Mathematical Statistics
Stockholm University

A network with tunable clustering, degree correlation and degree distribution, and an epidemic thereon

Frank Ball
Tom Britton
David Sirl

Research Report 2012:10

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Research Report **2012:10**,
http://www.math.su.se/matstat

A network with tunable clustering, degree correlation and degree distribution, and an epidemic thereon

Frank Ball, Tom Britton and David Sirl

July 2012

Abstract

A random network model which allows for tunable, quite general forms of clustering, degree correlation and degree distribution is defined. The model is an extension of the configuration model, in which stubs (half-edges) are paired to form a network. Clustering is obtained by forming small completely connected subgroups, and positive (negative) degree correlation is obtained by connecting a fraction of the stubs with stubs of similar (dissimilar) degree. An SIR (Susceptible → Infective → Recovered) epidemic model is defined on this network. Asymptotic properties of both the network and the epidemic, as the population size tends to infinity, are derived: the degree distribution, degree correlation and clustering coefficient, as well as a reproduction number R_*, the probability of a major outbreak and the relative size of such an outbreak. The theory is illustrated by Monte Carlo simulations and numerical examples. The main findings are that clustering tends to decrease the spread of disease, the effect of degree correlation is appreciably greater when the disease is close to threshold than when it is well above threshold and disease spread broadly increases with degree correlation ρ when R_* is just above its threshold value of one and decreases with ρ when R_* is well above one.

A network with tunable clustering, degree correlation and degree distribution, and an epidemic thereon

Frank Ball[*], Tom Britton[†] and David Sirl[‡]

July 13, 2012

Abstract

A random network model which allows for tunable, quite general forms of clustering, degree correlation and degree distribution is defined. The model is an extension of the configuration model, in which stubs (half-edges) are paired to form a network. Clustering is obtained by forming small completely connected subgroups, and positive (negative) degree correlation is obtained by connecting a fraction of the stubs with stubs of similar (dissimilar) degree. An SIR (Susceptible → Infective → Recovered) epidemic model is defined on this network. Asymptotic properties of both the network and the epidemic, as the population size tends to infinity, are derived: the degree distribution, degree correlation and clustering coefficient, as well as a reproduction number R_*, the probability of a major outbreak and the relative size of such an outbreak. The theory is illustrated by Monte Carlo simulations and numerical examples. The main findings are that clustering tends to decrease the spread of disease, the effect of degree correlation is appreciably greater when the disease is close to threshold than when it is well above threshold and disease spread broadly increases with degree correlation ρ when R_* is just above its threshold value of one and decreases with ρ when R_* is well above one.

Keywords: Branching process, configuration model, epidemic size, random graph, SIR epidemic, threshold behaviour.

MSC codes: 92D30, 05C80, 60J80.

1 Introduction

Ever since the pioneering work of Erdős and Rényi (1959) on a simple random graph there have been numerous important contributions on random graph models with the aim of

[*] School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK. Email: frank.ball@nottingham.ac.uk

[†] Department of Mathematics, Stockholm University, 106 91 Stockholm, Sweden. Email: tomb@math.su.se

[‡] Mathematics Education Centre, Loughborough University, Loughborough, Leicestershire, LE11 3TU, UK. E-mail: d.sirl@lboro.ac.uk

making them more flexible and realistic. For example, the configuration model (Molloy and Reed (1995) and Newman et al. (2001)) defines a network allowing for more or less arbitrary *degree distribution* F_D, the distribution describing the number of neighbours D of a randomly selected node (which in the epidemic context represents an individual) in the network. (For simplicity, from now on we refer to D as the degree distribution.) This extension was important for two reasons: most empirical networks tend to have much heavier tailed degree distributions than the Poisson distribution of the Erdős-Rényi (E-R) network, and networks with heavy tail degree distributions have been shown to exhibit rather different properties when compared with the E-R network; for example, if an epidemic outbreak takes place on the network the epidemic threshold R_0 is much higher (or even infinite) as compared to the same epidemic taking place on an E-R network with the same mean degree (Andersson (1999)).

Two other properties of real world networks that are not present in E-R networks are *clustering* and *degree correlation*. The clustering coefficient c measures how likely it is that two neighbours of a randomly selected node are neighbours themselves. The E-R network has no clustering whereas nearly all empirical networks have positive clustering, with typical values in the range 0.1–0.5 out of the possible range 0–1 (see Newman (2003), Table 3.1). The degree correlation ρ instead measures the correlation between the degrees of the adjacent individuals of a randomly selected *edge*. The E-R network has $\rho = 0$ whereas 'random' networks with heavy tail degree distribution tend to have $\rho > 0$ (van der Hofstad and Litvak (2012)). Empirical networks, on the other hand, have both positive and negative degree correlation: there seems to be a tendency for computer networks to have $\rho < 0$ whereas social networks (our main interest) typically have $\rho > 0$ (see Newman (2003), Table 3.1). There are numerous network models studied in the literature, with the aim of allowing one or several of these three extensions (of *local* properties) from the original E-R-network (see some references below where the focus is also on epidemics evolving on the network); the term 'local' refers to the fact that it is sufficient to observe nodes and their neighbourhoods to determine/estimate such properties (the complete network need not be observed in order to evaluate them). The current paper defines a model in which D, c and ρ can be made more or less arbitrary.

There are of course other important extensions in addition to allowing for arbitrary degree distribution, degree correlation and clustering. Further local properties considered in many models for social networks are households and other fully connected smaller units (e.g. Ball et al. (1997)), and models in which nodes and/or edges are of different types (e.g. Britton et al. (2007), Ball and Sirl (2012)). Several models have also been proposed which combine household and network structure, for example Trapman (2007), Gleeson (2009), Ball et al. (2010) and Ma et al. (2012). Other models aim to study and extend the range of global properties, such as small world networks (Watts and Strogatz (1998)) and dynamic network models (Barabási and Albert, (1999)). This paper does not address these (or any other) extensions; the focus being on degree distribution, degree correlation and clustering.

Our main motivation for studying networks is to investigate social networks and to examine what effect the three above-mentioned properties have in the event of an infectious disease entering the community; both in terms of the possibility and probability of an epidemic outbreak taking off, and also how large such an outbreak will be if it does take off. We study the class of SIR epidemics (e.g. Andersson and Britton (2000)) in which individuals are at first Susceptible (except for some introductory infectious cases) and those who get infected become Infectious for a random period of time when they may infect their network neighbours, after which they Recover and become immune to further infection. See, for instance, Diekmann et al. (1998), Andersson (1999) and Diekmann and Heesterbeek (2000, Ch. 10) for early analytical contributions in this area.

As mentioned above there have been many contributions to this area of research, in particular over the last decade or two. Allowing for arbitrary degree distribution, and studying its effect on an epidemic, dates back longer. May and Anderson (1987) concluded (when modelling the spread of HIV) that a heavy tail degree distribution makes the reproduction number R_0 large or even infinite. The important insight from their analysis was that diseases with very low transmission probability still may be at risk of epidemics taking off in networks having small mean degree, if the *variance* of the degree distribution is very large. The effect of clustering on epidemics has been studied in, for example, Britton et al. (2008), Miller (2009) and Newman (2009). Degree correlation has often been analysed in combination with clustered networks (e.g. Gleeson et al. (2010)). The impact of clustering and degree correlation on epidemics on networks has been studied empirically using simulation by Badham and Stocker (2010) and Isham et al. (2011). The main focus of most papers concerning epidemics on networks considering clustering, degree correlation and/or degree distribution lies in studying how these features affect the basic reproduction number R_0, i.e. the *possibility* of having an major outbreak. To derive the *probability* of such an outbreak, and its likely size in the event that it takes off, requires significantly deeper analysis; which for several of the above-mentioned models still is missing.

The current paper introduces a network model which (i) allows for more or less arbitrary clustering, degree correlation and degree distribution, and (ii) permits theoretical analysis of epidemics defined on the network. As in the configuration model, the network is formed by attaching stubs (i.e. half-edges) to individuals, which are then paired to form the edges of the network. The degree of an individual is the number of stubs emanating from it. The desired clustering and degree distribution is obtained by having two types of stubs going out from individuals. A fraction of stubs is local (which fraction being closely related to the desired clustering); the remaining stubs are global and are connected randomly (as described below) among stubs from all individuals. The local stubs are connected by grouping individuals into small local groups (‘households’). For example, an individual with four local stubs is connected to four other individuals having local degree 4, thus forming a group of 5 completely connected individuals (contributing to increased clustering). The degree distribution is given by the distribution of the sum of the local and global degree of a typical individual. Finally, the desired degree correlation ρ is obtained by manipulating how the global stubs are connected, which is controlled by a parameter r satisfying $-1 \leq r \leq 1$. With probability $1 - |r|$ a stub is connected uniformly at random among all global stubs. With probability $|r|$ the stub is connected to a stub having very similar total degree (if $r > 0$) or ‘opposite’ total degree (if $r < 0$).

The remainder of the paper is organised as follows. A more rigorous definition of the model appears in Section 2, where a continuous-time SIR epidemic on the network is also defined. In Section 3, we derive expressions for the degree distribution D, clustering coefficient c and degree correlation ρ, as functions of the model parameters, and discuss the more relevant reverse problem of choosing model parameters to obtain a desired c, ρ

and D, using a Poisson total degree distribution as a template. We also describe a simple rewiring algorithm, motivated by Miller (2009) and Gleeson et al. (2010), which permits the clustering in a network to be reduced in a controlled fashion without changing ρ or D. In Section 4, we analyse the main characteristics of epidemics defined on the network for suitably large population sizes, by exploiting approximating branching processes. Specifically, in Section 4.1, we obtain a threshold parameter R_* which determines whether or not a major outbreak is possible, and derive the probability that a major outbreak occurs (assuming that the infectious period is constant) and, in Section 4.2, we derive the relative final size (i.e. the proportion of the population that is ultimately removed) of a major outbreak. In Section 5, we describe how these results on epidemics are modified to incorporate rewiring and prove that, if all other parameters are held fixed, such rewiring increases the threshold parameter R_* and both the probability and relative final size of a major outbreak. In Section 6 we illustrate the theory with some numerical examples which demonstrate that the effect of degree correlation on epidemic properties is appreciably greater when the disease is just above threshold than when it is well above threshold. Moreover, both the probability and size of a major outbreak broadly increase with ρ when the disease is just above threshold, while they broadly decrease with ρ when the disease is well above threshold. However, this behaviour is not monotonic, particularly when clustering is low and R_* is close to one. We conclude with a brief discussion in Section 7.

2 The network model and the epidemic

2.1 The network model

Consider a network of undirected edges with n nodes (individuals). Below we define how to construct the network. First we define a set of random variables and briefly explain their interpretation in the network.

Let G be a discrete non-negative random variable with distribution $\{p_k\}$ referred to as the 'global degree', let H be another strictly positive discrete random variable with distribution $\{\pi_h\}$. In some cases H will reflect the household distribution in the community, but in applications where the underlying network has no household structure H is simply a device to introduce clustering into the network. Finally, let r be a real number satisfying $-1 \leq r \leq 1$. The value of $|r|$ reflects how often outgoing global edges connect to nodes of similar (if $r > 0$) or 'opposite' (if $r < 0$) 'total degree'. Let X be a Bernoulli random variable with parameter $|r|$, so $P(X = 1) = |r| = 1 - P(X = 0)$, this variable will determine if a stub will connect to a random stub or a stub with similar/'opposite' degree.

The network is constructed as follows. Let $H_1, H_2, \cdots$ be independent and identically distributed copies of the random variable H. Label the n nodes $1, 2, \cdots, n$ and group the first H_1 nodes into local group (household) one, nodes $H_1 + 1, H_1 + 2, \cdots, H_1 + H_2$ into group 2 and so on until all individuals belong to a local group (the last group will have a 'truncated' size). All nodes of a local group are connected to each other (for example, the first H_1 nodes make up a fully connected component with all individuals having local degree $H_1 - 1$). Let $G_1, G_2, \cdots, G_n$ be independent and identically distributed copies G; G_i denotes the global degree of node i. The total degree of individual i, D_i, equals the global degree plus the local degree, the local degree being one less than the group size. For example, a node residing in a local triangle ($H = 3$) and having global degree k has total degree $k+2$, whereas a local singleton ($H = 1$) with global degree j has total degree j. A node having global degree k has k outgoing stubs, and each of these stubs is labelled with an independent copy of X (stubs having independent and identically distributed X-variables with $P(X = 1) = |r| = 1 - P(X = 0)$) and the *total* degree of the node from which it emanates. All outgoing stubs in the network with label $X = 0$ are connected pairwise completely at random. The remaining stubs (having $X = 1$) are also connected randomly but in a different manner. This is done by ordering all global stubs having label $X = 1$ (suppose that there are n_1 such stubs) according to their total degree, and then separating the empirical distribution of global degrees so generated into n_Q (a fixed and freely chosen positive integer) equally sized quantiles. (If n_1/n_Q is not an integer then the n_Q quantiles are made as equal in size as possible.) The first such quantile hence consists of the n_1/n_Q stubs having smallest label (i.e. total degree) and so on. If $r > 0$, each quantile is treated in turn and all the stubs in that quantile are paired uniformly at random. If $r < 0$, the stubs in the first quantile are paired uniformly at random with those in the n_Qth quantile, the stubs in the second quantile are paired uniformly at random with those in the $n_Q - 1$th quantile, and so on. Thus, if n_Q is odd, the stubs in the middle quantile are paired uniformly at random with each other. The effect of this pairwise connection is that nodes of similar total degree will be connected if $r > 0$, whereas nodes of rather different total degree will be connected if $r < 0$; in both cases leading to correlated degrees (but of different sign). There may be one unattached stub having label $X = 0$ and at most n_Q unattached stubs having label $X = 1$ following the above pairings. These are simply ignored. This has no effect on the asymptotic properties of the network, nor on epidemics defined thereon, as $n \to \infty$. In the above construction, all the H, G and X random variables are assumed to be independent.

The network is hence made up of local completely connected groups having groups size distribution $\{\pi_h\}$ (as n goes to infinity the effect of the last group having a truncated household size is negligible). On top of this, each individual has global edges, the number being distributed as G. Some of these will be formed by connecting to other random stubs, the others will be formed by connecting to other stubs having similar or 'opposite' degree, thus creating positive or negative degree correlation. The construction of global edges may result in the presence of multiple edges and self-loops. However, if the degree distribution D has finite variance, the fraction of these will be negligible as $n \to \infty$, so removing them has negligible effect on the degree distribution and how stubs are connected (cf. Durrett (2006, Theorem 3.1.2) and Janson (2009)). The special case where $r = 0$ or $n_Q = 1$ is the network and households model (without degree correlation beyond that induced by the presence of households) studied by Ball et al. (2010), since in either of these situations all global stubs are simply paired uniformly at random.

2.2 An epidemic model on the network

We now define a continuous-time epidemic model for the spread of an SIR-type infectious disease upon the network defined in Section 2.1. We suppose that there is one initial infective, chosen uniformly at random from the n individuals (nodes) in the population and that the remainder of the population is susceptible. The infectious periods of different infectives are each distributed according to a random variable I, having an arbitrary but specified distribution. Throughout its infectious period, a given infective makes infectious contacts with any given neighbour (either local or global) in the network at the points of a homogeneous Poisson process having rate λ. A susceptible becomes infective as soon as it is contacted by an infective and an infective becomes removed (and plays no further part in the epidemic) at the end of its infectious period. Contacts between an infective and an infective or removed individual have no effect. All Poisson processes describing infectious contacts (whether or not either or both individuals involved are the same) and all infectious periods are mutually independent; they are also independent of the random variables used to construct the network. The epidemic ends when there is no infective remaining in the population.

3 Properties of the network model

We now derive the total degree distribution D, the clustering coefficient c and the degree correlation ρ for the network defined in Section 2.1. We treat the asymptotic case where the number of nodes n tends to infinity.

3.1 The degree distribution

We start with the degree distribution. From the construction it follows immediately that a node has global degree G. The local degree is one less than the household size, and the household size of a randomly selected node has distribution $\{\tilde{\pi}_h\}$, where $\tilde{\pi}_h = h\pi_h/\mu_H$ and $\mu_H = \sum_j j\pi_j$, i.e. the size-biased local group-size distribution. Let $\tilde{H}$ denote a random variable having the size-biased household distribution. It then follows that the total degree distribution (in the network) is given by

$$D \overset{D}{=} G + \tilde{H} - 1, \tag{1}$$

where $\overset{D}{=}$ means equal in distribution and G and $\tilde{H}$ are independent. In particular it follows that the mean total degree is

$$\mu_D = \mu_G + \frac{\sigma_H^2}{\mu_H} + \mu_H - 1.$$

(Throughout the paper, for a random variable, X say, μ_X and σ_X^2 denote respectively the mean and variance of X.)

3.2 The clustering coefficient

There are several measures of clustering used in the literature. We use a 'probabilistic' one (see, for example, Trapman (2007)) where an ordered triplet of nodes (i, j, k) is selected completely at random among all such ordered triplets for which i is directly connected to j and j is directly connected to k. The clustering coefficient c is then defined as the probability that i and k are also directly connected (i.e. that i, j and k form a triangle).

Thus c is given by the fraction of ordered triplets in the network that are triangles. The clustering coefficient of the present network model is identical to that of the model in Ball et al. (2010), since the models differ only in the way that global stubs are paired. For large n, the proportion of ordered triangles that are not wholly within households is small and zero in the limit as $n \to \infty$. Thus, asymptotically, the global pairings do not yield triangles in either of the two models, explaining why the clustering coefficients are the same for the two models. Hence, from equation (14) of Ball et al. (2010), the clustering coefficient $c = c(G, H, r)$ is given by

$$c = \frac{\mathrm{E}[H(H-1)(H-2)]}{\mathrm{E}[(H(G+H-1)(G+H-2)]}, \tag{2}$$

where G and H are the household and global degree distributions of the network.

3.3 The degree correlation

We now formulate an expression for the degree correlation ρ of the current network model. One way to define ρ is to pick a random edge in the network and let ρ be the correlation between the total degrees of the nodes adjacent to this edge (Newman, 2002a). The derivation of ρ involves long but standard computations which are given in the appendix. A key step in the derivation is to first condition on whether the chosen edge is a global or a local edge, the former having probability p_G given by

$$p_G = \frac{\mu_G}{\mu_G + \mu_{\tilde{H}} - 1}. \tag{3}$$

If the edge is global the degree covariance (of the right and left node adjacent to the edge) comes from the two stubs having the same (or 'opposite') quantile(s), which happens with probability $|r|$, and if the edge is local the degree covariance stems from the nodes having the same local degree.

Before giving the expression for the degree correlation $\rho = \rho(G, H, r)$ some more notation is required. Let $\hat{H}$ denote a random variable giving the household size of a household edge chosen uniformly at random from all household edges. Since a household of size h contains $\binom{h}{2}$ edges, $\mathrm{P}(\hat{H} = h) \propto \binom{h}{2}\pi_h$ $(h = 2, 3, \cdots)$, so

$$\mathrm{P}(\hat{H} = h) = \frac{h(h-1)\pi_h}{\mathrm{E}[H(H-1)]} \quad (h = 2, 3, \cdots).$$

Let $\tilde{D}$ and $\tilde{Q}$ denote respectively the *total* degree and quantile of a stub chosen uniformly at random from all stubs in the limit as $n \to \infty$. Then $\tilde{D} \overset{D}{=} \tilde{G} + \tilde{H} - 1$, where $\tilde{G}$ and $\tilde{H}$ are independent, and $\tilde{G}$ denotes a random variable having the size-biased global degree distribution $\{\tilde{p}_g\}$, where $\tilde{p}_g = gp_g/\mu_G$ $(g = 1, 2, \cdots)$. For $i = 1, 2, \cdots, n_Q$ and $d = 1, 2, \cdots$, let $p_{\tilde{Q}|\tilde{D}}(i|d) = \mathrm{P}(\tilde{Q} = i|\tilde{D} = d)$ and $p_{\tilde{D}|\tilde{Q}}(d|i) = \mathrm{P}(\tilde{D} = d|\tilde{Q} = i)$. (These conditional probabilities are derived easily from the probability mass function of $\tilde{D}$, noting that if $\tilde{u}_0 = 0$ and $\tilde{u}_d = \mathrm{P}(\tilde{D} \leq d)$ $(d = 1, 2, \cdots)$ then $\mathrm{P}(\tilde{D} = d, \tilde{Q} = i) = \max\left\{\min(\tilde{u}_d, \frac{i}{n_Q}) - \max(\tilde{u}_{d-1}, \frac{i-1}{n_Q}), 0\right\}$ $(d = 1, 2, \cdots; i = 1, 2, \cdots, n_Q)$.) Define the function $g_{\tilde{D},n_Q}(r)$ by

$$g_{\tilde{D},n_Q}(r) = \begin{cases} r\left(\frac{1}{n_Q}\sum_{i=1}^{n_Q}(\mu_{\tilde{D}}^{(i)})^2 - \mu_{\tilde{D}}^2\right) & \text{if } r \geq 0, \\ |r|\left(\frac{1}{n_Q}\sum_{i=1}^{n_Q}\mu_{\tilde{D}}^{(i)}\mu_{\tilde{D}}^{(n_Q+1-i)} - \mu_{\tilde{D}}^2\right) & \text{if } r < 0, \end{cases} \tag{4}$$

where

$$\mu_{\tilde{D}}^{(i)} = \sum_{d=1}^{\infty} d p_{\tilde{D}|\tilde{Q}}(d|i) \quad (i = 1, 2, \cdots, n_Q). \tag{5}$$

It is shown in the appendix that

$$\rho = \frac{(1-p_G)\sigma_{\hat{H}}^2 + p_G g_{\tilde{D},n_Q}(r) + p_G(1-p_G)\left(\mu_{\hat{H}} - \mu_{\tilde{H}} - \frac{\sigma_G^2}{\mu_G}\right)^2}{(1-p_G)\left(\sigma_{\hat{H}}^2 + \sigma_G^2\right) + p_G\left(\sigma_{\tilde{H}}^2 + \sigma_{\tilde{G}}^2\right) + p_G(1-p_G)\left(\mu_{\hat{H}} - \mu_{\tilde{H}} - \frac{\sigma_G^2}{\mu_G}\right)^2}. \tag{6}$$

3.4 Rewiring

Note that for household size and global degree distributions H and G, the degree distribution D and the clustering coefficient c are both independent of the parameter r. Thus, by letting r vary between -1 and $+1$ and keeping the distributions of H and G fixed, it is straightforward to tune the degree correlation in our network model without changing the degree distribution or clustering coefficient of the network. However, if we keep r fixed and vary, for example, the household size distribution to change the clustering coefficient of the network, then its degree distribution D and degree correlation ρ change also. This observation means that it is more difficult to tune just the clustering coefficient in a network. One way around this problem is to extend the rewiring construction of Gleeson et al. (2010) (see also Miller (2009), where the idea first originated) to our model.

Suppose that we construct a realisation of our network model and then colour all global edges green and all household edges red. Household edges are also labelled according to their household size. Let p_{RW} be a real number satisfying $0 \leq p_{RW} \leq 1$. Then, independently for each household, with probability p_{RW} the red edges in a household are each broken into two stubs, which retain their colour and household-size labels. For each $h = 2, 3, \cdots$, the red stubs with label h are now joined uniformly at random, which, together with the green edges and unbroken red edges creates a new network.

Observe that the above rewiring does not alter the degree distribution or the correlation structure (and in particular the degree correlation) of the network but it does change its clustering coefficient. Let $c(G, H, r, p_{RW})$ denote the clustering coefficient for the model with rewiring probability p_{RW}, so $c(G, H, r, 0)$ is the clustering coefficient of our model without rewiring. In the limit as $n \to \infty$, the proportion of triangles that are not wholly within unbroken households tends to zero, whence $c(G, H, r, p_{RW}) = (1 - p_{RW})c(G, H, r, 0)$. Thus, given our network model without rewiring, it is straightforward to use the above rewiring to tune the clustering coefficient to be any value between 0 and that of the model without rewiring.

3.5 Tuning

The formulae given in Sections 3.2 and 3.3 are fairly long but simplify appreciably for the special situation where both the household sizes and the global degrees follow Poisson-based distributions. Specifically, suppose that, with $0 \leq \mu < \gamma$, G follows a Poisson distribution with mean $\gamma - \mu$, which we denote by $\text{Poi}(\gamma - \mu)$, and H follows a Poisson distribution with mean μ that is conditioned on being strictly positive, which we denote by $\text{Poi}^+(\mu)$. Here we interpret $\text{Poi}^+(0)$ to be $\lim_{\mu \downarrow 0} \text{Poi}^+(\mu)$, the distribution identically equal to 1. Thus $\pi_h = (1 - \mathrm{e}^{-\mu})^{-1} \mu^h \mathrm{e}^{-\mu}/h!$ $(h = 1, 2, \cdots)$. Then $\tilde{H} - 1 \sim \text{Poi}(\mu)$ and it follows from (1) that the total degree $D \sim \text{Poi}(\gamma)$. Further, $1 - p_G = \mu/\gamma$ and $\hat{H} - 2 \sim \text{Poi}(\mu)$, so using (2) and (6), the formulae for the clustering and degree correlation are given by:

$$c = \left(\frac{\mu}{\gamma}\right)^2 \qquad \text{and} \qquad \rho = \frac{1}{\gamma^2}\left[\mu^2 + (\gamma - \mu) g_{\gamma, n_Q}(r)\right], \tag{7}$$

where $g_{\gamma, n_Q}(r)$ is given by (4) with $\tilde{D} \sim 1 + \text{Poi}(\gamma)$.

Observe that $g_{\gamma, n_Q}(0) = 0$, so $c = \rho$ when $r = 0$, i.e. for the model studied in Ball et al. (2010), Sections 4.3 and 4.4. Suppose that γ and μ are held fixed, so the clustering coefficient c is also held fixed. Then as r varies from -1 to $+1$ the degree correlation ρ varies between the values obtained by setting $r = -1$ and $r = 1$ in the formula for ρ in (7). These lower and upper values for ρ are shown in Figure 1 as functions of c for different choices of the number of quantiles n_Q, for the case when $\gamma = 10$. In the limit as $n_Q \to \infty$, if $r > 0$ then a stub with label $X = 1$ is paired, almost surely, with a stub having the same total degree and $g_{\gamma, n_Q}(1) \to \text{var}(\tilde{D}) = \gamma$ (recall $\tilde{D} \sim 1 + \text{Poi}(\gamma)$). It follows that the corresponding upper value for ρ is $1 + c - \sqrt{c}$. In the same limiting situation, if $r < 0$ then a stub with label $X = 1$ is paired, almost surely, with a stub having the 'opposite' total degree. There is no simple expression for $\lim_{n_Q \to \infty} g_{\gamma, n_Q}(-1)$, though it is easily computed. Observe from Figure 1 that very little extra is gained, in terms of the range of possible (c, ρ), by choosing a large value of n_Q. In practice, a small value of n_Q is beneficial as the proportions of self-loops and parallel edges between nodes, resulting from the pairing of stubs, both increase with n_Q. Additionally, large values of n_Q mean that the approximating branching processes have many types and numerical calculation of quantities of interest becomes more computationally intensive.

Write $c = c(\gamma, \mu, r)$ and $\rho = \rho(\gamma, \mu, r)$ to show explicitly their dependence on the parameters and, for $\gamma > 0$, let $A_\gamma = \{(c(\gamma, \mu, r), \rho(\gamma, \mu, r)) : 0 \leq \mu \leq \gamma, -1 \leq r \leq 1\}$ be the set of possible values (c, ρ) in our model when the total degree is $\text{Poi}(\gamma)$. For any $(c, \rho) \in A_\gamma$ there is a unique (μ, r) such that $(c(\gamma, \mu, r), \rho(\gamma, \mu, r)) = (c, \rho)$, so the model without rewiring can be tuned uniquely to any attainable (c, ρ). If we allow rewiring, it is easily seen that by choosing the rewiring probability p_{RW} appropriately, for each (c, ρ) lying strictly above the lower boundary of A_γ, there is a continuum of models with clustering coefficient c and degree correlation ρ.

A similar analysis to the above holds for other choices of total degree distribution D, though note that not all distributions D can be decomposed as in (1) in such a way that the clustering may be tuned continuously. Distributions D for which this is possible include negative binomial and compound Poisson. Indeed any distribution D that is infinitely divisible may be decomposed so that the clustering coefficient is any rational number in $[0, 1)$.

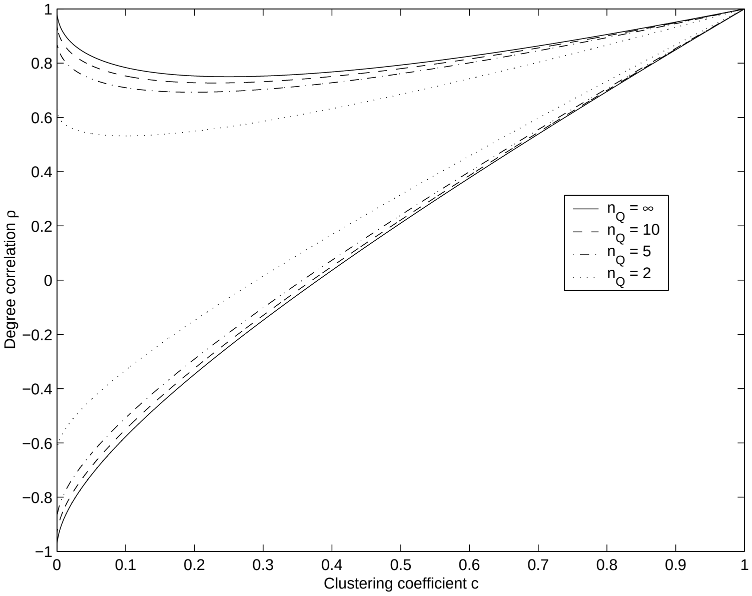

Figure 1: Possible values of (c, ρ) when $D \sim \text{Poi}(10)$.

4 Epidemics on network without rewiring

4.1 Establishment of the epidemic

4.1.1 Approximating forward branching process

The initial infective triggers a local (i.e within-household) epidemic in its household. Each infective in that local epidemic (including the initial infective) may make (global) infectious contact with individuals in other households. If the population size n is large, the probability that such global infectious contacts are all with individuals in previously uninfected households is close to one, owing to the random way in which the underlying network is formed. It follows that in the early stages of an epidemic the process of infected households may be approximated by a branching process, with individuals in the branching process corresponding to infectious households in the epidemic process. Unless $r = 0$ or $n_Q = 1$, this branching process needs to be multitype, since the degrees of endpoints of a global edge with $X = 1$ are correlated. Except for the ancestor, the type of an individual in the branching process is obtained by considering the primary infective, i^* say, in the corresponding single-household epidemic. The type of the individual is given by the total-degree quantile of the stub used in constructing the global edge along which i^* was infected in the epidemic. Thus there are n_Q types of individual in the branching process. The ancestor of the branching process is not typed in this fashion since the initial infective in the epidemic is chosen uniformly at random from the population and not infected along a global edge in the network. Nevertheless, the offspring distribution of the ancestor in the branching process depends on the household size and global degree of the initial infective in the epidemic.

Following Ball et al. (2009), the above branching process is termed a forward branching process as it approximates the forward spread of an epidemic process. In Section 4.2 we consider a backward branching process, which approximates an inverse epidemic process.

The approximation of the early stages of the epidemic process by the forward branching process can be made precise by constructing the branching process and, for each $n = 1, 2, \cdots$, a realisation of the epidemic process on a common probability space and using a coupling argument to show that, as $n \to \infty$, the process of infected households in the epidemic process converges almost surely to the multitype branching process; cf. Ball and Sirl (2012). Thus, if the population size n is sufficiently large, the probability that the epidemic becomes established and leads to a major outbreak is given approximately by the probability that the branching process survives (i.e. does not go extinct). Moreover, whether or not a major outbreak can occur with non-zero probability is determined by whether or not the branching process is supercritical.

We now determine the means and probability generating functions (PGFs) of the offspring distributions of the branching process, which determine respectively whether a major outbreak can occur and, if so, its probability. The offspring distribution is different in the initial generation from that of all subsequent generations, since the initial infective is chosen uniformly at random from the population (so its local and global degrees are independent), while subsequent primary infectives are infected through the network and their local and global degrees are dependent. We focus first on the offspring means for a non-initial generation, since they determine whether or not the branching process is supercritical.

4.1.2 Offspring mean matrix and threshold parameter R_*

Let $\mathcal{B}_F$ denote the above multitype forward branching process and let $\tilde{\mathcal{B}}_F$ be the multitype branching process describing the descendants of a typical first-generation individual in $\mathcal{B}_F$. Thus the type-dependent offspring law is the same for *all* generations in $\tilde{\mathcal{B}}_F$. For $i = 1, 2, \cdots, n_Q$, let $\tilde{\boldsymbol{C}}_i = (\tilde{C}_{i1}, \tilde{C}_{i2}, \cdots, \tilde{C}_{in_Q})$ be a vector random variable describing the numbers of offspring of different types of a typical type-i individual in the branching process $\tilde{\mathcal{B}}_F$. Thus, $\tilde{C}_{ij}$ is the number of type-j primary infectives generated by a typical single-household epidemic, whose primary infective is of type i. Let $\tilde{M} = [\tilde{m}_{ij}]$ be the $n_Q \times n_Q$ matrix with elements $\tilde{m}_{ij} = \mathrm{E}[\tilde{C}_{ij}]$ and let R_* be the dominant eigenvalue of $\tilde{M}$. Then by standard multitype branching process theory (see Mode (1971), Chapter 1, Theorem 7.1), the branching process $\tilde{\mathcal{B}}_F$ survives with strictly positive probability if and only if $R_* > 1$. Thus R_* serves as a threshold parameter for our epidemic model. Note that this and subsequent results using the theory of multitype branching processes require assumptions regarding the irreducibility and/or positive regularity of the mean matrix $\tilde{M}$, which are met for all but highly pathological choices of G, H and n_Q.

In order to compute $\tilde{M}$, and hence R_*, we need a further probability distribution. For $d = 1, 2, \cdots$ and $h = 1, 2, \cdots, d$, let $\tilde{\pi}_h^{(d)}$ be the probability that a stub chosen uniformly at random from all stubs having total degree d belongs to an individual who resides in a household of size h. Note that this probability is the same for stubs with label $X = 0$ and stubs with label $X = 1$, and that

$$\tilde{\pi}_h^{(d)} = \frac{\pi_h h \tilde{p}_{d-h+1}}{\sum_{h'=1}^{d+1} \pi_{h'} h' \tilde{p}_{d-h'+1}} = \frac{\tilde{\pi}_h \tilde{p}_{d-h+1}}{\sum_{h'=1}^{d+1} \tilde{\pi}_{h'} \tilde{p}_{d-h'+1}}.$$

To obtain $\tilde{m}_{ij}$, we condition first on the total degree of a typical type-i primary infective and then on the size of its household yielding

$$\tilde{m}_{ij} = \sum_{d=1}^{\infty} p_{\tilde{D}|\tilde{Q}}(d|i) \sum_{h=1}^{d} \tilde{\pi}_h^{(d)} \mathrm{E}[\tilde{C}_{ij}^{(h,d)}], \tag{8}$$

where $\tilde{\boldsymbol{C}}_i^{(h,d)} = (\tilde{C}_{i1}^{(h,d)}, \tilde{C}_{i2}^{(h,d)}, \cdots, \tilde{C}_{in_Q}^{(h,d)})$ is defined analogously to $\tilde{\boldsymbol{C}}_i$, except we condition on the type-i individual residing in a household of size h and having total degree d. (Note also that $p_{\tilde{D}|\tilde{Q}}(d|i)$ is is independent of the X-label of the individual concerned.)

Consider a typical size-h single-household epidemic, with one initial infective, who is of type i and has total degree d, and label the household members $0, 1, \cdots, h-1$, where 0 is the initial infective. For $k = 1, 2, \cdots, h-1$, let $\chi_k = 1$ if individual l is infected by the single-household epidemic and let $\chi_k = 0$ otherwise. Then

$$\tilde{\boldsymbol{C}}_i^{(h,d)} = \tilde{\boldsymbol{C}}_i^{(h,d)}(0) + \sum_{k=1}^{h-1} \chi_k \tilde{\boldsymbol{C}}_i^{(h,d)}(k), \tag{9}$$

where, for $k = 0, 1, \cdots, h-1$, $\tilde{\boldsymbol{C}}_i^{(h,d)}(k) = (\tilde{C}_{i1}^{(h,d)}(k), \tilde{C}_{i2}^{(h,d)}(k), \cdots, \tilde{C}_{in_Q}^{(h,d)}(k))$, with $\tilde{C}_{ij}^{(h,d)}(k)$ being the number of type-j primary infectives generated by individual k in the single-household epidemic if it becomes infected. (Throughout the paper, sums are zero if vacuous.)

Let $T^{(h)} = \sum_{k=1}^{h-1} \chi_k$ be the final size of the above single-household epidemic, not including the initial case, and let $\mu^{(h)}(\lambda) = \mathrm{E}[T^{(h)}]$. Then, see Ball (1986) equations (2.25) and (2.26),

$$\mu^{(h)}(\lambda) = h - 1 - \sum_{k=0}^{h-1} \binom{h-1}{k} \alpha_k \phi_I(k\lambda)^{h-k} \quad (h = 1, 2, \cdots),$$

where $\phi_I(\theta) = \mathrm{E}[\exp(-\theta I)]$ $(\theta \geq 0)$ is the moment generating function of I and $\alpha_0, \alpha_1, \cdots$ are defined recursively by

$$\sum_{l=0}^{k} \binom{k}{l} \alpha_l \phi_I(l\lambda)^{k-l} = k \quad (k = 0, 1, \cdots).$$

Note that χ_k and $\tilde{\boldsymbol{C}}_i^{(h,d)}(k)$ are independent, because whether or not an individual is infected by the single-household epidemic is independent of its infectious period, so taking expectations of (9) and noting that $\tilde{\boldsymbol{C}}_i^{(h,d)}(1), \tilde{\boldsymbol{C}}_i^{(h,d)}(2), \cdots, \tilde{\boldsymbol{C}}_i^{(h,d)}(h-1)$ are identically distributed yields

$$\mathrm{E}[\tilde{C}_{ij}^{(h,d)}] = \mathrm{E}[\tilde{C}_{ij}^{(h,d)}(0)] + \mu^{(h)}(\lambda)\mathrm{E}[\tilde{C}_{ij}^{(h,d)}(1)]. \tag{10}$$

To determine $\mathrm{E}[\tilde{C}_{ij}^{(h,d)}(k)]$ $(k = 0, 1)$, for $i, j = 1, 2, \cdots, n_Q$ and $l = 0, 1$, let $p_{i,j}^{(l)}(r)$ be the probability that, when constructing the network, a given stub with X-label l and total degree quantile i is paired with a stub having total degree quantile j. Then, $p_{i,j}^{(0)} = 1/n_Q$ and

$$p_{i,j}^{(1)}(r) = \begin{cases} \delta_{i,j} & \text{if } r > 0, \\ \delta_{i,n_Q+1-j} & \text{if } r < 0, \end{cases}$$

where $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ if $i \neq j$. Further, for $d = 1, 2, \cdots, j = 1, 2, \cdots, n_Q$ and $l = 0, 1$, let $\tilde{p}^{(l)}_{d,j}(r)$ be the probability that a stub chosen uniformly from all stubs having total degree d and X-label l is paired with a stub from quantile j. Then $\tilde{p}^{(0)}_{d,j}(r) = 1/n_Q$ and

$$\tilde{p}^{(1)}_{d,j}(r) = \sum_{i=1}^{n_Q} p_{\tilde{Q}|\tilde{D}}(i|d) p^{(1)}_{i,j}(r).$$

Consider the individual labelled 0, i.e. the primary case, in the above single-household epidemic. This individual has total degree d and resides in a household of size h, so it has $d - h + 1$ global neighbours, one of whom infected it. Thus the individual has $d - h$ global edges along which it can spread the epidemic. Each of the corresponding stubs independently has X-label 1 with probability $|r|$, so

$$\mathrm{E}[\tilde{C}^{(h,d)}_{ij}(0)] = (d - h)p_I[(1 - |r|)n_Q^{-1} + |r|p^{(1)}_{i,j}], \tag{11}$$

where $p_I = 1 - \phi_I(\lambda)$ is the unconditional probability that a given infective infects a given susceptible neighbour.

Now consider the individual labelled 1 in the single-household epidemic and suppose that it becomes infected. The global degree of individual 1 is distributed according to G. Thus, for $g = 1, 2, \cdots$, with probability p_g, individual 1 has g global neighbours and hence total degree $g + h - 1$. Each of these g global neighbours is infected with probability p_I and the X-labels of the corresponding outgoing stubs from individual 1 are independent Bernoulli random variables with success probability $|r|$. Summing over g and taking expectations yields

$$\mathrm{E}[\tilde{C}^{(h,d)}_{ij}(1)] = \sum_{g=1}^{\infty} p_g g p_I[(1 - |r|)n_Q^{-1} + |r|\tilde{p}^{(l)}_{g+h-1,j}(r)]. \tag{12}$$

Note that if $g = 0$ then individual 1 has no global neighbour to infect. Note also that $\mathrm{E}[\tilde{C}^{(h,d)}_{ij}(1)]$ is independent of both d and i, as indeed is the distribution of $\tilde{\boldsymbol{C}}^{(h,d)}_{i}(1)$.

Combining (8), (10), (11) and (12) gives

$$\begin{aligned}\tilde{m}_{ij} = p_I \sum_{d=1}^{\infty} p_{\tilde{D}|\tilde{Q}}(d|i) \sum_{h=1}^{d} \tilde{\pi}^{(d)}_h \Bigg\{ &(d - h)p_I \left[(1 - |r|)n_Q^{-1} + |r|p^{(1)}_{i,j}\right] \\ &+ \mu^{(h)}(\lambda)\left[(1 - |r|)\mathrm{E}[G]n_Q^{-1} + |r|\sum_{g=1}^{\infty} p_g g \tilde{p}^{(l)}_{g+h-1,j}(r)\right]\Bigg\}. \end{aligned} \tag{13}$$

To summarise, equation (13) defines the elements of the mean matrix $\tilde{M} = [\tilde{m}_{ij}]$ of the branching process $\tilde{\mathcal{B}}_F$. The dominant eigenvalue of $\tilde{M}$, denoted by R_*, determines whether or not a major outbreak is possible, as described at the beginning of the section.

4.1.3 Offspring PGFs and major outbreak probability

We now derive the offspring PGFs for the multitype branching processes $\mathcal{B}_F$ and $\tilde{\mathcal{B}}_F$, which enable their extinction probabilities (and hence the probability of a major outbreak) to be determined. Observe that if the infectious periods are not constant, i.e. there does not exist $\iota > 0$ such that $\mathrm{P}(I = \iota) = 1$, then the infectious periods of individuals infected by a single-household epidemic are not independent of the final size of that epidemic, which complicates, for example, using the decomposition (9) to determine the offspring PGFs of $\tilde{\mathcal{B}}_F$. As in Ball et al. (2010), it is possible to use the theory of final state random variables developed in Ball and O'Neill (1999) to obtain expressions for these offspring PGFs in terms of Gontcharoff polynomials, though the details are rather involved and we do not present them here. Instead, we consider the special case of a constant infection period, when the above-mentioned difficulties do not arise. Thus in this subsection, but not elsewhere in Section 4, we assume that $I \equiv \iota$ (i.e. $\mathrm{P}(I = \iota) = 1$), so any given infective infects each of its neighbours (local or global) independently with probability $p_I = 1 - \exp(-\lambda\iota)$. The epidemic model is then an extension of the standard Reed-Frost epidemic (see, for example, Andersson and Britton (2000), Chapter 1) to our network model. Note also that, in a physics setting, this Reed-Frost type model can be viewed as an extension, to incorporate degree correlation, of the bond percolation model of Gleeson (2009) for a class of clustered networks. Recall also that, as is well known for Reed-Frost type epidemics, the probability and the expected relative final size of a major outbreak are equal (cf. final paragraph of Section 4.2).

As noted previously, the forward branching process $\mathcal{B}_F$ has a different offspring distribution in the initial generation than in all subsequent generations. We consider first a non-initial generation. For $i = 1, 2, \cdots, n_Q$ and $\boldsymbol{s} = (s_1, s_2, \cdots, s_{n_Q})$ with $0 \leq s_i \leq 1$ $(i = 1, 2, \cdots, n_Q)$, let

$$f_{\tilde{\boldsymbol{C}}_i}(\boldsymbol{s}) = \mathrm{E}\left[\prod_{j=1}^{n_Q} s_j^{\tilde{C}_{ij}}\right]$$

be the joint PGF of $\tilde{\boldsymbol{C}}_i$. (Throughout the paper, for a vector random variable, $\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_{n_Q})$ say, we use $f_{\boldsymbol{Y}}(\boldsymbol{s})$ to denote its joint PGF.) Conditioning on the household size and total degree of a typical type-i primary infective, as at (8), yields

$$f_{\tilde{\boldsymbol{C}}_i}(\boldsymbol{s}) = \sum_{d=1}^{\infty} p_{\tilde{D}|\tilde{Q}}(d|i) \sum_{h=1}^{d} f_{\tilde{\boldsymbol{C}}_i^{(h,d)}}(\boldsymbol{s}). \tag{14}$$

The decomposition (9) may be expressed as

$$\tilde{\boldsymbol{C}}_i^{(h,d)} = \tilde{\boldsymbol{C}}_i^{(h,d)}(0) + \sum_{k=1}^{T^{(h)}} \tilde{\boldsymbol{C}}_i^{(h,d)}(k), \tag{15}$$

where now $\tilde{\boldsymbol{C}}_i^{(h,d)}(1), \tilde{\boldsymbol{C}}_i^{(h,d)}(2), \cdots, \tilde{\boldsymbol{C}}_i^{(h,d)}(T^{(h)})$ give the offspring vectors for the $T^{(h)}$ secondary cases in the single-household epidemic. Further, since the infectious period is constant, conditional upon $T^{(h)}$, the random vectors $\tilde{\boldsymbol{C}}_i^{(h,d)}(1), \tilde{\boldsymbol{C}}_i^{(h,d)}(2), \cdots, \tilde{\boldsymbol{C}}_i^{(h,d)}(T^{(h)})$ are independent and identically distributed copies of a random vector whose distribution is independent of $T^{(h)}$. Hence, (15) implies that

$$f_{\tilde{\boldsymbol{C}}_i^{(h,d)}}(\boldsymbol{s}) = f_{\tilde{\boldsymbol{C}}_i^{(h,d)}(0)}(\boldsymbol{s}) f_{T^{(h)}}\left(f_{\tilde{\boldsymbol{C}}_i^{(h,d)}(1)}(\boldsymbol{s})\right), \tag{16}$$

where $f_{T^{(h)}}(s)$ $(0 \leq s \leq 1)$ is the PGF of $T^{(h)}$, which, using Ball (1986), Theorem 2.6, is given by

$$f_{T^{(h)}}(s) = s^{h-1} \sum_{k=0}^{h-1} \binom{h-1}{k} \alpha_k(s)(1-p_I)^{k(h-k)} \quad (h = 1, 2, \cdots), \tag{17}$$

where $\alpha_0(s), \alpha_1(s), \cdots$ are defined recursively by

$$\sum_{l=0}^{k} \binom{k}{l} (1-p_I)^{l(k-l)} \alpha_l(s) = s^{-k} \quad (k = 0, 1, \cdots). \tag{18}$$

To complete the derivation of $f_{\tilde{\boldsymbol{C}}_i}(\boldsymbol{s})$, we obtain expressions for $f_{\tilde{\boldsymbol{C}}_i^{(h,d)}(0)}(\boldsymbol{s})$ and $f_{\tilde{\boldsymbol{C}}_i^{(h,d)}(1)}(\boldsymbol{s})$. Consider a typical type-i primary infective, i^* say, and let j^* be a susceptible global neighbour of i^*. Let $\boldsymbol{\chi}_i = (\chi_{i1}, \chi_{i2}, \cdots, \chi_{in_Q})$, where $\chi_{ik} = 1$ if i^* infects j^* *and* the edge between i^* and j^* was formed by connecting to a stub from j^* belonging to quantile k, and $\chi_{ik} = 0$ otherwise. (Note that if i^* does not infect j^* then every element of $\boldsymbol{\chi}_i$ is zero, and if i^* does infect j^* then precisely one element of $\boldsymbol{\chi}_i$ is one and all other elements of $\boldsymbol{\chi}_i$ are zero.) For $i = 1, 2, \cdots, n_Q$ and $\boldsymbol{s} \in [0,1]^{n_Q}$, define the PGF of $\boldsymbol{\chi}_i$

$$g_i(\boldsymbol{s}) = \mathrm{E}\left[\prod_{j=1}^{n_Q} s_j^{\chi_{ij}}\right] = 1 - p_I + p_I \sum_{j=1}^{n_Q} \left[(1-|r|)\frac{s_j}{n_Q} + |r| p_{i,j}^{(1)}(r) s_j\right]. \tag{19}$$

Then using a similar argument to the derivation of (11) yields

$$f_{\tilde{\boldsymbol{C}}_i^{(h,d)}(0)}(\boldsymbol{s}) = \left(g_d(\boldsymbol{s})\right)^{d-h}. \tag{20}$$

Now consider a typical individual, $\tilde{i}^*$ say, infected by a single-household epidemic and suppose that $\tilde{i}^*$ has total degree d. Let $\tilde{j}^*$ be a susceptible global neighbour of $\tilde{i}^*$ and define $\tilde{\boldsymbol{\chi}}_d = (\tilde{\chi}_{d1}, \tilde{\chi}_{d1}, \cdots, \tilde{\chi}_{dn_Q})$ in the same way as $\boldsymbol{\chi}_i$ but with i^* and j^* replaced by $\tilde{i}^*$ and $\tilde{j}^*$, respectively. Letting

$$\tilde{g}_d(\boldsymbol{s}) = \mathrm{E}\left[\prod_{j=1}^{n_Q} s_j^{\tilde{\chi}_{ij}}\right] = 1 - p_I + p_I \sum_{j=1}^{n_Q} \left[(1-|r|)\frac{s_j}{n_Q} + |r| \tilde{p}_{d,j}^{(1)}(r) s_j\right], \tag{21}$$

a similar argument to the derivation of (12) yields

$$f_{\tilde{\boldsymbol{C}}_i^{(h,d)}(1)}(\boldsymbol{s}) = \sum_{g=0}^{\infty} p_g \left(\tilde{g}_{g+h-1}(\boldsymbol{s})\right)^g. \tag{22}$$

Combining (14), (16), (20) and (22) gives the PGF of the offspring random variable $\tilde{\boldsymbol{C}}_i$ for a typical type-i individual in $\tilde{\mathcal{B}}_F$.

Consider now the initial generation of the forward branching process $\mathcal{B}_F$. Since the initial infective, i^* say, in the epidemic is not infected through the network, the ancestor in $\mathcal{B}_F$ is not typed according to its total degree. Let $\boldsymbol{C} = (C_1, C_2, \cdots, C_{n_Q})$ denote the offspring random variable for the ancestor in $\mathcal{B}_F$. Then, conditioning on i^*'s global degree and household size,

$$f_{\boldsymbol{C}}(\boldsymbol{s}) = \sum_{g=0}^{\infty} \sum_{h=1}^{\infty} p_g \tilde{\pi}_h f_{\boldsymbol{C}^{(h,g+h-1)}}(\boldsymbol{s}), \tag{23}$$

where, for $h = 1, 2, \cdots$ and $d = h+1, h+2, \cdots$, $\boldsymbol{C}^{(h,d)}$ denotes the offspring random variable for the ancestor given that i^* resides in a household of size h and has total degree d. Analogous to (15), $\boldsymbol{C}^{(h,d)}$ admits the decomposition

$$\boldsymbol{C}^{(h,d)} = \boldsymbol{C}^{(h,d)}(0) + \sum_{k=1}^{T^{(h)}} \boldsymbol{C}^{(h,d)}(k), \tag{24}$$

whence, as at (16),

$$f_{\boldsymbol{C}^{(h,d)}}(\boldsymbol{s}) = f_{\boldsymbol{C}^{(h,d)}(0)}(\boldsymbol{s}) f_{T^{(h)}}\left(f_{\boldsymbol{C}^{(h,d)}(1)}(\boldsymbol{s})\right). \tag{25}$$

Now $\boldsymbol{C}^{(h,d)}(1) \overset{D}{=} \tilde{\boldsymbol{C}}^{(h,d)}(1)$, so $f_{\boldsymbol{C}^{(h,d)}(1)}(\boldsymbol{s})$ is given by the right hand side of (22). Note that if i^* has household size h and total degree d, then, since all of its $d-h+1$ global neighbours are susceptible, its offspring distribution is the same as that of a secondary infective having total degree d in a single size-h household epidemic. Thus,

$$f_{\boldsymbol{C}^{(h,d)}(0)}(\boldsymbol{s}) = \left(\tilde{g}_{d-h+1}(\boldsymbol{s})\right)^{d-h+1}. \tag{26}$$

The offspring PGF $f_{\boldsymbol{C}}$ of the ancestor in $\mathcal{B}_F$ now follows using (23), (25), (22) and (26).

We now determine the probability of a major outbreak. Suppose that $R_* > 1$. For $i = 1, 2, \cdots, n_Q$, let σ_i be the probability that the branching process $\tilde{\mathcal{B}}_F$ goes extinct given that there is one ancestor whose type is i, and let $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \cdots, \sigma_{n_Q})$. Then, (see, for example, Mode (1971), Section 1.7.1), $\boldsymbol{\sigma}$ is the unique solution in $[0,1)^{n_Q}$ of the equations

$$f_{\tilde{\boldsymbol{C}}_i}(\boldsymbol{\sigma}) = \sigma_i \quad (i = 1, 2, \cdots, n_Q). \tag{27}$$

By conditioning on the number and type of offspring of the ancestor in $\mathcal{B}_F$, the probability that the branching process $\mathcal{B}_F$ survives (and hence the probability that a major outbreak occurs) is

$$p_{\text{maj}} = 1 - f_{\boldsymbol{C}}(\boldsymbol{\sigma}). \tag{28}$$

4.2 Final outcome of a major outbreak

We now consider the relative final size of a major outbreak. The main tool that we use is the *susceptibility set* (Ball (2000), Ball and Lyne (2001) and Ball and Neal (2002)), which we now define. Label the n nodes (individuals) $1, 2, \cdots, n$. For $i = 1, 2, \cdots, n$, by sampling from the infectious period distribution and the Poisson processes describing when i makes infectious contact with its neighbours, construct a (random) list of who i would have infectious contact with if i was to become infected. Then construct a directed random graph, with nodes $1, 2, \cdots, n$, in which for any pair of nodes (i, j), with $i \neq j$, there is a directed edge from i to j if and only if j is in i's list. For $i = 1, 2, \cdots, n$, the susceptibility set of node i is set of all nodes j from which there is a chain of directed edges to i (including i itself).

Observe that a node, i say, is ultimately infected by the epidemic if and only if the initial infective belongs to i's susceptibility set. Suppose that the population size n is large. Then, as with the early stages of the epidemic, we can approximate the susceptibility set of a node, i^* say, chosen uniformly at random from the population by a households-based multitype branching process. We first consider i^*'s *local* susceptibility set, i.e. the set of nodes in i^*'s household from which there is a chain of within-household directed edges to i^* (including i^* itself). We next consider each member, j^* say, of i^*'s local susceptibility set and determine which of j^*'s global neighbours have a directed edge joining them to j^*. The set of all such global neighbours of i^*'s household form the first generation of the (backward) approximating branching process, with each such global neighbour, k^* say, (generation-1 individual in the branching process) being typed by the quantile of the corresponding stub from k^*. The process is then repeated in the obvious fashion to obtain the second generation of the backward branching process, and so on. Denote this branching process by $\mathcal{B}_B$. As with the forward branching process, the offspring law of $\mathcal{B}_B$ is different in the initial generation from that of all subsequent generations. Let $\tilde{\mathcal{B}}_B$ be the multitype branching process describing the descendants of a typical first-generation individual in $\mathcal{B}_B$.

We conjecture that, subject to mild conditions on the household size and global degree distributions, the expected relative final size of a major outbreak converges to the survival probability of $\mathcal{B}_B$ as $n \to \infty$. This is proved formally in Ball et al. (2009) for the model with constant household size and no global degree correlation (i.e. $r = 0$); however, the proof in Ball et al. (2009) is long and we do not attempt here to adapt it to the present model. Further, assuming the conjecture is true, the argument in Ball et al. (2012) can be used to show that the relative final size of a major outbreak converges in probability to the survival probability of $\mathcal{B}_B$ as $n \to \infty$. The proof in Ball et al. (2012) is also quite long and we do not attempt to adapt it to the present model. The numerical illustrations in Section 6 (see Figure 2 and the surrounding commentary) support the above conjecture.

We determine now the offspring PGFs for $\mathcal{B}_B$ and $\tilde{\mathcal{B}}_B$. We do not assume that the infectious periods are constant. Let $\boldsymbol{B} = (B_1, B_2, \cdots, B_{n_Q})$ denote the offspring random variable for the ancestor in $\mathcal{B}_B$ and, for $i = 1, 2, \cdots, n_Q$, let $\tilde{\boldsymbol{B}}_i = (\tilde{B}_{i1}, \tilde{B}_{i2}, \cdots, \tilde{B}_{in_Q})$ denote the offspring random variable for a typical type-i individual in $\tilde{\mathcal{B}}_B$.

Consider $\tilde{\boldsymbol{B}}_i$ first. Let k^* be as above and assume it has type i. Then arguing as at (14) yields

$$f_{\tilde{\boldsymbol{B}}_i}(\boldsymbol{s}) = \sum_{d=1}^{\infty} p_{\tilde{D}|\tilde{Q}}(d|i) \sum_{h=1}^{d} f_{\tilde{\boldsymbol{B}}_i^{(h,d)}}(\boldsymbol{s}), \tag{29}$$

where $\tilde{\boldsymbol{B}}_i^{(h,d)}$ denotes the corresponding offspring random variable when k^* belongs to a household of size h and has total degree d. Let $M^{(h)} + 1$ denote the size of a typical local susceptibility set in a household of size h. For $l = 0, 1$, let $\tilde{\boldsymbol{B}}_i^{(h,d)}(l) = (\tilde{B}_{i1}(l), \tilde{B}_{i2}(l), \cdots, \tilde{B}_{in_Q}(l))$, where $\tilde{B}_{ij}(0)$ is the number of type-j global neighbours of k^* that would attempt to infect k^* if they become infected and $\tilde{B}_{ij}(1)$ is defined similarly but for any other member of k^*'s local susceptibility set. Then, noting that infectious global neighbours of an individual make infectious contact with that individual independently, each with probability p_I,

$$f_{\tilde{\boldsymbol{B}}_i^{(h,d)}}(\boldsymbol{s}) = f_{\tilde{\boldsymbol{B}}_i^{(h,d)}(0)}(\boldsymbol{s}) f_{M^{(h)}}\left(f_{\tilde{\boldsymbol{B}}_i^{(h,d)}(1)}(\boldsymbol{s})\right),$$

where, for $d = 1, 2, \cdots$ and $h = 1, 2, \cdots, d+1$,

$$f_{\tilde{\boldsymbol{B}}_i^{(h,d)}(0)}(\boldsymbol{s}) = (g_d(\boldsymbol{s}))^{d-h} \qquad \text{and} \qquad f_{\tilde{\boldsymbol{B}}_i^{(h,d)}(1)}(\boldsymbol{s}) = \sum_{g=0}^{\infty} p_g \left(\tilde{g}_{g+h-1}(\boldsymbol{s})\right)^g$$

and $g_i(\boldsymbol{s})$ and $\tilde{g}_i(\boldsymbol{s})$ are defined by (19) and (21).

Turning to the PGF of $\boldsymbol{B}$, similar arguments to the above show that, in an obvious notation,

$$f_{\boldsymbol{B}}(\boldsymbol{s}) = \sum_{g=0}^{\infty} \sum_{h=1}^{\infty} p_g \tilde{\pi}_h f_{\boldsymbol{B}^{(h,g+h-1)}(0)}(\boldsymbol{s}) f_{M^{(h)}} \left(f_{\boldsymbol{B}^{(h,g+h-1)}(1)}(\boldsymbol{s}) \right), \tag{30}$$

where, for $d = 0, 1, \cdots$, and $h = 1, 2, \cdots, d+1$,

$$f_{\boldsymbol{B}^{(h,d)}(0)} = \left(\tilde{g}_{d-h+1}(\boldsymbol{s})\right)^{d-h+1} \qquad \text{and} \qquad f_{\boldsymbol{B}^{(h,d)}(1)} = \sum_{g=0}^{\infty} p_g \left(\tilde{g}_{g+h-1}(\boldsymbol{s})\right)^g .$$

The probability mass function (and hence the PGF) of $M^{(h)}$ may be determined using the following result (see Ball and Neal (2002), Lemma 3.1). For $h = 2, 3, \cdots$,

$$\mathrm{P}(M^{(h)} = k) = \binom{h-1}{k} \phi_I((k+1)\lambda)^{h-1-k} \mathrm{P}(M^{(k)} = k-1) \qquad (k = 0, 1, \cdots, h-1),$$

where

$$\sum_{l=1}^{k} \binom{k-1}{l-1} \phi_I(l\lambda)^{k-l} \mathrm{P}(M^{(l)} = l-1) = 1 \qquad (k = 1, 2, \cdots).$$

It is readily shown that $\mathrm{E}[M^{(h)}] = \mathrm{E}[T^{(h)}]$ $(h = 1, 2, \cdots)$, see Lemma 1 in the appendix of Ball et al. (1997), using which it follows that $\tilde{\mathcal{B}}_B$ and $\tilde{\mathcal{B}}_F$ have the same offspring mean matrix. Thus the branching process $\mathcal{B}_B$ survives if and only if $R_* > 1$. For $i = 1, 2, \cdots, n_Q$, let ξ_i be the probability that the branching process $\tilde{\mathcal{B}}_F$ goes extinct given that there is one ancestor whose type is i, and let $\boldsymbol{\xi} = (\xi_1, \xi_2, \cdots, \xi_{n_Q})$. Then, if $R_* > 1$, $\boldsymbol{\xi}$ is the unique solution in $[0, 1)^{n_Q}$ of the equations

$$f_{\tilde{\boldsymbol{B}}_i}(\boldsymbol{\xi}) = \xi_i \quad (i = 1, 2, \cdots, n_Q)$$

and, for n suitably large, the relative final size of a major outbreak, z say, is given approximately by

$$z = 1 - f_{\boldsymbol{B}}(\boldsymbol{\xi}). \tag{31}$$

There does not appear to exist a similar recursive expression for the PGF $f_{M^{(h)}}(s)$ to that for $f_{T^{(h)}}(s)$ given by (17) and (18), except when the infectious period is constant. In this case $M^{(h)}$ and $T^{(h)}$ have the same distribution, from which it easily follows (using the PGF formulae in the preceding sections) that $p_{\text{maj}} = z$.

5 Epidemics on rewired networks

5.1 Properties of epidemics

We now extend the results of the previous section to the model in which the edges in a fraction p_{RW} of households are rewired.

Suppose first that $p_{RW} = 1$, so all household edges are rewired. The early stages of an epidemic in the rewired network may be approximated by a multitype branching process as in Section 4.1.1, except now a local epidemic is the spread of disease along red edges alone, each having the same household size label. Such local epidemics are realisations of the acquaintance model studied by Diekmann et al. (1998) and a special case of a standard SIR epidemic on a configuration-model random network, see, for example, Newman (2002b). Note that, if n is large, the graph of red edges in the rewired network is locally tree-like. For $h = 2, 3, \cdots$, let $\hat{\mathcal{E}}^{(h)}$ denote an SIR epidemic, with one initial infective, on a tree in which each node has degree $h-1$, with infectious period distributed according to I and infection rate λ. Then for large n, a local epidemic in the rewired process may be approximated by $\hat{\mathcal{E}}^{(h)}$ and all the results of Sections 4.1 and 4.2 continue to hold provided the single-household final size and susceptibility set random variables $T^{(h)}$ and $M^{(h)}$ are replaced by their corresponding rewired counterparts defined on $\hat{\mathcal{E}}^{(h)}$, which we denote by $\hat{T}^{(h)}$ and $\hat{M}^{(h)}$. As usual, the approximation of a local epidemic by $\hat{\mathcal{E}}^{(h)}$ can be made exact in the limit as $n \to \infty$ via a coupling argument.

Each individual in households of size 2 have precisely one red stub, so when the corresponding red stubs are paired up such individuals are partitioned into households of size 2 as before, whence $\hat{T}^{(2)} \overset{D}{=} T^{(2)}$ and $\hat{M}^{(2)} \overset{D}{=} M^{(2)}$. Fix $h \geq 2$ and consider a typical local epidemic $\hat{\mathcal{E}}^{(h)}$. The initial infective in $\hat{\mathcal{E}}^{(h)}$ has $h-1$ susceptible neighbours, while any subsequent infective in the local epidemic has $h-2$ susceptible neighbours. Any given infective infects any given susceptible neighbour with probability $p_I = 1 - \phi_I(\lambda)$. Thus in the (single-type) branching process, $\hat{\mathcal{B}}_F^{(h)}$ say, which gives the size of successive generations of infectives in $\hat{\mathcal{E}}^{(h)}$, the ancestor has offspring mean $(h-1)p_I$ and all subsequent individuals have offspring mean $(h-2)p_I$, whence

$$\hat{\mu}^{(h)}(\lambda) = \mathrm{E}[\hat{T}^{(h)}] = \begin{cases} (h-1)p_I[1-(h-2)p_I]^{-1} & \text{if } p_I < \frac{1}{h-2}, \\ \infty & \text{if } p_I \geq \frac{1}{h-2}. \end{cases} \tag{32}$$

Suppose now that $I \equiv \iota$, so any infective in $\hat{\mathcal{E}}^{(h)}$ infects each of its neighbours independently with probability p_I. Then the offspring distribution of the ancestor in $\hat{\mathcal{B}}_F^{(h)}$ is $\mathrm{Bin}(h-1, p_I)$ and the offspring distribution of any subsequent individual is $\mathrm{Bin}(h-2, p_I)$, where $\mathrm{Bin}(n, p)$ denotes a binomial distribution having n trials and success probability p. Standard branching process arguments then yield that, for $h = 1, 2, \cdots$,

$$f_{\hat{T}^{(h)}}(s) = \left(1 - p_I + p_I \tilde{f}^{(h)}(s)\right)^{h-1} \qquad (0 \leq s \leq 1), \tag{33}$$

where $\tilde{f}^{(h)}(s)$ is the unique solution in $[0,1]$ of the equation

$$\tilde{f}^{(h)}(s) = s\left(1 - p_I + p_I \tilde{f}^{(h)}(s)\right)^{h-2},$$

cf. equations (17) and (18) of Newman(2002b); note that $\tilde{f}^{(h)}(s)$ is the PGF of the total progeny of a typical non-ancestor in $\hat{\mathcal{B}}_F^{(h)}$.

Consider now the branching process, $\hat{\mathcal{B}}_B^{(h)}$ say, that describes on a generation basis a typical local susceptibility set associated with $\hat{\mathcal{E}}^{(h)}$ and return to the case of a general infectious period distribution. It is easily seen that the offspring distributions of the ancestor and any subsequent individual in $\hat{\mathcal{B}}_B^{(h)}$ are $\mathrm{Bin}(h-1,p_I)$ and $\mathrm{Bin}(h-2,p_I)$, respectively, where $p_I = \phi_I(\lambda)$, so $f_{\hat{M}^{(h)}}(s)$ is given by the right hand side of (33).

Finally we consider the case when the rewiring probability $p_{RW} \in (0,1)$. Then, for example, the size $T^{(h)}(p_{RW})$ of a typical local epidemic corresponding to households having size h is distributed according to $\hat{T}^{(h)}$, with probability p_{RW}, and to $T^{(h)}$, with probability $1-p_{RW}$. Thus, $\mathrm{E}[T^{(h)}(p_{RW})] = (1-p_{RW})\mu^{(h)}(\lambda) + p_{RW}\hat{\mu}^{(h)}(\lambda)$, $f_{T^{(h)}(p_{RW})}(s) = (1-p_{RW})f_{T^{(h)}}(s) + p_{RW}f_{\hat{T}^{(h)}}(s)$ and $f_{M^{(h)}(p_{RW})}(s) = (1-p_{RW})f_{M^{(h)}}(s) + p_{RW}f_{\hat{M}^{(h)}}(s)$. The threshold parameter R_*, probability of a major epidemic p_{maj} and relative final size of a major outbreak z now follow by appropriate substitution into the results in Sections 4.1 and 4.2.

5.2 Effect of rewiring

We now examine the qualitative effect of rewiring on the probability and relative final size of a major outbreak. For the model with $r=0$, constant infectious period and fixed household size (i.e. $\mathrm{P}(H=h)=1$ for some h), Gleeson et al. (2010) use an analytic argument to show that the bond percolation threshold (the value of p_I so that $R_*=1$) is larger for the model with full rewiring ($p_{RW}=1$) than for the model with no rewiring ($p_{RW}=0$). Miller (2009) proves a similar result, again using an analytic argument, for an alternative model of random clustered networks, involving triangles, and also shows that the relative final size z of a major outbreak is smaller for the fully rewired network than for the corresponding model without rewiring. Here we employ a coupling argument, similar to that in, for example, Mollison (1977) and Ball (1983), to prove that for our model R_*, p_{maj} and z are all increasing functions of the rewiring probability p_{RW}. The coupling argument is both intuitive and powerful. It may be extended to the model of Gleeson et al. (2010), without the restriction of a common household size, to the models of Miller (2009) and Newman (2009), and to the extension of the latter model proposed by Karrer and Newman (2010) that incorporates more general subgraphs than triangles.

For $h=1,2,\cdots$, let $\mathcal{E}^{(h)}$ denote the single size-h household epidemic introduced in Section 4.1.2, so $T^{(h)}$ is the final size of $\mathcal{E}^{(h)}$ not including the initial infective. For fixed $h \geq 2$, a realisation of $\mathcal{E}^{(h)}$, viewed in generations of infectives, may be constructed from a realisation of $\hat{\mathcal{B}}_F^{(h)}$ as follows. The ancestor of $\hat{\mathcal{B}}_F^{(h)}$ corresponds to the initial infective in $\mathcal{E}^{(h)}$. The number of individuals, Z_1 say, in the first generation in $\hat{\mathcal{B}}_F^{(h)}$ (i.e. the offspring of the ancestor) give the number of people directly infected by the initial infective in $\mathcal{E}^{(h)}$. The individuals so infected are obtained by sampling Z_1 individuals uniformly at random without replacement from the $h-1$ individuals in the household excluding the initial infective. The sampled individuals form the first generation of infectives in $\mathcal{E}^{(h)}$. We now consider each first-generation individual in the branching process $\hat{\mathcal{B}}_F^{(h)}$ in turn. The immediate offspring of such a first-generation individual give the number of people with which the corresponding infective in $\mathcal{E}^{(h)}$ makes infectious contact. The people so contacted are obtained by sampling uniformly at random without replacement from the $h-1$ individuals in the household excluding the infective under consideration. It is possible that a person so contacted has already been infected in $\mathcal{E}^{(h)}$, in which case the corresponding birth in $\hat{\mathcal{B}}_F^{(h)}$ and all of the descendants of that individual in $\hat{\mathcal{B}}_F^{(h)}$ are ignored in the construction of $\mathcal{E}^{(h)}$. The construction of $\mathcal{E}^{(h)}$ continues in the obvious fashion and terminates when there is no infective remaining in the household.

Observe that by construction the size of the epidemic $\mathcal{E}^{(h)}$ is not larger than that the total progeny of the branching process $\hat{\mathcal{B}}_F^{(h)}$, so $\hat{T}^{(h)} \overset{st}{\geq} T^{(h)}$, where $\overset{st}{\geq}$ denotes stochastic ordering, whence $\hat{\mu}^{(h)}(\lambda) \geq \mu^{(h)}(\lambda)$ and $f_{\hat{T}^{(h)}}(s) \leq f_{T^{(h)}}(s)$ $(0 \leq s \leq 1)$. Moreover, provided $\lambda\mu_I > 0$, these inequalities are strict for all $h \geq 3$ and all $s \in [0, 1)$. It follows that, if all other parameters are held fixed, the threshold parameter R_* is an increasing function of the rewiring probability p_{RW}, as is the probability of a major outbreak p_{maj} (assuming that the infectious period is constant). When the infectious period is not constant, the above coupling can be extended to include the global degrees of individuals in such a way that infectives in the household epidemic $\mathcal{E}^{(h)}$ have the same global degree and make the same global infectious contacts as the corresponding individuals in the branching process $\hat{\mathcal{B}}_F^{(h)}$, from which it follows that p_{maj} is increasing in p_{RW}. Moreover, if $\mathrm{P}(H \geq 3) > 0$ and $\lambda\mu_I > 0$ then both R_* and p_{maj} are strictly increasing in p_{RW}.

Turning to the final outcome of a major outbreak, for fixed $h \geq 2$, we can construct a realisation of the local susceptibility set $\mathcal{S}^{(h)}$ say, of an individual, i^* say, who resides in a household of size h, from a realisation of the branching process $\hat{\mathcal{B}}_B^{(h)}$ as follows. The local susceptibility set of i^* is constructed on a generation basis. The ancestor of $\hat{\mathcal{B}}_B^{(h)}$ corresponds to the individual i^*. The first generation of $\hat{\mathcal{B}}_B^{(h)}$ gives the number of individuals in i^*'s household who would make infectious contact with i^* if they were to become infected; who these individuals (who form the first generation of $\mathcal{S}^{(h)}$) are is then determined by sampling without replacement as above. We next consider in turn each member, j^* say, of the first generation of $\mathcal{S}^{(h)}$ and determine which of those individuals not currently in $\mathcal{S}^{(h)}$ would join the susceptibility set of i^* by virtue of making infectious contact with j^*. Suppose that j^* is the kth first-generation member of $\mathcal{S}^{(h)}$ to be considered in this fashion. Then any individual not currently in $\mathcal{S}^{(h)}$ has failed to infect k individuals, so the probability that it fails to infect j^* is given by $p_F(k) = \phi_I((k+1)\lambda)/\phi_I(k\lambda)$. Moreover, since such individuals are distinct, they each fail to infect j^* independently with probability $p_F(k)$. Let $p_F(0) = \phi_I(\lambda)$. We now prove that, as one would expect on intuitive grounds, for any $\lambda > 0$, $p_F(k) \geq p_F(0)$ $(k = 1, 2, \cdots)$, with strict inequality unless $I \equiv \iota$ for some $\iota \geq 0$.

Define the function η by $\eta(\theta) = \log \phi_I(\theta)$ $(\theta \geq 0)$. Then η is a convex function, since ϕ_I is a moment generating function, and $\eta(0) = 0$. Thus, $\eta(\lambda) \leq \frac{1}{k+1}\eta((k+1)\lambda)$ and $\eta(k\lambda) \leq \frac{k}{k+1}\eta((k+1)\lambda)$, whence

$$\eta(\lambda) + \eta(k\lambda) \leq \eta((k+1)\lambda), \tag{34}$$

which implies that $p_F(k) \geq p_F(0)$ $(k = 1, 2, \cdots)$. Moreover, if the infectious period random variable I is not almost surely constant then η is a strictly convex function, so, provided $\lambda > 0$, the inequality in (34) is strict and $p_F(k) > p_F(0)$ $(k = 1, 2, \cdots)$.

In view of the above result, the individuals who join the susceptibility set $\mathcal{S}^{(h)}$ by virtue of making infectious contact with j^* may be determined as follows. Let Z_{j^*} be the number of immediate offspring of the individual in $\hat{\mathcal{B}}_B^{(h)}$ that corresponds to j^* and note that $Z_{j^*} \sim \mathrm{Bin}(h-2, 1-p_F(0))$. Given Z_{j^*}, sample $\hat{Z}_{j^*}$ from the binomial distribution $\mathrm{Bin}\left(Z_{j^*}, \frac{1-p_F(k)}{1-p_F(0)}\right)$ and then sample Z_{j^*} individuals uniformly at random without replacement from the $h-1$ individuals in the household excluding j^*. Any individual in this latter sample that is not currently in $\mathcal{S}^{(h)}$ is added to $\mathcal{S}^{(h)}$. This process is repeated for all j^* belonging to the first generation of $\mathcal{S}^{(h)}$, thus yielding the second generation of $\mathcal{S}^{(h)}$, and so on. Observe that, by construction, any individual in $\mathcal{S}^{(h)}$ has a corresponding individual in $\hat{\mathcal{B}}_B^{(h)}$, so $\hat{M}^{(h)} \overset{st}{\geq} M^{(h)}$, whence $f_{\hat{M}^{(h)}}(s) \leq f_{M^{(h)}}(s)$ $(0 \leq s \leq 1)$, with strict inequality for $h \geq 3$ and $0 \leq s < 1$ provided $\lambda\mu_I > 0$. It follows that the relative final size z of a major outbreak is increasing in the rewiring probability p_{RW}, and strictly increasing if $\mathrm{P}(H \geq 3) > 0$ and $\lambda\mu_I > 0$.

6 Numerical examples

In this section we explore some properties of our network epidemic model numerically. We restrict our attention to the Reed-Frost type version of our model, i.e. we assume that $I \equiv \iota$ for some $\iota > 0$, which implies that $p_{\mathrm{maj}} = z$, and rather than dealing explicitly with I and the contact rate λ we refer to the marginal infection probability $p_I = 1 - \exp(\lambda\iota)$. Also, we use the notation Poi and Poi^+ for global degree and household size distributions, as in Section 3.5.

First we briefly investigate the convergence of p_{maj} and z for finite populations (derived empirically from simulations) to the asymptotic values (derived analytically) as the number of nodes/individuals n becomes large. Figure 2 shows this behaviour in p_{maj} and z, for fixed G, H, n_Q, p_I and varying $r \in [-1, 1]$, comparing the asymptotic results to empirical estimates from networks of size $n =$ 1,000 and 10,000 nodes/individuals. Each empirical estimate of a quantity of interest is based on $n_0 = 1,000$ simulations and is represented by an approximate 95.4% confidence interval, calculated as a point estimate ± 2 standard errors (SE). (Also note that each simulation consists of generating a network then running an epidemic on it; we do not just run 1,000 epidemics on a single randomly generated network.) Each point estimate of p_{maj} is simply the proportion $\hat{p}$ of simulations that took off into a major outbreak (the cutoff between minor and major outbreaks being determined by inspecting histograms of epidemic final size), and $\mathrm{SE} = (\hat{p}(1-\hat{p})/n_0)^{1/2}$. The point estimate of z is the mean fraction of the population ultimately infected by a major outbreak and here $\mathrm{SE} = \hat{\sigma} n_1^{-1/2}$, where $\hat{\sigma}^2$ is the sample variance of the fraction of the population ultimately infected by a major outbreak and n_1 is the number of simulations that resulted in a major outbreak. As was explained in the closing sentences of Section 5 of Ball et al. (2009), our simulation methods yield much tighter confidence bands for z than for p_{maj} since each simulation effectively gives a single realisation of the epidemic process but each simulation that results in a major outbreak gives $n-1$ (highly correlated) realisations of the susceptibility set process.

We see that for networks with only 1000 nodes the asymptotic values of p_{maj} seem to be very good approximations to the empirically calculated major outbreak probabilities across all values of r. The expected relative final size also seems to be well approximated by the asymptotic values even for $n = 1,000$; though there does appear to be some

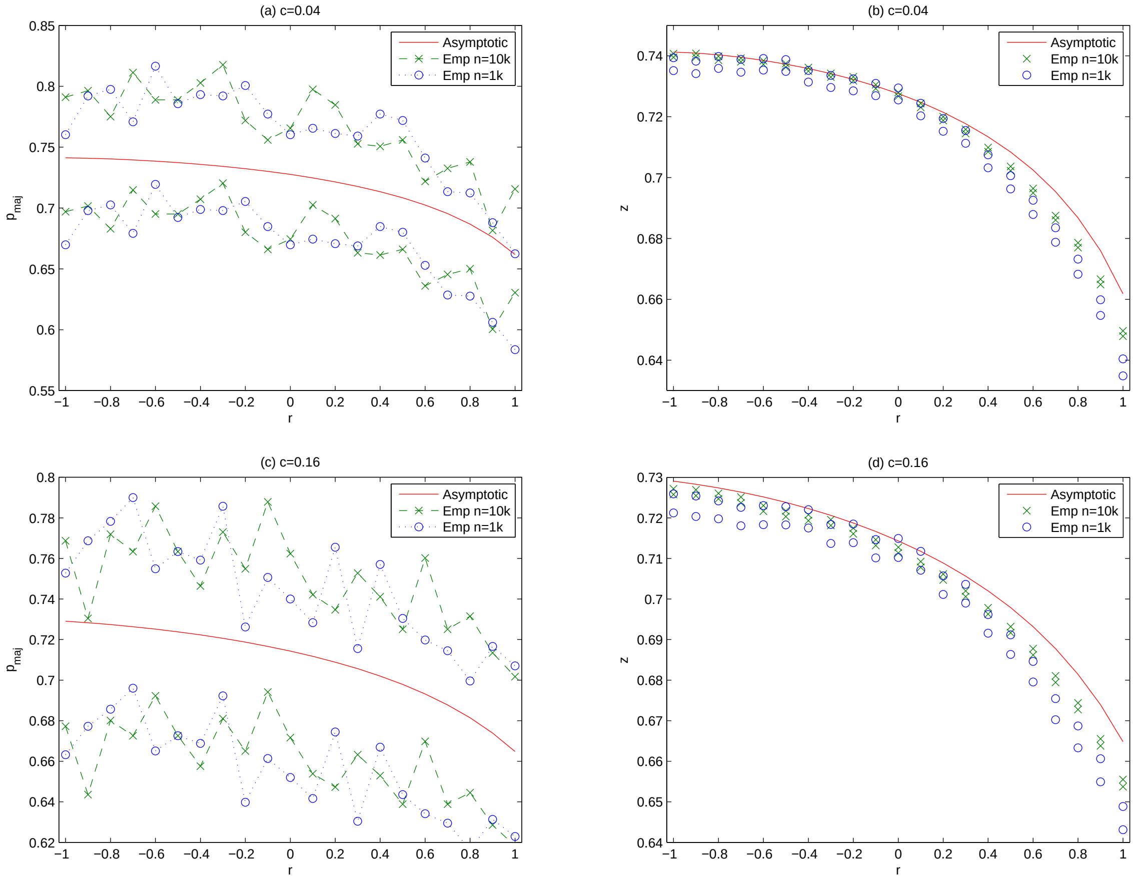

Figure 2: Plots comparing empirical estimates ($n < \infty$) and asymptotic values ($n \to \infty$) of p_{maj} and z, as a function of r, for our model with degree distributions $H \sim \text{Poi}^+(2)$ and $G \sim \text{Poi}(8)$ ($c = 0.04$) and $H \sim \text{Poi}^+(4)$ and $G \sim \text{Poi}(6)$ ($c = 0.16$). Other parameters are $n_Q = 10$ and $p_I = 0.2$. Empirical estimates are for network sizes $n = 1,000$ and $n = 10,000$, each estimate being based on 1,000 simulations. Note that the scales on the vertical axis on these plots is very variable.

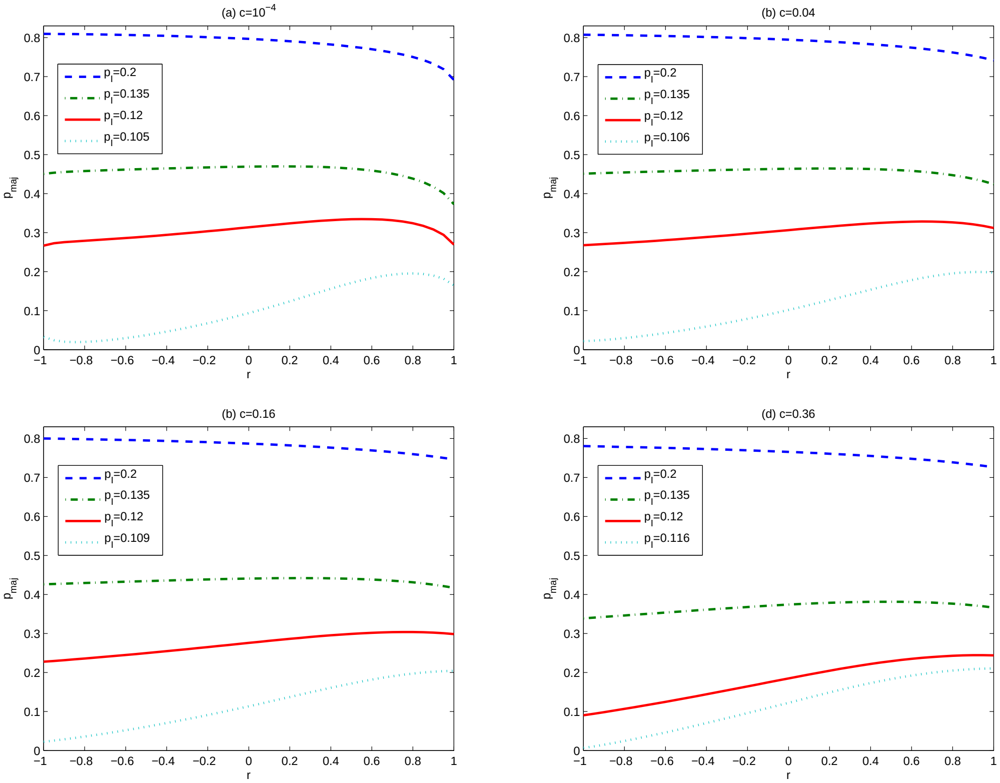

Figure 3: Plot of p_{maj} versus r for varying values of p_I. $G \sim \mathrm{Poi}(10-\mu)$ and $H \sim \mathrm{Poi}^+(\mu)$, with μ taking the values, in order, 0.1, 2, 4, 6; corresponding to clustering coefficients $10^{-4}, 0.04, 0.16, 0.36$. Note also that the p_I values used are the same in each plot except for the smallest value, which is chosen so that the epidemic is just supercritical for all values of r.

bias, which is more pronounced for more extreme values of r. One explanation for this is that when r is close to -1 or 1, there are more imperfections in the random graph (self-loops, household self-loops, etc.) and so the branching process approximation breaks down sooner. Nevertheless, the z plots lend considerable credence to our conjecture in Section 4.2 that the expected relative final size of a major outbreak converges to the survival probability of $\mathcal{B}_B$ as $n \to \infty$.

Having seen that our asymptotic results give reasonable descriptions of the behaviour of our epidemic model on a moderately sized finite network, we turn our attention to investigating the effect of some of the parameters of our model on its (asymptotic) behaviour. We focus initially on the qualitative behaviour of $p_{\mathrm{maj}}(= z)$ considered as a function of r (and p_I). Figure 3 illustrates this behaviour in the case where $G \sim \mathrm{Poi}(10-\mu)$, $H \sim \mathrm{Poi}^+(\mu)$, so $D \sim \mathrm{Poi}(10)$, and $n_Q = 10$, for various values of $\mu \in [0, 10)$ (and therefore $c = (\mu/10)^2$).

We see a variety of patterns in the dependance of p_{maj} on r as p_I and c are varied. Broadly, when the process is well above criticality the dependance is not very strong, but when the process is only just supercritical changes in r in particular (and thus in the degree correlation) can have a substantial impact on the epidemic model. The interesting (and somewhat unexpected) qualitative behaviour observed in the $p = 0.105$ line in plot (a) is explored in further detail in Figure 4. Note, however, that the model parameters that

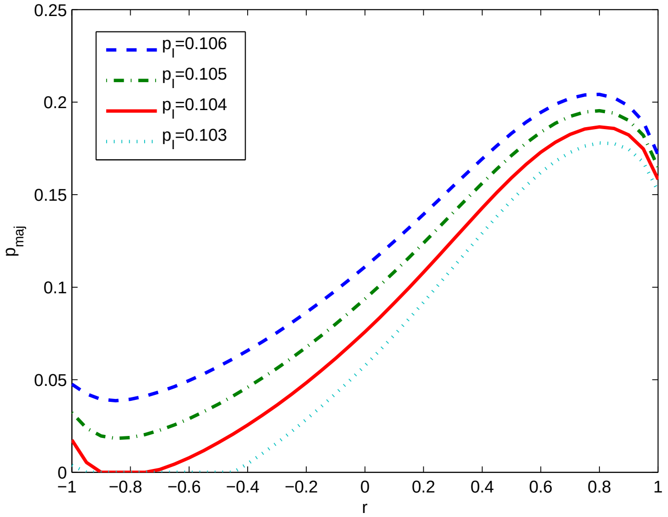

Figure 4: Plot of p_{maj} versus r for near-critical values of p_I, when $G \sim \text{Poi}(9.9)$, $H \sim \text{Poi}^+(0.1)$ and $n_Q = 10$. (Note that the $p_I = 0.103$ line is positive near $r = -1$.)

give rise to this behaviour are $\mu_G = 9.9$ and $\mu_H = 0.1$, so there is essentially no clustering in the network; clearly further work is required to determine whether the model behaves in such a way with other, more realistic parameter values. Nevertheless, the wide range of values of $p_{\text{maj}}(= z)$ for different values of r (i.e. degree correlation) are observed near criticality in all of the plots in Figure 3; even though the non-monotonicity is only observed in plot (a).

Finally, Figure 5 illustrates the effect on $p_{\text{maj}}(= z)$ of changing c, keeping r and p_I fixed, for the case when the total degree $D \sim \text{Poi}(10)$ and $n_Q = 10$. The degree correlation ρ is held fixed at $\rho = 0.2$ and, for the unrewired model, the clustering coefficient c is tuned to be any value in its feasible range (see Figure 1) by varying μ and using (7). The maximum value of c, consistent with $\rho = 0.2$, is $c = 0.4855$, which is attained when $r = -1$ and $\mu = 6.9676$. For the rewired model, the clustering coefficient is tuned by taking the unrewired model with $r = -1$ and $\mu = 6.9676$ and letting the rewiring probability p_{RW} vary in $[0, 1]$. Figure 5 shows how $p_{\text{maj}}(= z)$ varies with c for both the unrewired and rewired models. Note that, as one might expect, $p_{\text{maj}}(= z)$ decreases with c for both models; indeed this is proved formally for the rewired model in Section 5.2. Note also that $p_{\text{maj}}(= z)$ is different for the two models, illustrating that these epidemic properties depend on more than just the local properties of the network encapsulated in (D, c, r).

7 Discussion

In this paper we define a network model which allows for quite arbitrary clustering c, degree correlation ρ and degree distribution D, and asymptotic features of the model are derived. The main focus is on analysing an epidemic model on the network, and in particular what effect various network properties have on the epidemic in terms of its

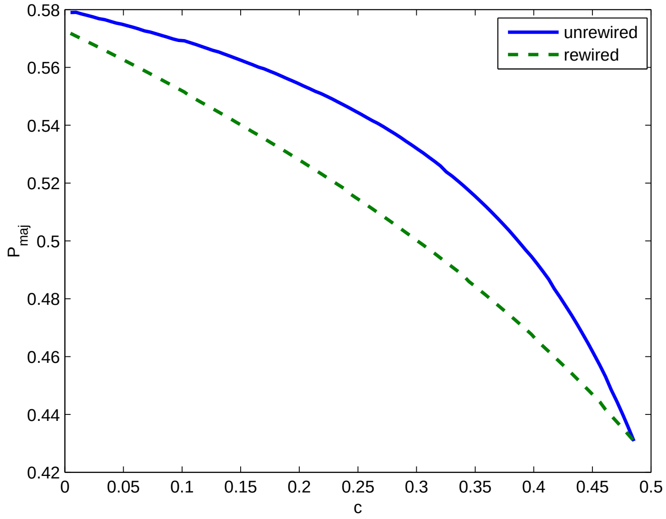

Figure 5: Plot of $p_{\text{maj}}(= z)$ versus c when $D \sim \text{Poi}(10), \rho = 0.2, n_Q = 10$ and $p_I = 0.15$.

threshold parameter R_*, the probability p_{maj} of a major outbreak, and the relative size z of a major outbreak. The main conclusion is that all three quantities R_*, p_{maj} and z are decreasing with the clustering coefficient c (when rewiring edges in the network thus keeping everything else fixed), whereas the dependence on the degree correlation ρ is not as easily expressed: the quantities may be either increasing or decreasing depending on which part of the parameter space is being investigated. To our knowledge this is the first network model having such general features for which the properties of an epidemic are analysed in this level of detail.

A disadvantage with the model is that, in general, there is no simple and explicit relation between the model parameters H, G, r, and n_Q and the more interesting network properties c, ρ and D. Note however the relation for D given in equation (1), and the facts that ρ is increasing with r and c is increasing in H (in the sense that $c(G, H_1, r) \geq c(G, H_2, r)$ if $H_1 \overset{st}{\geq} H_2$), keeping other parameters fixed. A model having simpler relationships to the local network properties could be more easily interpreted and would hence be of interest. The use of appropriate pairing of stubs to control degree correlation, as done in this paper, could be applied to other models of clustered networks, such as those in Newman (2009), Miller (2009) and Karrer and Newman (2010).

It is important to observe that, as illustrated in Figure 5, there may be distinct network models having the same local network features D, ρ and c but still giving different properties of an epidemic, the latter being a global property. In applications it is hence important to fit not only local properties of a network model to empirical network data, but also to study the definitions of the model and try to understand if the model mechanism seems to agree realistically with how the empirical network may have been constructed.

Acknowledgments This research was supported by the UK Engineering and Physical Sciences Research Council, under research grant number EP/E038670/1 (FB and DS), and by the Swedish Research Council (TB).

Appendix: Derivation of degree correlation ρ

In the appendix we derive the formula for the degree correlation ρ for our model given in equation (6). Let E denote an edge chosen uniformly at random from all edges in the network, and let X_L and X_R denote the total degrees of the nodes adjacent to E. Then $\rho = \mathrm{corr}(X_L, X_G)$, i.e. the correlation between X_L and X_R. Let $I_G = 1$ if E is a global edge and $I_G = 0$ if E is a household edge, so $\mathrm{P}(I_G = 1) = p_G = 1 - \mathrm{P}(I_G = 0)$. We determine first the probability p_G that E is a global edge.

Let N_G and N_H denote respectively the number of global and household edges in the network. Then $\mu_{N_G} = \frac{n}{2}\mu_G$, since each stub contributes to half an edge, and $\mu_{N_H} = \frac{n}{2}\mu_{\tilde{H}-1}$, since the household size of an individual chosen unifomly at random from the population is distributed according to $\tilde{H}$ and if such an individual resides in a household of size h it has $h-1$ household neighbours. Letting $n \to \infty$ and using the strong law of large numbers shows that p_G is given by (3).

Note that

$$\mathrm{cov}(X_L, X_R) = \mathrm{E}[\mathrm{cov}(X_L, X_R|I_G)] + \mathrm{cov}(\mathrm{E}[X_L|I_G], \mathrm{E}[X_R|I_G]). \tag{35}$$

We calculate the two quantities on the right hand side of (35) in turn.

Suppose that $I_G = 0$, so E is a household edge. Then $X_L = H_E - 1 + G_L$ and $X_R = H_E - 1 + G_R$, where H_E is the size of the household that contains the edge E, and G_L and G_R are the global degrees of the nodes adjacent to E. Observe that H_E is distributed as $\hat{H}$ and, since $I_G = 0$, G_L and G_R are independent copies of G. Thus,

$$\mathrm{cov}(X_L, X_R|I_G = 0) = \sigma^2_{\hat{H}}. \tag{36}$$

Suppose that $I_G = 1$, so E is a global edge. Let Q_L and Q_R be the total degree quantiles of the two stubs used to form the edge E. Then, for $i, j = 1, 2, \cdots, n_Q$,

$$\mathrm{P}(Q_L = i, Q_R = j) = \begin{cases} \frac{1-r}{n_Q^2} + \delta_{i,j}\frac{r}{k} & \text{if } r \geq 0, \\ \frac{1-|r|}{n_Q^2} + \delta_{i,n_Q+1-j}\frac{|r|}{k} & \text{if } r < 0. \end{cases} \tag{37}$$

Now,

$$\begin{aligned} \mathrm{cov}(X_L, X_R|I_G = 1) = {} & \mathrm{E}[\mathrm{cov}(X_L, X_R|I_G = 1, Q_L, Q_R)] \\ & + \mathrm{cov}(\mathrm{E}[X_L|I_G = 1, Q_L], \mathrm{E}[X_R|I_G = 1, Q_R]). \end{aligned} \tag{38}$$

Given (Q_L, Q_R), the total degrees X_L and X_R are independent, so

$$\mathrm{cov}(X_L, X_R|I_G = 1, Q_L, Q_R) = 0. \tag{39}$$

Further, for $i = 1, 2, \cdots, n_Q$, $\mathrm{E}[X_L|I_G = 1, Q_L = i] = \mathrm{E}[X_L|I_G = 1, Q_R = i] = \mu^{(i)}_{\tilde{D}}$ (see equation (5)). Using the distribution (37) and noting that $\mu_{\tilde{D}} = n_Q^{-1}\sum_{i=1}^{n_Q}\mu^{(i)}_{\tilde{D}}$ yields

$$\mathrm{cov}(\mathrm{E}[X_L|I_G = 1, Q_L], \mathrm{E}[X_R|I_G = 1, Q_R]) = g_{\tilde{D},n_Q}(r), \tag{40}$$

where $g_{\tilde{D},n_Q}(r)$ is defined at (4).

Note that $\mathrm{P}(I_G = 1) = p_G = 1 - \mathrm{P}(I_G = 0)$. Then, equations (36), (38), (39) and (40) yield

$$\mathrm{E}[\mathrm{cov}(X_L, X_R | I_G)] = (1 - p_G)\sigma^2_{\hat{H}} + p_G g_{\tilde{D},n_Q}(r). \tag{41}$$

We turn now to the second quantity on the right hand side of (35). Note that $\mathrm{E}[X_L|I_G] = \mathrm{E}[X_R|I_G]$, so $\mathrm{cov}(\mathrm{E}[X_L|I_G], \mathrm{E}[X_R|I_G]) = \mathrm{var}(\mathrm{E}[X_L|I_G])$. Suppose that $I_G = 0$. Then, in the above notation, $X_L = H_E - 1 + G_L$, where $G_L \overset{D}{=} G$. Thus,

$$\mathrm{E}[X_L|I_G = 0] = \mu_{\hat{H}-1} + \mu_G. \tag{42}$$

Suppose that $I_G = 1$. Then $X_L \overset{D}{=} \tilde{D}$ and recall that $\tilde{D} \overset{D}{=} \tilde{H} - 1 + \tilde{G}$. Thus,

$$\mathrm{E}[X_L|I_G = 1] = \mu_{\tilde{H}-1} + \mu_{\tilde{G}}. \tag{43}$$

Recalling that $\mathrm{P}(I_G = 1) = p_G = 1 - \mathrm{P}(I_G = 0)$ and that $\mu_{\tilde{G}} = \mathrm{E}[G^2]/\mu_G$, equations (42) and (43) yield

$$\mathrm{cov}(\mathrm{E}[X_L|I_G], \mathrm{E}[X_R|I_G]) = p_G(1 - p_G)\left(\mu_{\hat{H}} - \mu_{\tilde{H}} - \frac{\sigma^2_G}{\mu_G}\right)^2. \tag{44}$$

Combining equations (35), (41) and (44) gives

$$\mathrm{cov}(X_L, X_R) = (1 - p_G)\sigma^2_{\hat{H}} + p_G g_{\tilde{D},n_Q}(r) + p_G(1 - p_G)\left(\mu_{\hat{H}} - \mu_{\tilde{H}} - \frac{\sigma^2_G}{\mu_G}\right)^2. \tag{45}$$

We now derive $\mathrm{var}(X_L)$. First note that

$$\mathrm{var}(X_L) = \mathrm{E}[\mathrm{var}(X_L|I_G)] + \mathrm{var}(\mathrm{E}[X_L|I_G]). \tag{46}$$

As above, if $I_G = 0$ then $X_L = H_E - 1 + G_L$, where $H_E \overset{D}{=} \hat{H}$ and $G_L \overset{D}{=} G$ are independent, so $\mathrm{var}(X_L|I_G = 0) = \sigma^2_{\hat{H}} + \sigma^2_G$; and if $I_G = 1$ then $X_L \overset{D}{=} \tilde{H} - 1 + \tilde{G}$, where $\tilde{H}$ and $\tilde{G}$ are independent, so $\mathrm{var}(X_L|I_G = 1) = \sigma^2_{\tilde{H}} + \sigma^2_{\tilde{G}}$. Hence,

$$\mathrm{E}[\mathrm{var}(X_L|I_G)] = (1 - p_G)\left(\sigma^2_{\hat{H}} + \sigma^2_G\right) + p_G\left(\sigma^2_{\tilde{H}} + \sigma^2_{\tilde{G}}\right),$$

which on substituting into (46), recalling that $\mathrm{var}(\mathrm{E}[X_L|I_G]) = \mathrm{cov}(\mathrm{E}[X_L|I_G], \mathrm{E}[X_R|I_G])$ and using (44) yields

$$\mathrm{var}(X_L) = (1 - p_G)\left(\sigma^2_{\hat{H}} + \sigma^2_G\right) + p_G\left(\sigma^2_{\tilde{H}} + \sigma^2_{\tilde{G}}\right) + p_G(1 - p_G)\left(\mu_{\hat{H}} - \mu_{\tilde{H}} - \frac{\sigma^2_G}{\mu_G}\right)^2. \tag{47}$$

The expression (6) for the degree correlation ρ, given in Section 3.3, follows from equations (45) and (47), since $\mathrm{var}(X_L) = \mathrm{var}(X_R)$.

References

Andersson, H. (1999), Epidemic models and social networks, *The Mathematical Scientist* **24(2)** 128–147.

Andersson, H. and Britton, T. (2000), Stochastic epidemic models and their statistical analysis, *Springer Lecture Notes in Statistics* **151**, New York: Springer Verlag.

Badham, J. and Stocker, R. (2010), The impact of network clustering and assortativity on epidemic behaviour, *Theor. Pop. Biol.* **77** 71–75.

Ball, F.G. (1983), The threshold behaviour of epidemic models, *J. Appl. Prob.* **20** 227–241.

Ball, F.G. (1986), A unified approach to the distribution of total size and total area under the trajectory of the infectives in epidemic models, *Adv. Appl. Prob.* **18** 289–310.

Ball, F.G. (2000), Susceptibility sets and the final outcome of stochastic SIR epidemic models. Research Report 00-09. Division of Statistics, School of Mathematical Sciences, University of Nottingham.

Ball, F.G. and Lyne, O.D. (2001), Stochastic multitype SIR epidemics among a population partitioned into households, *Adv. Appl. Prob.* **33** 99–123.

Ball, F.G.; Mollison, D. and Scalia-Tomba, G. (1997), Epidemics with two levels of mixing, *Ann. Appl. Prob.* **7** 46-89.

Ball, F.G. and Neal, P. (2002), A general model for stochastic SIR epidemics with two levels of mixing, *Math. Biosci.* **180** 73–102.

Ball, F.G. and O'Neill, P.D. (1999), The distribution of general final state random variables for stochastic epidemic models, *J. Appl. Prob.* **36** 473–491.

Ball, F.G. and Sirl, D.J. (2012), An SIR epidemic model on a population with random network and household structure, and several types of individuals, *Adv. Appl. Prob.* **44** 63–86.

Ball, F.G.; Sirl, D.J. and Trapman, P. (2009), Threshold behaviour and final outcome of an epidemic on a random network with household structure, *Adv. Appl. Prob.* **41** 765–796.

Ball, F.G.; Sirl, D.J. and Trapman, P. (2010), Analysis of a stochastic SIR epidemic on a random network incorporating household structure, *Math. Biosci.* **224(2)** 53–73.

Ball, F.G.; Sirl, D.J. and Trapman, P. (2012), Epidemics on random intersection graphs, Submitted.

Barabási, A. and Albert, R. (1999), Emergence of scaling in random networks, *Science* **286** 509–512.

Britton T.; Nordvik, M.K. and Liljeros, F. (2007) Modelling sexually transmitted infections: the effect of partnership activity and number of partners on R_0, *Theor. Pop. Biol.* **72** 389-399.

Britton T.; Deijfen, M.; Lindholm, M. and Lagerås, A.N. (2008), Epidemics on random graphs with tunable clustering, *J. Appl. Prob.* **45** 743–756.

Diekmann, O. and Heesterbeek, J.A.P. (2000), *Mathematical Epidemiology of Infectious Diseases*, Chichester: John Wiley & Son.

Diekmann, O.; de Jong, M.C.M. and Metz, J.A.J. (1998), A deterministic epidemic model taking account of repeated contacts between the same individuals, *J. Appl. Prob.* **35** 448–462.

Erdős, P. and Rényi, A. (1959), On random graphs. *Publicationes Mathematicae* **6**, 290-297.

Gleeson, J.P. (2009), Bond percolation on a class of clustered random networks, *Phys. Rev. E* **80**, 036107.

Gleeson, J.P.; Melnik, S. and Hackett, A. (2010), How clustering affects the bond percolation threshold in complex networks, *Phys. Rev. E* **81**, 066114.

van der Hofstad, R. and Litvak, N. (2012), Degree-degree correlations in random graphs with heavy-tailed degrees, arXiv:1202.307v3.

Isham, V., Kaczmarska, J. and Nekovee, M. (2011), Spread of information and infection on finite random networks. *Phys. Rev. E* **83**, 046128.

Karrer, B. and Newman, M.E.J. (2010), Random graphs containing arbitrary distributions of subgraphs, *Phys. Rev. E* **82**, 066118.

Ma, J.; van den Driessche, P. and Willeboordse, F.H. (2012), Effective degree household network disease model, *J. Math. Biol.* Published online 18th January 2012. DOI 10.1007/s00285-011-0502-9.

May, R.M. and Anderson, R.M. (1987), Transmission dynamics of HIV infections, *Nature* **326**, 137–142.

Miller, J.C. (2009), Spread of infectious disease through clustered populations, *J. R. Soc. Interface* **6** 1121–1134.

Mode, C.J. (1971), *Multitype branching processes. Theory and applications.* Modern Analytic and Computational Methods in Science and Mathematics, **34**. Elsevier, New York.

Mollison, D. (1977), Spatial contact models for ecological and epidemic spread, *J. Roy. Stat. Soc. B* **39**(3) 283–326.

Molloy, M. and Reed, B. (1995), A critical point for random graphs with a given degree sequence. *Rand. Struct. Alg.* **6** 161–179.

Newman, M.E.J., Strogatz, S.H. and Watts, D.J. (2001), Random graphs with arbitrary degree distributions and their applications, *Phys. Rev. E* **64**, 026118.

Newman, M.E.J. (2002a), Assortative mixing in networks, *Phys. Rev. Lett.* **89** 208701.

Newman, M.E.J. (2002b), Spread of epidemic disease on networks, *Phys. Rev. E* **66** 016128.

Newman, M.E.J. (2003), The structure and function of complex networks, *SIAM Review* **45** 167–256.

Newman, M.E.J. (2009), Random graphs with clustering, *Phys. Rev. Lett.* **103** 058701.

Trapman, P. (2007), On analytical approaches to epidemics on networks, *Theor. Pop. Biol.* **71** 160–173.

Watts, S.C. and Strogatz, S.H. (1998), Collective dynamics of 'small-world' networks, *Nature* **393** 440–442.