181116000043 Published online by Cambridge University Press

Probability of passing through transient states:

$$\mathcal{P}_{\mathrm{SII}} = \frac{2\mathcal{R}}{2\mathcal{R}+1}\mathrm{E}_{\mathrm{SIS}} + \frac{\mathcal{R}}{\mathcal{R}+1}\mathrm{E}_{\mathrm{SSI}},$$

$$\mathcal{P}_{\mathrm{ISR}} = \frac{2}{2\mathcal{R}+2}\mathrm{E}_{\mathrm{ISI}},$$

$$\mathcal{P}_{\mathrm{SIR}} = \frac{1}{\mathcal{R}+2}\mathcal{P}_{\mathrm{SII}},$$

$$\mathcal{P}_{\mathrm{SRI}} = \frac{1}{\mathcal{R}+2}\mathcal{P}_{\mathrm{SII}},$$

$$\mathcal{P}_{\mathrm{III}} = \frac{\mathcal{R}}{\mathcal{R}+2}\mathcal{P}_{\mathrm{SII}} + \frac{2\mathcal{R}}{2\mathcal{R}+2}\mathrm{E}_{\mathrm{ISI}},$$

$$\mathcal{P}_{\mathrm{IIR}} = \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{\mathrm{SIR}} + \frac{2}{3}\mathcal{P}_{\mathrm{III}} + \frac{\mathcal{R}}{\mathcal{R}+1}\mathcal{P}_{\mathrm{ISR}},$$

$$\mathcal{P}_{\mathrm{IRI}} = \tfrac{1}{3}\mathcal{P}_{\mathrm{III}},$$

$$\mathcal{P}_{\mathrm{RIR}} = \tfrac{1}{2}\mathcal{P}_{\mathrm{IIR}},$$

$$\mathcal{P}_{\mathrm{IRR}} = \mathcal{P}_{\mathrm{IRI}} + \tfrac{1}{2}\mathcal{P}_{\mathrm{IIR}}.$$

Probability of terminating in absorbing states:

$$\mathcal{P}_{\mathrm{SRS}} = \frac{1}{2\mathcal{R}+1}\mathrm{E}_{\mathrm{SIS}},$$

$$\mathcal{P}_{\mathrm{SSR}} = \frac{1}{\mathcal{R}+1}\mathrm{E}_{\mathrm{SSI}},$$

$$\mathcal{P}_{\mathrm{RSR}} = \frac{1}{\mathcal{R}+1}\mathcal{P}_{\mathrm{ISR}},$$

$$\mathcal{P}_{\mathrm{SRR}} = \frac{1}{\mathcal{R}+1}\mathcal{P}_{\mathrm{SIR}} + \mathcal{P}_{\mathrm{SRI}},$$

$$\mathcal{P}_{\mathrm{RRR}} = \mathcal{P}_{\mathrm{RIR}} + \mathcal{P}_{\mathrm{IRR}}.$$

To find the equations for the final size probabilities, we evaluated the following:

$$\mathbb{P}(\text{Final size} = 1) = \mathcal{P}_{\mathrm{SRS}} + \mathcal{P}_{\mathrm{SSR}},$$

$$\mathbb{P}(\text{Final size} = 2) = \mathcal{P}_{\mathrm{RSR}} + \mathcal{P}_{\mathrm{SRR}},$$

$$\mathbb{P}(\text{Final size} = 3) = \mathcal{P}_{\mathrm{RRR}}.$$

Simplifying the above, we derive three sets of analytical expressions for the final size probabilities (see Table 1), which reflect the different initial conditions possible for the line network. For example, to illustrate how we found $\mathbb{P}(\text{Final size} = 2)$, we make the

https://doi.org/10.1017/S1446181116000043 Published online by Cambridge University Press

TABLE 1. PMFs for the final size (FS).

Initial state	SSI	SIS	ISI
$\mathbb{P}(\text{FS} = 1)$	$\frac{1}{\mathcal{R}+1}$	$\frac{1}{2\mathcal{R}+1}$	0
$\mathbb{P}(\text{FS} = 2)$	$\frac{\mathcal{R}}{(\mathcal{R}+1)^2}$	$\frac{2\mathcal{R}}{(\mathcal{R}+1)(2\mathcal{R}+1)}$	$\frac{1}{(\mathcal{R}+1)^2}$
$\mathbb{P}(\text{FS} = 3)$	$\frac{\mathcal{R}^2}{(\mathcal{R}+1)^2}$	$\frac{2\mathcal{R}^2}{(\mathcal{R}+1)(2\mathcal{R}+1)}$	$\frac{\mathcal{R}(\mathcal{R}+2)}{(\mathcal{R}+1)^2}$
Expected FS	$\frac{3\mathcal{R}^2+3\mathcal{R}+1}{(\mathcal{R}+1)^2}$	$\frac{3\mathcal{R}+1}{\mathcal{R}+1}$	$\frac{3\mathcal{R}^2+6\mathcal{R}+2}{(\mathcal{R}+1)^2}$

following substitutions:

$$\begin{aligned}
\mathbb{P}(\text{Final size} = 2) &= \mathcal{P}_{\text{RSR}} + \mathcal{P}_{\text{SRR}} \\
&= \frac{1}{\mathcal{R}+1}\mathcal{P}_{\text{ISR}} + \frac{1}{\mathcal{R}+1}\mathcal{P}_{\text{SIR}} + \mathcal{P}_{\text{SRI}} \\
&= \frac{2}{(\mathcal{R}+1)(2\mathcal{R}+2)}\text{E}_{\text{ISI}} + \frac{1}{(\mathcal{R}+1)(\mathcal{R}+2)}\mathcal{P}_{\text{SII}} + \frac{1}{\mathcal{R}+2}\mathcal{P}_{\text{SII}} \\
&= \frac{1}{(\mathcal{R}+1)^2}\text{E}_{\text{ISI}} + \frac{2\mathcal{R}(\mathcal{R}+2)}{(\mathcal{R}+1)(\mathcal{R}+2)(2\mathcal{R}+1)}\text{E}_{\text{SIS}} \\
&\quad + \frac{\mathcal{R}(\mathcal{R}+2)}{(\mathcal{R}+1)^2(\mathcal{R}+2)}\text{E}_{\text{SSI}} \\
&= \frac{1}{(\mathcal{R}+1)^2}\text{E}_{\text{ISI}} + \frac{2\mathcal{R}}{(\mathcal{R}+1)(2\mathcal{R}+1)}\text{E}_{\text{SIS}} + \frac{\mathcal{R}}{(\mathcal{R}+1)^2}\text{E}_{\text{SSI}}.
\end{aligned}$$

The probability mass function of the final size distribution for the line network is shown in Figure 7. We independently verify our final size calculations using a stochastic SIR model. See Appendix A for an illustration of the stochastic simulations on the line network with $N = 3$ nodes.

3.2. Progression of infection over time In the following, we use P_{XYZ} to denote the probability that the line network is in the state XYZ at time t. Thus, the equation for the time derivative, $\dot{P}_{XYZ}$, shows how the network can enter and leave the state XYZ. The rate the network enters and leaves each state can be found from the transition diagram. These equations allow us to simulate the time course of the epidemic and to check our final size calculations. Equations describing the probability that the network

https://doi.org/10.1017/S1446181116000043 Published online by Cambridge University Press

is in a given state at time t for an SIR model on the line network are

initial states:

$$\dot{P}_{\mathrm{SSS}} = 0, \tag{3.1}$$
$$\dot{P}_{\mathrm{SIS}} = -(2\mathcal{R}+1)P_{\mathrm{SIS}}, \tag{3.2}$$
$$\dot{P}_{\mathrm{SSI}} = -(\mathcal{R}+1)P_{\mathrm{SSI}}, \tag{3.3}$$
$$\dot{P}_{\mathrm{ISI}} = -2(\mathcal{R}+1)P_{\mathrm{ISI}}, \tag{3.4}$$

transient states:

$$\dot{P}_{\mathrm{SII}} = 2\mathcal{R}P_{\mathrm{SIS}} + \mathcal{R}P_{\mathrm{SSI}} - (\mathcal{R}+2)P_{\mathrm{SII}}, \tag{3.5}$$
$$\dot{P}_{\mathrm{ISR}} = 2P_{\mathrm{ISI}} - (\mathcal{R}+1)P_{\mathrm{ISR}}, \tag{3.6}$$
$$\dot{P}_{\mathrm{III}} = \mathcal{R}P_{\mathrm{SII}} + 2\mathcal{R}P_{\mathrm{ISI}} - 3P_{\mathrm{III}}, \tag{3.7}$$
$$\dot{P}_{\mathrm{SIR}} = P_{\mathrm{SII}} - (\mathcal{R}+1)P_{\mathrm{SIR}}, \tag{3.8}$$
$$\dot{P}_{\mathrm{SRI}} = P_{\mathrm{SII}} - P_{\mathrm{SRI}}, \tag{3.9}$$
$$\dot{P}_{\mathrm{IIR}} = \mathcal{R}P_{\mathrm{SIR}} + \mathcal{R}P_{\mathrm{ISR}} + 2P_{\mathrm{III}} - 2P_{\mathrm{IIR}}, \tag{3.10}$$
$$\dot{P}_{\mathrm{IRI}} = P_{\mathrm{III}} - 2P_{\mathrm{IRI}}, \tag{3.11}$$
$$\dot{P}_{\mathrm{RIR}} = P_{\mathrm{IIR}} - P_{\mathrm{RIR}}, \tag{3.12}$$
$$\dot{P}_{\mathrm{IRR}} = (P_{\mathrm{IIR}} + 2P_{\mathrm{IRI}}) - P_{\mathrm{IRR}}, \tag{3.13}$$

absorbing states:

$$\dot{P}_{\mathrm{SSR}} = P_{\mathrm{SSI}}, \tag{3.14}$$
$$\dot{P}_{\mathrm{SRS}} = P_{\mathrm{SIS}}, \tag{3.15}$$
$$\dot{P}_{\mathrm{RSR}} = P_{\mathrm{ISR}}, \tag{3.16}$$
$$\dot{P}_{\mathrm{SRR}} = P_{\mathrm{SIR}} + P_{\mathrm{SRI}}, \tag{3.17}$$
$$\dot{P}_{\mathrm{RRR}} = P_{\mathrm{IRR}} + P_{\mathrm{RIR}}. \tag{3.18}$$

We have included the equation for the initial state SSS for completeness, even though it is disjoint from the transition diagram as no infection is present. As $t \to \infty$, the infection will die out and the system will end up in one of the absorbing states. For a given transition rate, the initial state determines the probability of tending to each of the possible absorbing states. To find the final size probabilities, we numerically solve the system of differential equations with specified initial conditions and sum up the appropriate absorbing state probabilities once a steady state has been reached. Figure 4 shows the numerical results for solving the system of differential equations (3.1)–(3.18). Numerical results confirm our analytical expressions for the final epidemic size.

4. Epidemics on networks of three or four nodes

We analyse an SIR model on each of the small networks of three and four nodes (see Figure 1) following the method illustrated above shown in Section 3 for the line

https://doi.org/10.1017/S1446181116000043 Published online by Cambridge University Press

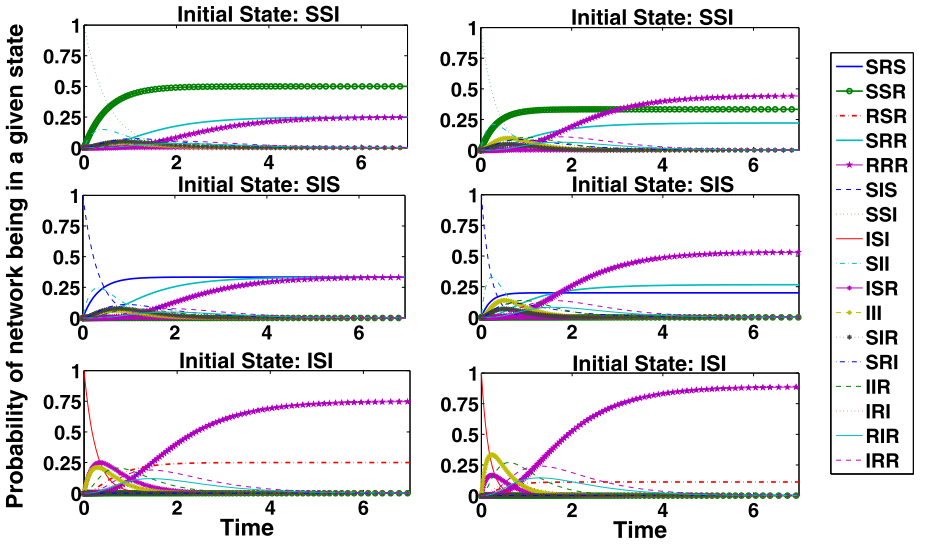

FIGURE 4. Numerical results for solving the system of differential equations (3.1)–(3.18) which describes the progression of infection over time for an SIR model on a line network with $N = 3$ nodes. Left and right columns contain graphical results for $\mathcal{R} = 1$ and $\mathcal{R} = 2$, respectively, for the specified initial conditions. The numerical results are in agreement with the analytical expressions for the same set of initial conditions (colour available online).

network of three nodes. Here, we present the results from all networks considered and refer the reader to Appendix B for the full details of the analysis for each network. Figures 5 and 6 show plots of the expected final size for all networks of sizes $N = 3$ and $N = 4$, respectively, over a range of values for $\mathcal{R}$.

It can be seen clearly in Figure 5 that the expected final size for the triangle network is always higher than the expected final size for the line network with $N = 3$ nodes starting with one infectious node. Similarly, in Figure 6, the expected final size for the complete network is always higher than every other network of size $N = 4$ for the specified range of $\mathcal{R}$ values. In Table 2, we give expressions for the expected final size of each of the small networks with specified initial conditions. In Figure 7, we show the probability mass function for the final epidemic size of each network considered. For each network there is a probability mass function corresponding to each initial state. In Table 3, we give the expected final size corresponding to each of the probability mass functions in Figure 7. Table 3 also shows the degree of the initial infectious node(s) and clustering coefficient for each of the small networks of three and four nodes considered. Not surprisingly, the triangle network has a clustering coefficient of $\phi = 1$ and all complete networks have a clustering coefficient equal to 1. All line networks have a clustering coefficient of $\phi = 0$, as there are no closed loops of any size in a line network. Note that the square network has a closed loop of order four, which is not included in the derivation of the clustering coefficient.

https://doi.org/10.1017/S1446181116000043 Published online by Cambridge University Press

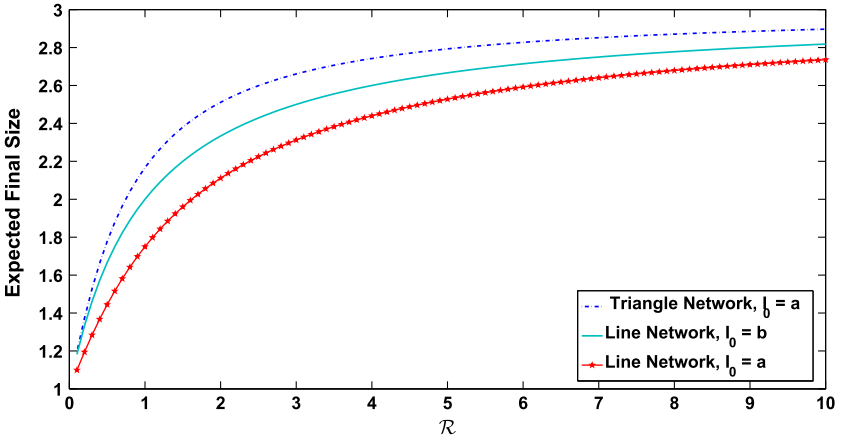

FIGURE 5. Expected final size functions of $\mathcal{R}$ for all networks of size $N = 3$; these were evaluated for the range of values $\mathcal{R} = 0.1$ to $\mathcal{R} = 10$. Each SIR epidemic started with one infectious node, $\mathbf{I}_0$, as specified (colour available online).

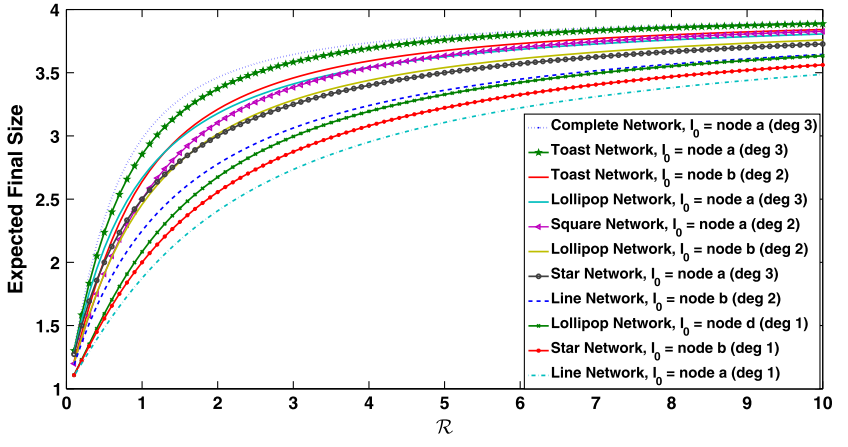

FIGURE 6. Expected final size functions of $\mathcal{R}$ for all networks of size $N = 4$; these were evaluated for the range of values $\mathcal{R} = 0.1$ to $\mathcal{R} = 10$. Each SIR epidemic started with one infectious node, $\mathbf{I}_0$, as specified (colour available online).

We have derived exact analytical expressions for the probability mass functions of the final epidemic size on eight small networks and investigated the effect that network structure and the degree of the initial infectious node has on the spread of an infection. Increasing the complexity of the network structure reduced the effect of grouping states together based on symmetries of the network. Consequently, this increased the complexity of the analysis, which is evident in the probability mass function expressions and is due to the increasing number of possible infection paths. The results presented here form the basis for finding tractable analytical results which

https://doi.org/10.1017/S1446181116000043 Published online by Cambridge University Press

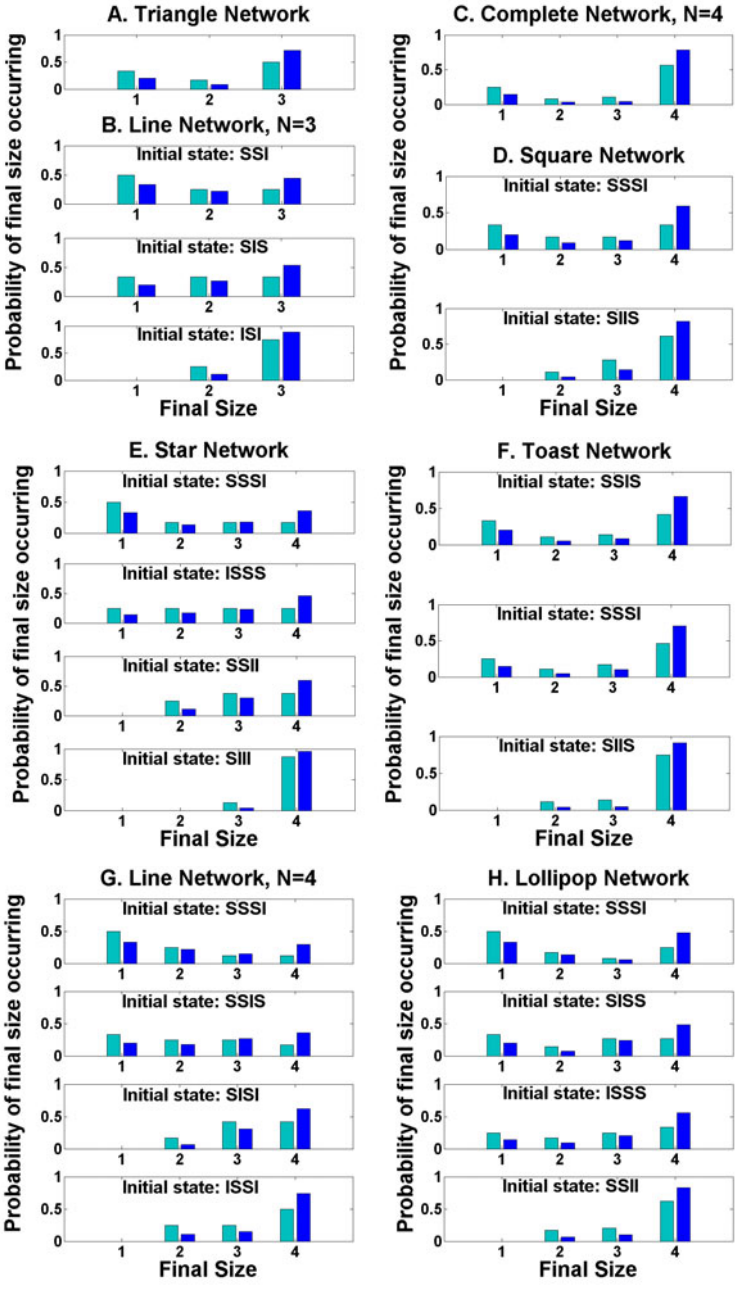

FIGURE 7. Probability mass functions for the eight small networks shown in Figure 1. The PMFs in light and dark blue (colour available online) are for $\mathcal{R} = 1$ and $\mathcal{R} = 2$, respectively. The final size probabilities were validated against results found from stochastic simulations.

https://doi.org/10.1017/S1446181116000043 Published online by Cambridge University Press

TABLE 2. Expressions for the expected final size of an SIR epidemic starting with one infectious node.

	Initial state and initial infectious node, I_0	Degree of I_0	Expected final size
	Networks with $N = 3$		
A. Triangle	SSI ($I_0 = a$)	2	$\frac{6\mathcal{R}^3 + 13\mathcal{R}^2 + 6\mathcal{R} + 1}{(\mathcal{R}+1)^2(2\mathcal{R}+1)}$
B. Line	SSI ($I_0 = a$ or c)	1	$\frac{3\mathcal{R}^2 + 3\mathcal{R} + 1}{(\mathcal{R}+1)^2}$
	SIS ($I_0 = b$)	2	$\frac{3\mathcal{R}+1}{\mathcal{R}+1}$
	Networks with $N = 4$		
C. Complete	SSSI ($I_0 = a, b, c$ or d)	3	$\frac{48\mathcal{R}^6 + 196\mathcal{R}^5 + 310\mathcal{R}^4 + 217\mathcal{R}^3 + 73\mathcal{R}^2 + 13\mathcal{R} + 1}{(\mathcal{R}+1)^3(2\mathcal{R}+1)^2(3\mathcal{R}+1)}$
D. Square	SSSI ($I_0 = a, b, c$ or d)	2	$4 - \frac{3\mathcal{R}^3 + 17\mathcal{R}^2 + 13\mathcal{R} + 3}{(\mathcal{R}+1)^3(2\mathcal{R}+1)}$
E. Star	SSSI ($I_0 = b, c$ or d)	1	$4 - \frac{5\mathcal{R}+3}{(\mathcal{R}+1)^2}$
	ISSS ($I_0 = a$)	3	$4 - \frac{3}{\mathcal{R}+1}$
F. Toast	SSIS ($I_0 = b$ or c)	2	$4 - \frac{6\mathcal{R}^4 + 29\mathcal{R}^3 + 4\mathcal{R}^2 + 19\mathcal{R} + 3}{(2\mathcal{R}+1)^2(\mathcal{R}+1)^3}$
	SSSI ($I_0 = a$ or d)	3	$4 - \frac{6\mathcal{R}^4 + 33\mathcal{R}^3 + 47\mathcal{R}^2 + 21\mathcal{R} + 3}{(2\mathcal{R}+1)(3\mathcal{R}+1)(\mathcal{R}+1)^3}$
G. Line	SSSI ($I_0 = a$ or d)	1	$4 - \frac{6\mathcal{R}^2 + 8\mathcal{R} + 3}{(\mathcal{R}+1)^3}$
	SSIS ($I_0 = b$ or c)	2	$\frac{(2\mathcal{R}+1)^2}{(\mathcal{R}+1)^2}$
H. Lollipop	SSSI ($I_0 = d$)	1	$4 - \frac{8\mathcal{R}^3 + 21\mathcal{R}^2 + 14\mathcal{R} + 3}{(\mathcal{R}+1)^3(2\mathcal{R}+1)}$
	SISS ($I_0 = b$ or c)	2	$4 - \frac{5\mathcal{R}^3 + 16\mathcal{R}^2 + 13\mathcal{R} + 3}{(\mathcal{R}+1)^3(2\mathcal{R}+1)}$
	ISSS ($I_0 = a$)	3	$4 - \frac{4\mathcal{R}^2 + 9\mathcal{R} + 3}{(\mathcal{R}+1)^2(2\mathcal{R}+1)}$

describe the spread of an infection through large networks which are composed of the small networks discussed in this paper. We calculated the clustering coefficient, ϕ, for each of the small networks in order to see if there was any correlation between final size of the epidemic and how clustered the network is. The expected final size for the triangle network is higher than that for the line network of $N = 3$ nodes for all parameter values and, although not shown here, this can be proved analytically. Small networks with a higher clustering coefficient had a higher expected final size when the epidemic was started with one infectious node. The probability that each node in a network would acquire infection at some point during the epidemic increased when the degree of the initial infectious node or the transmission parameter (β, and hence $\mathcal{R}$) was increased for these small networks. Similarly, the probability that the infection would die out before infecting an initially susceptible node was higher when

https://doi.org/10.1017/S1446181116000043 Published online by Cambridge University Press

TABLE 3. Expected final size (EFS) for an SIR epidemic and network clustering coefficients.

	Clustering coefficient, ϕ	Initial state	Degree of initial infectious node(s)	EFS $\mathcal{R}=1$	EFS $\mathcal{R}=2$
Networks with $N=3$					
A. Triangle	1	SSI	2	2.17	2.51
B. Line	0	SSI	1	1.75	2.11
		SIS	2	2	2.33
		ISI	1 & 1	2.75	2.89
Networks with $N=4$					
C. Complete	1	SSSI	3	2.98	3.46
D. Square	0	SSSI	2	2.5	3.1
		SIIS	2 & 2	3.5	3.78
E. Star	0	SSSI	1	2	2.56
		ISSS	3	2.5	3
		SSII	1 & 1	3.125	3.48
		SIII	1, 1 & 1	3.875	3.96
F. Toast	0.75	SSIS	2	2.64	3.21
		SSSI	3	2.85	3.37
		SIIS	2 & 2	3.64	3.87
G. Line	0	SSSI	1	1.88	2.41
		SSIS	2	2.25	2.78
		SISI	2 & 1	3.25	3.56
		ISSI	1 & 1	3.25	3.63
H. Lollipop	0.6	SSSI	1	2.08	2.67
		SISS	2	2.46	3.02
		ISSS	3	2.67	3.18
		SSII	2 & 1	3.46	3.76

the degree of the initial infectious node was 1 when compared with initial infectious nodes of degrees 2 and 3.

5. Discussion

The results presented in this paper give us a good indication that both the network topology and the degree of the initial infectious node are key factors in understanding how an infection might spread through small networks, which could represent small populations or communities. Some of the results presented here are intuitive; however, we emphasize the importance of having found exact analytical expressions for the probability mass functions of the final epidemic size of the small networks. Having the probability mass function for the final epidemic size provides us with more detail than a single expression for the expected final size of an epidemic.

Ongoing investigations involve generalizing the probability mass function of the final epidemic size of larger networks composed of the small networks shown in Figure 1. With the correct assemblage of smaller networks, we aim to describe how an infection can spread through a larger network, based on the results found here for the small networks. Based on this idea, it will not be necessary to construct the transition diagrams for each network. The correct assemblage of smaller networks to generalize

https://doi.org/10.1017/S1446181116000043 Published online by Cambridge University Press

the probability mass function of the final epidemic size is crucial. For example, if we were to join two triangle networks together to assemble a larger network there are three ways we could do this; the first way is by joining the two triangle networks at an apex node; this would create a bow tie network of five nodes. We could also join two triangle networks together by allowing them to share two nodes; this would create a toast network (see Figure 1(F)). We could also have two triangle networks connected by one edge, which would create a line of triangles network of six nodes. Connecting small networks together with overlapping edges becomes problematic, as discussed by Kiss et al. [10], when trying to generalize results based on those already found for the small networks. However, connecting small networks together with overlapping nodes is a feasible approach for the generalization of some time-independent results such as the final epidemic size, which depends only upon the infection parameters and initial conditions.

One limitation of this study is the use of static networks, when in many situations it would be more realistic to consider a dynamic network in which each individual's connections can change over time. Therefore, future work could include investigating the spread of an infection through small dynamic networks of three and four nodes with the intention of composing a larger dynamic network made up of the smaller networks. Another way this research could be extended would be to consider investigating how an infection spreads through small directed networks with the same topological structure as in Figure 1. There are many possibilities for extending the work presented in this paper and we look forward to seeing how this area of research unfolds.

The emphasis of this research is on finding tractable analytical expressions which describe how an infection may spread through a contact network. If we can understand which network properties impact the transmission of infection through a population, then we can use this information to aid the planning and implementation of control strategies such as vaccination campaigns.

Acknowledgements

This research was supported by the Marsden Fund (MAU1106), which we acknowledge and thank for their support. We also wish to thank Dr Roslyn Hickson for helping with the Gillespie algorithm code and for many insightful discussions.

References

[1] F. Ball and P. Neal, "On methods for studying stochastic disease dynamics", *J. R. Soc. Interface* **5** (2008) 171–181; doi:10.1016/j.mbs.2008.01.001.

[2] S. Bansal, B. Grenfell and L. Meyers, "When individual behaviour matters: homogeneous and network models in epidemiology", *J. R. Soc. Interface* **4** (2007) 879–891; doi:10.1098/rsif.2007.1100.

[3] A. Barabási and R. Albert, "Emergence of scaling in random networks", *Science* **286** (1999) 509–512; doi:10.1126/science.286.5439.509.

[4] L. Danon, A. Ford, T. House, C. Jewell, M. Kelling, G. Roberts, J. Ross and M. Vernon, "Networks and the epidemiology of infectious disease", *Interdiscip. Perspect. Infect. Dis.* **2011** (2011) 1–28; doi:10.1155/2011/284909.

[5] J. Gleeson, "High-accuracy approximation of binary-state dynamics on networks", *Phys. Rev. Lett.* **107** (2011) 068701; doi:10.1103/physrevlett.107.068701.
[6] T. House and M. Keeling, "Epidemic prediction and control in clustered populations", *J. Theoret. Biol.* **272** (2010) 1–7; doi:10.1016/j.jtbi.2010.12.009.
[7] M. Keeling, "The effects of local spatial structures on epidemiological invasions", *Proc. R. Soc. Lond. B* **266** (1999) 859–867; doi:10.1098/rspb.1999.0716.
[8] M. Keeling and K. Eames, "Networks and epidemic models", *J. R. Soc. Interface* **2** (2005) 295–307; doi:10.1098/rsif.2005.0051.
[9] M. Keeling and P. Rohani, *Modeling infectious diseases in humans and animals* (Princeton University Press, Princeton, NJ, 2008).
[10] I. Kiss, C. Morris, F. Selley, P. Simon and R. Wilkinson, "Exact deterministic representation of Markovian SIR epidemics on networks with and without loops", *J. Math. Biol.* **70** (2015) 437–464; doi:10.1007/s00285-014-0772-0.
[11] J. Lindquist, J. Ma, P. van den Driessche and F. Willeboordse, "Effective degree network disease models", *J. Math. Biol.* **62** (2011) 143–164; doi:10.1007/s00285-010-0331-2.
[12] A. Lloyd and S. Valeika, "Network models in epidemiology: an overview", in: *World Scientific Lecture Notes in Complex Systems* (eds B. Blasius, J. Kurths and L. Stone), (World Scientific, Singapore, 2007) 189–214; doi:10.1142/9789812771582_0008.
[13] J. Ma, P. van den Driessche and F. Willeboordse, "The importance of contact network topology for the success of vaccination strategies", *J. Theoret. Biol.* **325** (2013) 12–21; doi:10.1016/j.jtbi.2013.01.006.
[14] V. Marceau, P. Noel, L. Hebert-Dufresne, A. Allard and L. Dube, "Adaptive networks: coevolution of disease and topology", *Phys. Rev. E* **82** (2010) 1–9; doi:10.1103/physreve.82.036116.
[15] J. Miller, "A note on a paper by Erik Volze: SIR dynamics in random networks", *J. Math. Biol.* **62** (2010) 349–358; doi:10.1007/s00285-010-0337-9.
[16] J. Miller and I. Kiss, "Epidemic spread in networks: existing methods and current challenges", *Math. Model. Nat. Phenom.* **9** (2014) 4–42; doi:10.1051/mmnp/20149202.
[17] J. Miller, A. Slim and E. Volz, "Edge-based compartmental modelling for infectious disease spread", *J. R. Soc. Interface* **9** (2012) 890–906; doi:10.1098/rsif.2011.0403.
[18] J. Miller and E. Volz, "Incorporating disease and population structure into models of SIR disease in contact networks", *PLoS ONE* **8** (2013) e69162; doi:10.1371/journal.pone.0069162.
[19] M. Newman, *Networks: an introduction* (Oxford University Press, New York, 2010).
[20] R. Pastor-Satorras and A. Vespignani, "Epidemic dynamics and endemic states in complex networks", *Phys. Rev. E* **63** (2001) 066117; doi:10.1103/physreve.63.066117.
[21] L. Pellis, F. Ball, S. Bansal, K. Eames, T. House, V. Isham and P. Trapman, "Eight challenges for network epidemic models", *Epidemics* **10** (2015) 58–62; doi:10.1016/j.epidem.2014.07.003.
[22] M. Ritchie, L. Berthouze, T. House and I. Kiss, "Higher-order structure and epidemic dynamics in clustered networks", *J. Theoret. Biol.* **348** (2014) 21–32; doi:10.1016/j.jtbi.2014.01.025.
[23] P. Simon, M. Taylor and I. Kiss, "Exact epidemic models on graphs using graph-automorphism driven lumping", *J. Math. Biol.* **62** (2010) 479–508; doi:10.1007/s00285-010-0344-x.
[24] M. Taylor, T. Taylor and I. Kiss, "Epidemic threshold and control in a dynamic network", *Phys. Rev. E* **85** (2012) 016103; doi:10.1103/physreve.85.016103.
[25] E. Volz, "SIR dynamics in random networks with heterogenous connectivity", *J. Math. Biol.* **56** (2008) 293–310; doi:10.1007/s00285-007-0116-4.
[26] D. Watts and S. Strogatz, "Collective dynamics of 'small-world' networks", *Nature* **393** (1998) 440–442; doi:10.1038/30918.
[27] Y. Zhao and M. Roberts, "Simulating epidemics on networks", *Res. Lett. Inform. Math. Sci.* **6** (2007) 101–103; http://www.massey.ac.nz/massey/learning/colleges/college-of-sciences/research/natural-mathematical-sciences/mathematics-research/en/simulating-epidemics-on-networks.cfm.