Mathematical Biology

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

Frank Ball · Tom Britton · David Sirl

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 (1) clustering tends to decrease the spread of disease, (2) the effect of degree correlation is appreciably greater when the disease is close to threshold than when it is well above threshold and (3) disease spread broadly

F. Ball
School of Mathematical Sciences, University of Nottingham,
University Park, Nottingham, NG7 2RD, UK
e-mail: frank.ball@nottingham.ac.uk

T. Britton
Department of Mathematics,
Stockholm University, Stockholm 106 91, Sweden
e-mail: tomb@math.su.se

D. Sirl (✉)
Mathematics Education Centre, Loughborough University,
Loughborough, Leicestershire, LE11 3TU, UK
e-mail: d.sirl@lboro.ac.uk

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

Mathematic subject classification   92D30 · 05C80 · 60J80

1 Introduction

Odo Diekmann has made many highly influential contributions to epidemic theory, for example to the geographical spread of infection (Diekmann 1978), to the definition and computation of the basic reproduction number R_0 (Diekmann et al. 1990) and, as discussed further later, to epidemics on random networks (Diekmann et al. 1998), which is the topic of the present paper.

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 making them more flexible and realistic. For example, the configuration model (Molloy and Reed 1995; 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 (ER) network, and networks with heavy tail degree distributions have been shown to exhibit rather different properties when compared with the ER network; for example, if an epidemic outbreak takes place on the network the basic reproduction number R_0 is much higher (or even infinite) as compared to the same epidemic taking place on an ER network with the same mean degree (Andersson 1999).

Two other properties of real world networks that are not present in ER 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 ER 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 ER 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 ER 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 tuned with considerable flexibility.

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 (e.g. Britton et al. 2011). 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 (i.e. becoming established and infecting a non-neglibile proportion of the population, an event we call a *major outbreak*), and also in terms of the size of a major outbreak. We study the class of SIR (Susceptible $\rightarrow$ Infective $\rightarrow$ Recovered) epidemics (see, 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, Chapter 10), for early analytical contributions in this area.

The paper by Diekmann et al. (1998) is one of the first papers on epidemics on networks and contains many ideas that have subsequently proved to be very fruitful. It is concerned with the analysis of an SIR epidemic on a network in which each individual is connected to k acquaintances. A branching process approximation (cf. Sect. 4.1.1 of this paper) is used to study the initial phase of an epidemic and a deterministic analysis is used to determine rigorously the final outcome of an epidemic in an infinite population. A susceptibility set argument (cf. Sect. 4.2 of this paper) is used to give a heuristic derivation of the final outcome. As indicated by its title, the model in Diekmann et al. (1998) is deterministic and defined as an infinite population limit; the authors point out that it can be realised as a limit of a stochastic model defined on a finite population (see, for example, Andersson 1997, 1998). Extension to the case where individuals may have differing numbers of acquaintances is mentioned in the concluding remarks of Diekmann et al. (1998), the situation studied in the seminal paper of Newman (2002b) on epidemics on networks.

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 basic 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 with controllable 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 a 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; this deeper analysis is missing for several of the above-mentioned models.

The current paper introduces a network model which (1) allows for a wide range of clustering, degree correlation and degree distribution (the extent of which is discussed in Sect. 3.5), and (2) 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 five 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 $-1 \le r \le 1$. 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 $r < 0$).

It is worth mentioning that, as in other cited work, there are many other possible constructions resulting in the same local properties (such as D, c and ρ) but having different global properties (such as the possibility and size of a giant component). Hence, in order to justify the conclusions made from the analysis of an empirical network it is also important to argue why the model construction should resemble the real world network.

The remainder of the paper is organised as follows. A more rigorous definition of the model appears in Sect. 2, where a continuous-time SIR epidemic on the network is also defined. In Sect. 3, we derive expressions for the degree distribution D, clustering coefficient c and degree correlation ρ, as functions of the model parameters (see Propositions 1–3) 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 Sect. 4, we analyse the main characteristics of epidemics defined on the network for suitably large population sizes, by exploiting approximating branching processes. Specifically, in Sect. 4.1, we obtain a threshold parameter R_* which determines whether or not a major outbreak is possible (Theorem 2) and, assuming that the infectious period is constant, derive the probability that a major outbreak occurs (Theorem 3) and, in Sect. 4.2, we derive the relative final size (i.e. the proportion of the population that is ultimately removed) of a major outbreak (Theorem 4). In Sect. 5, we describe how these results on epidemics are modified to incorporate rewiring (see Corollary 1) 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 (Theorem 5). In Sect. 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 Sect. 7.

2 The network model and the epidemic

2.1 The network model

2.1.1 Definition of network model

Consider a network of undirected edges with n nodes (individuals) labelled $1, 2, \ldots, n$. 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' and 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. Let $G_1, G_2, \ldots$ be independent and identically distributed copies of the random variable G and define $H_1, H_2, \ldots$ similarly for H. Let r be a real number satisfying $-1 \le r \le 1$. 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. Let n_Q be a strictly positive integer, it will denote the *number* of equal sized quantiles that we separate a sample of degrees into.

Definition 1 The network model, with random variables $G \sim \{p_k\}$ and $H \sim \{\pi_h\}$, and with parameters r and n_Q specified above, is defined as follows. Group the first

H_1 nodes into local group (household) 1, nodes $H_1 + 1, H_1 + 2, \ldots, H_1 + H_2$ into local group 2 and so on until all individuals belong to a local group (the last local group may 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_i denote 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 $k + 2$. A node having global degree k has k outgoing stubs, and each of these stubs is labelled with an independent copy of X and with the *total* degree of the node from which it emanates. (All the G, H and X random variables are assumed to be independent.) 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 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 the pairing of global stubs 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). An example of the construction of a (very small) network of this type is given in Sect. 2.1.2.

The network is hence made up of local completely connected groups having group-size distribution $\{\pi_h\}$ (as n goes to infinity the effect of the last group possibly having a truncated 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, self-loops and (at most $n_Q + 1$) un-paired stubs. However, if both the global degree and household size distributions G and H have finite variance, the fraction of these will be negligible as $n \to \infty$, so removing them has no effect on the asymptotic properties of the network and epidemics thereon considered in this paper (cf. Durrett 2006, Theorem 3.1.2; Janson 2009). The special case where $r = 0$ (or equivalently $n_Q = 1$) 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.
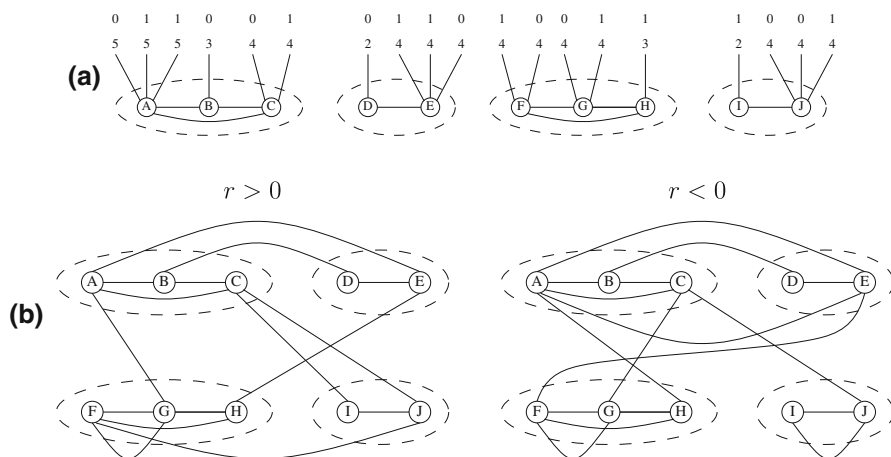
Fig. 1 (**a**) Initial stages of network construction; rows of numbers are the stubs' X-labels and total degrees. (**b**) Constructed networks with $r > 0$ and $r < 0$

2.1.2 Example of network construction

Figure 1(a) shows individuals (here $n = 10$, with individuals labelled A, B, ..., J) in households (here of sizes $\{H_i\} = \{3, 2, 3, 2\}$). The individuals have been assigned global degrees (the number of stubs emanating from them; here 3, 1, 2, 1, 3, 2, 2, 1, 1, 3). Each stub has then been labelled with the appropriate X-label and total degree (the top and bottom rows of numbers, respectively).

The next step in the construction is to pair the 0-stubs uniformly at random; for example this might give pairings of individuals A–E, B–D, C–J and F–G (with a left-over stub emanating from J). Next we deal with the 1-stubs: first construct $n_Q = 2$ quantiles of the 1-stubs (the ordering of the stubs with the same global degree is in general not important, though it does have quite an impact on this very small network):

total degree	2	3	4	4	4	4	4	4	5	5
individual	I	H	C	E	E	F	G	J	A	A

Now, if $r > 0$, we pair the 1-stubs within each quantile uniformly at random or, if $r < 0$, pair the 1-stubs with a stub chosen uniformly at random from the 'opposite' quantile. For example if $r > 0$, the pairings might be, from the first quantile, I–C and H–E (with a left-over stub emanating from E) and from the second quantile, F–J and G–A (with a leftover stub emanating from A). If on the other hand $r < 0$, the pairings might be I–J, H–A, C–G, E–A and E–F. These outcomes result in the networks depicted in Fig. 1(b).

Because this example is on a very small population there are a number of imperfections in the graph (e.g. parallel edges between F and G); there will be (relatively) fewer of these in a larger network.

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 given by Definition 1. Let I be a non-negative random variable and λ a strictly positive real number.

Definition 2 The continuous time SIR epidemic (with parameters I and λ) on the network is defined as follows. Initially there is one randomly selected infectious individual and all other $n - 1$ individuals are susceptible. A susceptible individual who is contacted by an infective immediately becomes infectious and remains so for a period I after which the individual becomes recovered/immune (and plays no further part in the epidemic). 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. Contacts between an infective and an infective or recovered individual have no effect. All Poisson processes describing infectious contacts 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.

Having defined the network model and an epidemic thereon, we now analyse some key properties of the network and then the epidemic.

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 Sect. 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. We hence have the following proposition

Proposition 1 *The total degree distribution (in the network) as $n \to \infty$ is given by*

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

where $\stackrel{d}{=}$ 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.)

Note that Proposition 1 implies that a degree distribution D may be obtained using our model if and only if D is the convolution of two distributions having support in the non-negative integers.

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 $c^{(n)}$ 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 $c^{(n)} = N_\triangle^{(n)}/N_T^{(n)}$, where $N_\triangle^{(n)}$ and $N_T^{(n)}$ are respectively the total numbers of ordered triangles and ordered triples in the network. Note that $c^{(n)}$ is an empirical measure of clustering and may be viewed as a random variable.

Let $N_{\triangle,H}^{(n)}$ be the total number of ordered triangles that are wholly within households. Suppose that σ_G^2 is finite and H has finite third moment. Then by the strong law of large numbers, as $n \to \infty$, the number of individuals per household converges almost surely to μ_H and (cf. Ball et al. 2010, Section 4.2) $n^{-1}N_{\triangle,H}^{(n)}$ and $n^{-1}N_T^{(n)}$ converge almost surely to $\mu_H^{-1}E[H(H-1)(H-2)]$ and $E[H(G+H-1)(G+H-2)]$, respectively. The present network model differs from the that of Ball et al. (2010) only in the way that global stubs are paired. Suppose that the number of quantiles n_Q is fixed and independent of n. Then the number of triangles that are not wholly within households, $N_\triangle^{(n)} - N_{\triangle,H}^{(n)}$ is small and, as outlined in Appendix A, the proof in Ball et al. (2010, Appendix C) may be adapted to show that $n^{-1}(N_\triangle^{(n)} - N_{\triangle,H}^{(n)}) \overset{a.s.}{\to} 0$ as $n \to \infty$, where $\overset{a.s.}{\to}$ denotes almost sure convergence. It then follows that $c^{(n)} \overset{a.s.}{\to} c$ as $n \to \infty$, where $c = c(G, H, r)$ is given by

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

$$= 1 - \frac{E[H\{2G(H-1) + G(G-1)\}]}{E[H(G+H-1)(G+H-2)]}. \tag{3}$$

Note that the expression (3) for c is well-defined provided σ_G^2 and σ_H^2 are both finite and that $c = 1$ if $E[H^3] = \infty$. Further, as indicated in Appendix A, both σ_G^2 and σ_H^2 being finite is sufficient for $c^{(n)} \overset{a.s.}{\to} c$ as $n \to \infty$. We thus have the following theorem.

Theorem 1 *Suppose that $E[H^2]$ and $E[G^2]$ are both finite and that n_Q is fixed and independent of n. Then*

$$c^{(n)} \overset{a.s.}{\to} c \quad as \quad n \to \infty.$$

We call c the clustering coefficient of our model and thus have the following proposition.

Proposition 2 *The clustering coefficient $c = c(G, H, r)$ of the network model is given by* (3), *where G and H are the global degree and household size distributions of the network.*

Note that c is independent of r and n_Q, essentially because asymptotically the global pairings do not yield triangles, and in particular the clustering coefficient of our model is identical to that of the model in Ball et al. (2010). Note also that $c = 0$ if $P(H \geq 3) = 0$ and that, for fixed G (with $\sigma_G^2 < \infty$), c can be made arbitrarily close to its maximum possible value of 1 by choosing H 'sufficiently large' (i.e. by making $E[H^3]/E[H^2]$ as large as necessary). For computational purposes, (2) and (3) admit simple representations in terms of factorial moments of G and H; see Eq. (16) of Ball et al. (2010).

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 Appendix C. 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

$$p_G = \frac{\mu_G}{\mu_G + \mu_{\hat{H}} - 1}. \tag{4}$$

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, $P(\hat{H} = h) \propto \binom{h}{2}\pi_h$ $(h = 2, 3, \ldots)$, so

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

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 $\tilde{p}_g = g p_g / \mu_G$ $(g = 1, 2, \ldots)$. For $i = 1, 2, \ldots, n_Q$ and $d = 1, 2, \ldots$, let $p_{\tilde{Q}|\tilde{D}}(i|d) = P(\tilde{Q} = i | \tilde{D} = d)$ and $p_{\tilde{D}|\tilde{Q}}(d|i) = P(\tilde{D} =$

$d|\tilde{Q} = i)$. (These conditional probabilities are derived easily from the probability mass function of $\tilde{D}$, noting that if $\tilde{u}_0 = 0$ and $\tilde{u}_d = P(\tilde{D} \leq d)$ $(d = 1, 2, \ldots)$ then $P(\tilde{D} = d, \tilde{Q} = i) = \max\{\min(\tilde{u}_d, \frac{i}{n_Q}) - \max(\tilde{u}_{d-1}, \frac{i-1}{n_Q}), 0\}$ $(d = 1, 2, \ldots; i = 1, 2, \ldots, n_Q)$.) Define the function $g_{\tilde{D}, n_Q}(r)$ by

$$g_{\tilde{D}, n_Q}(r) = \begin{cases} r\left(\frac{1}{n_Q} \sum_{i=1}^{n_Q} (\mu_{\tilde{D}}^{(i)})^2 - \mu_{\tilde{D}}^2\right) & \text{if } r \geq 0, \\ |r|\left(\frac{1}{n_Q} \sum_{i=1}^{n_Q} \mu_{\tilde{D}}^{(i)} \mu_{\tilde{D}}^{(n_Q+1-i)} - \mu_{\tilde{D}}^2\right) & \text{if } r < 0, \end{cases} \tag{5}$$

where

$$\mu_{\tilde{D}}^{(i)} = \sum_{d=1}^{\infty} d \, p_{\tilde{D}|\tilde{Q}}(d|i) \quad (i = 1, 2, \ldots, n_Q). \tag{6}$$

The following proposition is proved in Appendix C.

Proposition 3 *The degree correlation $\rho = \rho(G, H, r)$ of the network model is given by*

$$\rho = \frac{(1 - p_G)\sigma_{\hat{H}}^2 + p_G g_{\tilde{D}, n_Q}(r) + p_G(1 - p_G)\left(\mu_{\hat{H}} - \mu_{\tilde{H}} - \frac{\sigma_G^2}{\mu_G}\right)^2}{(1 - p_G)\left(\sigma_{\hat{H}}^2 + \sigma_G^2\right) + p_G\left(\sigma_{\hat{H}}^2 + \sigma_G^2\right) + p_G(1 - p_G)\left(\mu_{\hat{H}} - \mu_{\tilde{H}} - \frac{\sigma_G^2}{\mu_G}\right)^2}. \tag{7}$$

The proposition together with (5) imply that ρ is an increasing function of r, if all other parameters are held fixed, a result that is also expected from the construction.

3.4 Rewiring

Note that for given 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 (and also of n_Q). 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 $h = 2, 3, \ldots$, the red stubs with label h are now paired uniformly at random, which, together with the green edges and unbroken red edges creates a new network. The rewiring process may yield self-loops and multiple edges, which are not deleted when determining the properties of the network and epidemics thereon. However, as with the original network, if both the global degree and household size distributions, G and H, have finite variance, then the fraction of such imperfections is negligible as $n \to \infty$ and removing them has no effect on the asymptotic properties of the network and epidemics thereon studied in this paper.

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. Having added one parameter p_{RW} we now 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 (previously denoted $c(G, H, r)$). In the limit as $n \to \infty$, the number of triangles that are not wholly within unbroken households, per individual, tends to zero. This implies the following proposition.

Proposition 4 *For the network model with rewiring fraction p_{RW}, as described above, the clustering coefficient is given by*

$$c(G, H, r, p_{RW}) = (1 - p_{RW})c(G, H, r, 0). \tag{8}$$

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 $c(G, H, r, 0)$, i.e. that of the model without rewiring.

As for the unrewired model, $c(G, H, r, p_{RW})$ may be defined formally as the almost sure limit of its associated empirical clustering coefficient $c^{(n)}(G, H, r, p_{RW})$. We show in Appendix B, that under the slightly stronger condition that $E[H^3] < \infty$, Theorem 1 holds with $c^{(n)}$ and c replaced by $c^{(n)}(G, H, r, p_{RW})$ and $c(G, H, r, p_{RW})$, respectively.

3.5 Tuning with Poisson degree distributions

The formulae given in Sects. 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 $\text{Poi}^+(\mu)$, so $\pi_h = (1 - e^{-\mu})^{-1}\mu^h e^{-\mu}/h!$ $(h = 1, 2, \ldots)$. Here we interpret $\text{Poi}(0)$ to be the distribution identically equal to 0 and $\text{Poi}^+(0)$ to be $\lim_{\mu \downarrow 0} \text{Poi}^+(\mu)$, the distribution identically equal to 1. 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 (7), the formulae for the clustering and degree correlation are given by:
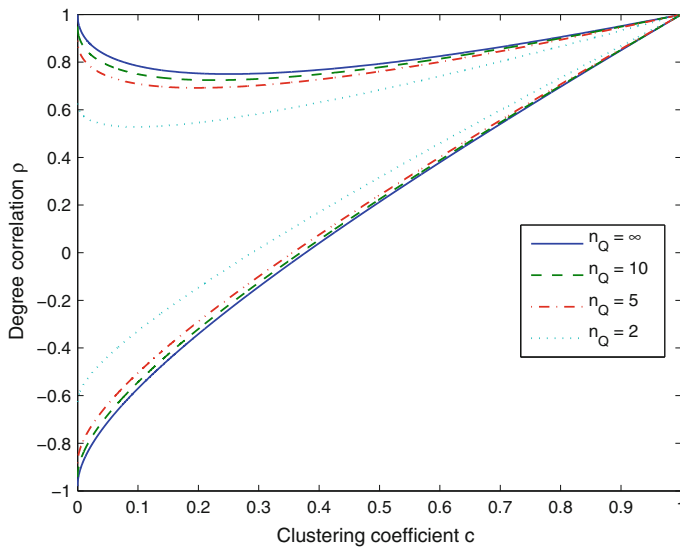
Fig. 2 Plot showing bounds on possible values of (c, ρ) when $D \sim \text{Poi}(10)$

$$c = \left(\frac{\mu}{\gamma}\right)^2 \quad \text{and} \quad \rho = \frac{1}{\gamma^2}\left[\mu^2 + (\gamma - \mu)g_{\gamma,n_Q}(r)\right], \tag{9}$$

where $g_{\gamma,n_Q}(r)$ is given by (5) 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 Poisson model studied in Ball et al. (2010, Sects. 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 (9). These lower and upper values for ρ are shown in Fig. 2 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 Fig. 2 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 Poi(γ). 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 total degree distributions 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 $[0, 1)$. See, for example, Feller (1971, page 176) for the definition of infinitely divisible and Warde and Katti (1971) for discussion of infinite divisibility of discrete distributions having support in the positive integers.

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 Sect. 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 $n = 1, 2, \ldots$, 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 occurs with non-zero probability is determined by whether or not the branching process is supercritical.

The connection between a major outbreak and survival of the approximating branching process can be made precise by defining a major outbreak to be one that infects at least $\log n$ individuals and noting that, as $n \to \infty$, the probability of a major outbreak tends to the survival probability of the approximating branching process (cf. the discussion following Corollary 6.1 in Ball et al. 2009). Moreover, as discussed in Sect. 4.2, we conjecture that if a major outbreak occurs then, as $n \to \infty$, its relative final size converges in probability to the survival probability of another supercritical branching process. This implies that a major outbreak is one that infects a strictly positive fraction of the population in the limit as $n \to \infty$ and we use this practical interpretation in the statements of Theorems 2–4 below.

We now determine the means and probability generating functions (PGFs) of the offspring distributions of the forward 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 $i = 1, 2, \ldots, n_Q$, let $\tilde{C}_i = (\tilde{C}_{i1}, \tilde{C}_{i2}, \ldots, \tilde{C}_{in_Q})$ be a random vector 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 Theorem 7.1 in Chapter 1 of Mode 1971), 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 $d = 1, 2, \ldots$ and $h = 1, 2, \ldots, d$, 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

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

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

$$\tilde{m}_{ij} = \sum_{d=1}^{\infty} p_{\tilde{D}|\tilde{Q}}(d|i) \sum_{h=1}^{d} \tilde{\pi}_h^{(d)} \mathrm{E}[\tilde{C}_{ij}^{(h,d)}], \tag{10}$$

where $\tilde{C}_i^{(h,d)} = (\tilde{C}_{i1}^{(h,d)}, \tilde{C}_{i2}^{(h,d)}, \ldots, \tilde{C}_{in_Q}^{(h,d)})$ is defined analogously to $\tilde{C}_i$, 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 independent of the X-label of the edge along which the type-i primary individual was infected.)

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

$$\tilde{C}_i^{(h,d)} = \tilde{C}_i^{(h,d)}(0) + \sum_{k=1}^{h-1} \chi_k \tilde{C}_i^{(h,d)}(k), \tag{11}$$

where, for $k = 0, 1, \ldots, h-1$, $\tilde{C}_i^{(h,d)}(k) = (\tilde{C}_{i1}^{(h,d)}(k), \tilde{C}_{i2}^{(h,d)}(k), \ldots, \tilde{C}_{in_Q}^{(h,d)}(k))$, 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 single-household epidemic $\mathcal{E}^{(h)}$, not including the initial case, and let $\mu^{(h)}(\lambda) = \mathrm{E}[T^{(h)}]$. Then, see Eqs. (2.25) and (2.26) of Ball (1986),

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

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

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

Note that χ_k and $\tilde{C}_i^{(h,d)}(k)$ are independent, because whether or not an individual is infected by $\mathcal{E}^{(h)}$ is independent of its infectious period, so taking expectations of (11) and noting that $\tilde{C}_i^{(h,d)}(1), \tilde{C}_i^{(h,d)}(2), \ldots, \tilde{C}_i^{(h,d)}(h-1)$ are identically distributed yields

$$\mathrm{E}[\tilde{C}_{ij}^{(h,d)}] = \mathrm{E}[\tilde{C}_{ij}^{(h,d)}(0)] + \mu^{(h)}(\lambda)\mathrm{E}[\tilde{C}_{ij}^{(h,d)}(1)]. \tag{12}$$

To determine $\mathrm{E}[\tilde{C}_{ij}^{(h,d)}(k)](k = 0, 1)$, for $i, j = 1, 2, \ldots, n_Q$ 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 $i \neq j$. This follows immediately because when $X = 1$ an i-stub is joined to another i-stub if $r > 0$, whereas if $r < 0$ it is joined to an $(n_Q + 1 - i)$-stub. Further, for $d = 1, 2, \ldots, j = 1, 2, \ldots, n_Q$ and $l = 0, 1$, let $\tilde{p}_{d,j}^{(l)}(r)$ 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 $\tilde{p}_{d,j}^{(0)}(r) = 1/n_Q$ and

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

Consider the individual labelled 0, i.e. the primary case, in $\mathcal{E}^{(h)}$. 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

$$\mathrm{E}[\tilde{C}_{ij}^{(h,d)}(0)] = (d - h)p_I[(1 - |r|)n_Q^{-1} + |r|\tilde{p}_{d,j}^{(1)}(r)], \tag{13}$$

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 $\mathcal{E}^{(h)}$ and suppose that it becomes infected. The global degree of individual 1 is distributed according to G. Thus, for $g = 1, 2, \ldots$, 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

$$\mathrm{E}[\tilde{C}_{ij}^{(h,d)}(1)] = \sum_{g=1}^{\infty} p_g g p_I [(1 - |r|) n_Q^{-1} + |r| \tilde{p}_{g+h-1,j}^{(1)}(r)]. \tag{14}$$

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

Combining (10), (12), (13) and (14) gives

$$\tilde{m}_{ij} = p_I \sum_{d=1}^{\infty} p_{\tilde{D}|\tilde{Q}}(d|i) \sum_{h=1}^{d} \tilde{\pi}_h^{(d)} \left\{ (d - h) p_I \left[(1 - |r|) n_Q^{-1} + |r| \tilde{p}_{d,j}^{(1)}(r) \right] \right.$$
$$\left. + \mu^{(h)}(\lambda) \left[(1 - |r|) \mu_G n_Q^{-1} + |r| \sum_{g=1}^{\infty} p_g g \tilde{p}_{g+h-1,j}^{(1)}(r) \right] \right\}. \tag{15}$$

To summarise, Eq. (15) defines the elements of the mean matrix $\tilde{M} = [\tilde{m}_{ij}]$ of the branching process $\tilde{\mathcal{B}}_F$. We have the following theorem stating when a major epidemic outbreak is possible.

Theorem 2 *For the epidemic model taking place on the network model with community size $n \to \infty$, a major epidemic outbreak (infecting a strictly positive community-fraction) is possible if and only if the dominant eigenvalue of $\tilde{M} = [\tilde{m}_{ij}]$, denoted by R_*, satisfies $R_* > 1$.*

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 (11) 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 Sect. 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 (see final paragraph of Sect. 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 $i = 1, 2, \ldots, n_Q$ and $\boldsymbol{s} = (s_1, s_2, \ldots, s_{n_Q})$ with $0 \leq s_i \leq 1$ $(i = 1, 2, \ldots, n_Q)$, 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, $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_{n_Q})$ 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 (10), yields

$$
f_{\tilde{\boldsymbol{C}}_i}(\boldsymbol{s}) = \sum_{d=1}^{\infty} p_{\tilde{D}|\tilde{Q}}(d|i) \sum_{h=1}^{d} \tilde{\pi}_h^{(d)} f_{\tilde{\boldsymbol{C}}_i^{(h,d)}}(\boldsymbol{s}). \tag{16}
$$

The decomposition (11) may be expressed as

$$
\tilde{\boldsymbol{C}}_i^{(h,d)} = \tilde{\boldsymbol{C}}_i^{(h,d)}(0) + \sum_{k=1}^{T^{(h)}} \tilde{\boldsymbol{C}}_i^{(h,d)}(k), \tag{17}
$$

where now $\tilde{\boldsymbol{C}}_i^{(h,d)}(1), \tilde{\boldsymbol{C}}_i^{(h,d)}(2), \ldots, \tilde{\boldsymbol{C}}_i^{(h,d)}(T^{(h)})$ give the offspring vectors for the $T^{(h)}$ secondary cases in the single-household epidemic $\mathcal{E}^{(h)}$. Further, since the infectious period is constant, conditional upon $T^{(h)}$, the random vectors $\tilde{\boldsymbol{C}}_i^{(h,d)}(1), \tilde{\boldsymbol{C}}_i^{(h,d)}(2), \ldots, \tilde{\boldsymbol{C}}_i^{(h,d)}(T^{(h)})$ are independent and identically distributed copies of a random vector whose distribution is independent of $T^{(h)}$. Hence, (17) implies that

$$
f_{\tilde{\boldsymbol{C}}_i^{(h,d)}}(\boldsymbol{s}) = f_{\tilde{\boldsymbol{C}}_i^{(h,d)}(0)}(\boldsymbol{s}) f_{T^{(h)}}\left(f_{\tilde{\boldsymbol{C}}_i^{(h,d)}(1)}(\boldsymbol{s})\right), \tag{18}
$$

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

$$
f_{T^{(h)}}(s) = s^{h-1} \sum_{k=0}^{h-1} \binom{h-1}{k} \alpha_k(s)(1 - p_I)^{k(h-k)} \quad (h = 1, 2, \ldots), \tag{19}
$$

where $\alpha_0(s)(= 1), \alpha_1(s), \ldots$ are defined recursively by

$$\sum_{l=0}^{k} \binom{k}{l}(1 - p_I)^{l(k-l)}\alpha_l(s) = s^{-k} \quad (k = 0, 1, \ldots). \tag{20}$$

To complete the derivation of $f_{\tilde{C}_i}(s)$, we obtain expressions for $f_{\tilde{C}_i^{(h,d)}(0)}(s)$ and $f_{\tilde{C}_i^{(h,d)}(1)}(s)$. Consider a typical type-i primary infective, i^* say, and let j^* be a susceptible global neighbour of i^*. Suppose that i^* has total degree d and let $\chi_d = (\chi_{d1}, \chi_{d2}, \ldots, \chi_{dn_Q})$, where $\chi_{dk} = 1$ 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 $\chi_{dk} = 0$ otherwise. (Note that if i^* does not infect j^* then every element of χ_d is zero, and if i^* does infect j^* then precisely one element of χ_d is one and all other elements of χ_d are zero.) For $d = 1, 2, \ldots$ and $s \in [0, 1]^{n_Q}$, define the PGF of χ_d, viz.

$$g_d(s) = \mathrm{E}\left[\prod_{j=1}^{n_Q} s_j^{\chi_{dj}}\right] = 1 - p_I + p_I \sum_{j=1}^{n_Q}\left[(1 - |r|)\frac{s_j}{n_Q} + |r|\tilde{p}_{d,j}^{(1)}(r)s_j\right]. \tag{21}$$

Then using similar arguments to the derivations of (13) and (14) yields respectively

$$f_{\tilde{C}_i^{(h,d)}(0)}(s) = (g_d(s))^{d-h} \tag{22}$$

and

$$f_{\tilde{C}_i^{(h,d)}(1)}(s) = \sum_{g=0}^{\infty} p_g \left(g_{g+h-1}(s)\right)^g. \tag{23}$$

Combining (16), (18), (22) and (23) gives the PGF of the offspring random variable $\tilde{C}_i$ 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 $C = (C_1, C_2, \ldots, C_{n_Q})$ denote the offspring random variable for the ancestor in $\mathcal{B}_F$. Then, conditioning on i^*'s global degree and household size,

$$f_C(s) = \sum_{g=0}^{\infty}\sum_{h=1}^{\infty} p_g \tilde{\pi}_h f_{C^{(h,g+h-1)}}(s), \tag{24}$$

where, for $h = 1, 2, \ldots$ and $d = h+1, h+2, \ldots$, $C^{(h,d)}$ 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 (17), $C^{(h,d)}$ admits the decomposition

$$C^{(h,d)} = C^{(h,d)}(0) + \sum_{k=1}^{T^{(h)}} C^{(h,d)}(k), \tag{25}$$

whence, as at (18),

$$f_{\boldsymbol{C}^{(h,d)}}(\boldsymbol{s}) = f_{\boldsymbol{C}^{(h,d)}(0)}(\boldsymbol{s}) f_{T^{(h)}}\left(f_{\boldsymbol{C}^{(h,d)}(1)}(\boldsymbol{s})\right). \tag{26}$$

Now $\boldsymbol{C}^{(h,d)}(1) \stackrel{d}{=} \tilde{\boldsymbol{C}}^{(h,d)}(1)$, so $f_{\boldsymbol{C}^{(h,d)}(1)}(\boldsymbol{s})$ is given by the right hand side of (23). 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,

$$f_{\boldsymbol{C}^{(h,d)}(0)}(\boldsymbol{s}) = (g_d(\boldsymbol{s}))^{d-h+1}. \tag{27}$$

The offspring PGF $f_{\boldsymbol{C}}$ of the ancestor in $\mathcal{B}_F$ now follows using (24), (26), (23) and (27).

We now determine the probability of a major outbreak. Suppose that $R_* > 1$. For $i = 1, 2, \ldots, n_Q$, 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 $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_{n_Q})$. Then, see, for example, Mode (1971, Section 1.7.1), $\boldsymbol{\sigma}$ is the unique solution in $[0, 1)^{n_Q}$ of the equations

$$f_{\tilde{\boldsymbol{C}}_i}(\boldsymbol{\sigma}) = \sigma_i \quad (i = 1, 2, \ldots, n_Q). \tag{28}$$

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 $p_{\text{maj}} = 1 - f_{\boldsymbol{C}}(\boldsymbol{\sigma})$. We have just proved the following theorem.

Theorem 3 *Consider the epidemic taking place on the network starting with one randomly selected infective and the rest being susceptible. Assume that $R_* > 1$ and the infectious period is constant, viz. $I \equiv \iota$. Then, as the community size $n \to \infty$, the probability p_{maj} that a strictly positive fraction becomes infected is given by*

$$p_{\text{maj}} = 1 - f_{\boldsymbol{C}}(\boldsymbol{\sigma}), \tag{29}$$

where $\boldsymbol{\sigma}$ is the root of (28) and $f_{\boldsymbol{C}}$ is the joint PGF of $\boldsymbol{C}$ defined above. If $R_* \leq 1$ then $p_{\text{maj}} = 0$.

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; Ball and Neal 2002), which we now define. Label the n nodes (individuals) $1, 2, \ldots, n$. For $i = 1, 2, \ldots, n$, 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 $1, 2, \ldots, n$, in which for any ordered 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

$i = 1, 2, \ldots, n$, 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 households of all such global neighbours of i^*'s household form the first generation of the (backward) approximating branching process, with the household of a typical 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. Figure 3 depicts this process. 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 Sect. 6 (see Fig. 4 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 $\boldsymbol{B} = (B_1, B_2, \ldots, B_{n_Q})$ denote the offspring random variable for the ancestor in $\mathcal{B}_B$ and, for $i = 1, 2, \ldots, n_Q$, let $\tilde{\boldsymbol{B}}_i = (\tilde{B}_{i1}, \tilde{B}_{i2}, \ldots, \tilde{B}_{in_Q})$ 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 (an individual that has made global contact with a member of the susceptibility set of interest) and assume it has type i. Then arguing as at (16) yields

$$f_{\tilde{\boldsymbol{B}}_i}(\boldsymbol{s}) = \sum_{d=1}^{\infty} p_{\tilde{D}|\tilde{Q}}(d|i) \sum_{h=1}^{d} \tilde{\pi}_h^{(d)} f_{\tilde{\boldsymbol{B}}_i^{(h,d)}}(\boldsymbol{s}), \tag{30}$$
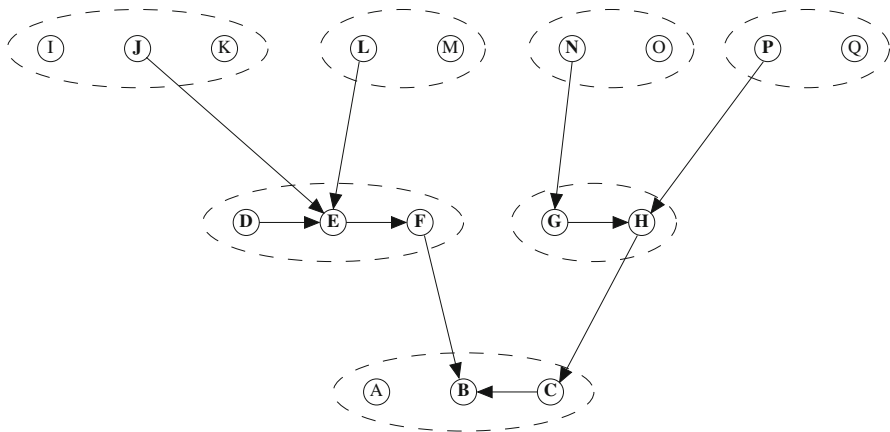
Fig. 3 Construction of susceptibility set. In this diagram we consider (part of) the susceptibility set of individual B; the first 2 generations in terms of the branching process approximation $\mathcal{B}_B$. To avoid clutter, we show only the contacts (edges in the directed random graph) that are strictly relevant to the susceptibility set. Individuals in B's susceptibility set have their labels in *bold* type. As we are considering B's susceptibility set, B's household corresponds to the ancestor in the branching process $\mathcal{B}_B$. Then B's local susceptibility set consists of B and C, and the first generation of $\mathcal{B}_B$ consists of the households that contain individuals who make (global) contact with B or C (i.e. B's local susceptibility set); these are the households of individuals F and H. (In the language of the second paragraph of Sect. 4.2, B plays the role of i^*, C of j^* and F and H of k^*.) Note that any individual who makes infectious contact with A does not join the susceptibility set by virtue of this contact (and its household is therefore not an individual in the first generation of the branching process), since A is not in B's local susceptibility set. To construct the next generation of $\mathcal{B}_B$ we repeat this process, focusing on global contacts with the local susceptibility sets of F and H in turn. These global contactors are J and L (they contact F's local susceptibility set {D, E, F}) and N and P (they contact H's local susceptibility set {H, G}). Construction of the susceptibility set and branching process approximation now continues by considering global contacts with the local susceptibility sets of J, L, N and P

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 $\tilde{\boldsymbol{B}}_i^{(h,d)}(l) = (\tilde{B}_{i1}(l), \tilde{B}_{i2}(l), \ldots, \tilde{B}_{in_Q}(l))$, 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 $d = 1, 2, \ldots$ and $h = 1, 2, \ldots, d$,

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

and $g_d(\boldsymbol{s})$ is defined by (21).

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

$$f_{\boldsymbol{B}}(\boldsymbol{s}) = \sum_{g=0}^{\infty} \sum_{h=1}^{\infty} p_g \tilde{\pi}_h f_{\boldsymbol{B}^{(h,g+h-1)}(0)}(\boldsymbol{s}) f_{M^{(h)}} \left(f_{\boldsymbol{B}^{(h,g+h-1)}(1)}(\boldsymbol{s}) \right), \qquad (31)$$

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

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

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 $h = 1, 2, \ldots,$

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

where

$$\sum_{l=0}^{k} \binom{k}{l} \phi_I((l+1)\lambda)^{k-l} P(M^{(l+1)} = l) = 1 \qquad (k = 0, 1, \ldots).$$

It is readily shown that $E[M^{(h)}] = E[T^{(h)}]$ $(h = 1, 2, \ldots)$, 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 $i = 1, 2, \ldots, n_Q$, let ξ_i be the probability that the branching process $\mathcal{B}_B$ goes extinct given that there is one ancestor whose type is i, and let $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_{n_Q})$. Then, if $R_* > 1$, $\boldsymbol{\xi}$ is the unique solution in $[0, 1)^{n_Q}$ of the equations

$$f_{\tilde{B}_i}(\boldsymbol{\xi}) = \xi_i \quad (i = 1, 2, \ldots, n_Q) \qquad (32)$$

and, for n suitably large, the relative final size of a major outbreak, z say, is given approximately by $z = 1 - f_{\boldsymbol{B}}(\boldsymbol{\xi})$. Thus, assuming that our conjecture in the third paragraph of this subsection is true, we have the following result which, for convenience of reference, we state as a theorem.

Theorem 4 *Consider the epidemic taking place on the network starting with one randomly selected infective and the rest being susceptible, and assume that $R_* > 1$. Then, as the community size $n \to \infty$, the relative final size of a major outbreak is given by*

$$z = 1 - f_{\boldsymbol{B}}(\boldsymbol{\xi}), \qquad (33)$$

where $\boldsymbol{\xi}$ is the root of (32) and $f_{\boldsymbol{B}}$ is the joint PGF of $\boldsymbol{B}$ defined above.

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 (19) and (20), 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 Sect. 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 $h = 2, 3, \ldots$, 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 Sects. 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.

Consider first households of size 1; note that $\hat{T}^{(1)} \stackrel{d}{=} T^{(1)} \stackrel{d}{=} \hat{M}^{(1)} \stackrel{d}{=} M^{(1)} \equiv 0$. Next, each individual in households of size 2 has 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 $\hat{T}^{(2)} \stackrel{d}{=} T^{(2)}$ and $\hat{M}^{(2)} \stackrel{d}{=} M^{(2)}$. Now fix $h \geq 3$ 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

$$\hat{\mu}^{(h)}(\lambda) = \mathrm{E}[\hat{T}^{(h)}] = \begin{cases} (h - 1)p_I[1 - (h - 2)p_I]^{-1} & \text{if } p_I < \frac{1}{h-2}, \\ \infty & \text{if } p_I \geq \frac{1}{h-2}. \end{cases} \quad (34)$$

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 $h \geq 3$,

$$f_{\hat{T}^{(h)}}(s) = \left(1 - p_I + p_I \tilde{f}^{(h)}(s)\right)^{h-1} \quad (0 \leq s \leq 1), \tag{35}$$

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

$$\tilde{f}^{(h)}(s) = s \left(1 - p_I + p_I \tilde{f}^{(h)}(s)\right)^{h-2}, \tag{36}$$

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 $\hat{\mathcal{B}}_F^{(h)}$ and further that (35) and (36) also hold for $h = 2$.

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, for $h \geq 3$, 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 $p_I = 1 - \phi_I(\lambda)$, so $f_{\hat{M}^{(h)}}(s)$ is given by the right hand side of (35).

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 $1 - p_{RW}$. This shows that we may also apply results from Sect. 4 to prove limiting properties of epidemics defined on the rewired network.

Corollary 1 *Consider the epidemic model on the network with rewiring probability p_{RW}, starting with one randomly selected infective and the rest being susceptible. Then, the threshold parameter R_* and, as $n \to \infty$, the probability of a major outbreak p_{maj} (for the special case that $I \equiv \iota$) and the relative final size z when a major outbreak occurs are given by the corresponding expressions in Sects. 4.1 and 4.2, by using the following substitutions: for R_* (see Eq. (15) and Theorem 2), replace $\mu^{(h)}(\lambda)$ by $\mathrm{E}[T^{(h)}(p_{RW})] = (1 - p_{RW})\mu^{(h)}(\lambda) + p_{RW}\hat{\mu}^{(h)}(\lambda)$; for p_{maj} (see Sect. 4.1.3 and Theorem 3), replace $f_{T^{(h)}}(s)$ by $f_{T^{(h)}(p_{RW})}(s) = (1 - p_{RW})f_{T^{(h)}}(s) + p_{RW}f_{\hat{T}^{(h)}}(s)$; and for z (see Sect. 4.2 and Theorem 4), replace $f_{M^{(h)}}(s)$ by $f_{M^{(h)}(p_{RW})}(s) = (1 - p_{RW})f_{M^{(h)}}(s) + p_{RW}f_{\hat{M}^{(h)}}(s)$.*

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 smaller 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 larger for the fully rewired network than for the corresponding model without rewiring. Here we extend these results by employing a coupling argument, similar to that in, for example, Mollison (1977) and Ball (1983), to prove that for our model R_*, p_{maj} (now with a general infectious period distribution) 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 $h = 1, 2, \ldots$, recall that $\mathcal{E}^{(h)}$ denotes the single size-h household epidemic introduced in Sect. 4.1.2 and $T^{(h)}$ 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 $P(H \geq 3) > 0$ and $\lambda \mu_I > 0$ then R_* is strictly increasing in p_{RW}, as is p_{maj} when $R_* > 1$.

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 (i^* and the previous $k-1$ first-generation members of $\mathcal{S}^{(h)}$), 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 $\lambda > 0$, $p_F(k) \geq p_F(0)$ ($k = 1, 2, \ldots$), 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

$$\eta(\lambda) + \eta(k\lambda) \leq \eta((k+1)\lambda), \tag{37}$$

which implies that $p_F(k) \geq p_F(0)$ ($k = 1, 2, \ldots$). 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 (37) is strict and $p_F(k) > p_F(0)$ ($k = 1, 2, \ldots$).

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 $\hat{Z}_{j^*}$ 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 $\mathrm{P}(H \geq 3) > 0$, $\lambda\mu_I > 0$ and $R_* > 1$.

Collecting the above results together, we have proved the following theorem concerning the effect of rewiring on asymptotic properties of epidemics defined on the

network. (Note that these results for p_{maj} are *not* restricted to the case of a constant infectious period.)

Theorem 5 *Consider the epidemic model defined on the network with rewiring probability p_{RW}, starting with one randomly selected infective and the rest being susceptible. Then, if all other parameters are held fixed, the threshold parameter, R_*, and the asymptotic probability and relative final size of a major outbreak, p_{maj} and z, are each increasing functions of the rewiring probability p_{RW}. Further, if $P(H \geq 3) > 0$ and $\lambda \mu_I > 0$, then R_* is strictly increasing in p_{RW} and, provided $R_* > 1$, so are p_{maj} and z.*

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 $p_{\text{maj}} = z$, and rather than dealing explicitly with I and the contact rate λ we refer to the marginal infection probability $p_I = 1 - \exp(-\lambda \iota)$. We use the notation Poi and Poi$^+$ for global degree and household size distributions, as in Sect. 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 4 shows this behaviour 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 the endpoints of 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 take off into a major outbreak (suitable cutoffs between minor and major outbreaks being determined by inspecting histograms of epidemic final size), and 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 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 result in a major outbreak. As 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 1,000 nodes the asymptotic values of p_{maj} seem to be very good approximations to the empirically calculated major outbreak probabilities across the various values of r and c. 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 bias, which reduces when n increases to 10,000 and is more pronounced for more extreme values of r. One possible explanation for this is
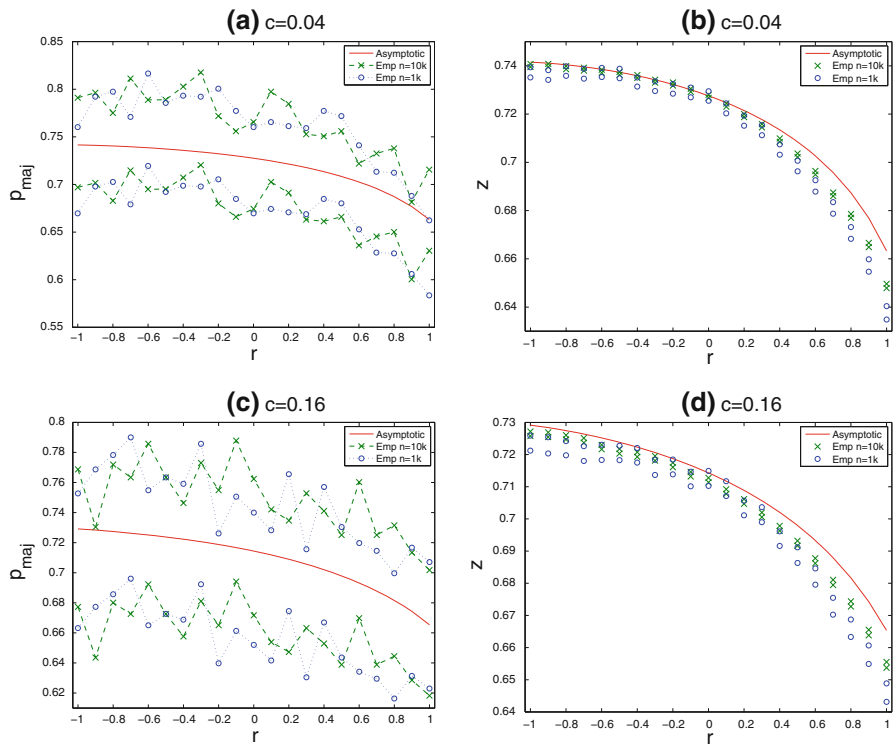
Fig. 4 Apparent convergence of empirically calculated quantities of interest on finite networks to asymptotic values. These plots compare empirical (simulation-based) estimates ($n < \infty$) and asymptotic values ($n \to \infty$) of p_{maj} and z, as a function of r, for our model with [plots (**a**) and (**b**)] degree distributions $H \sim \text{Poi}^+(2)$ and $G \sim \text{Poi}(8)$ ($c = 0.04$) and [plots (**c**) and (**d**)] $H \sim \text{Poi}^+(4)$ 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 are not the same on the vertical axes of these plots

that when $|r|$ is close to 1, there are more imperfections (self-loops, parallel edges, household self-loops, etc.) and global triangles in the random network (for example, there are roughly n_Q times as many self-loops when $r = 1$ than when $r = 0$), and consequently the branching process provides a coarser upper bound to the epidemic process. Nevertheless, the z plots in Fig. 4 lend considerable credence to our conjecture in Sect. 4.2 that the expected relative final size of a major outbreak converges to the survival probability of $\mathcal{B}_B$ as $n \to \infty$. In addition, preliminary investigations using heavy-tailed (power law) global degree distributions suggest that, as for the model with $r = 0$ (see Fig. 3 of Ball et al. 2009), the asymptotic values of p_{maj} and z provide good approximations to these quantities in modestly sized finite networks.

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 $p_{\text{maj}}(= z)$ considered as a function of r (and p_I). Recall (from the end of Sect. 3.3) that the
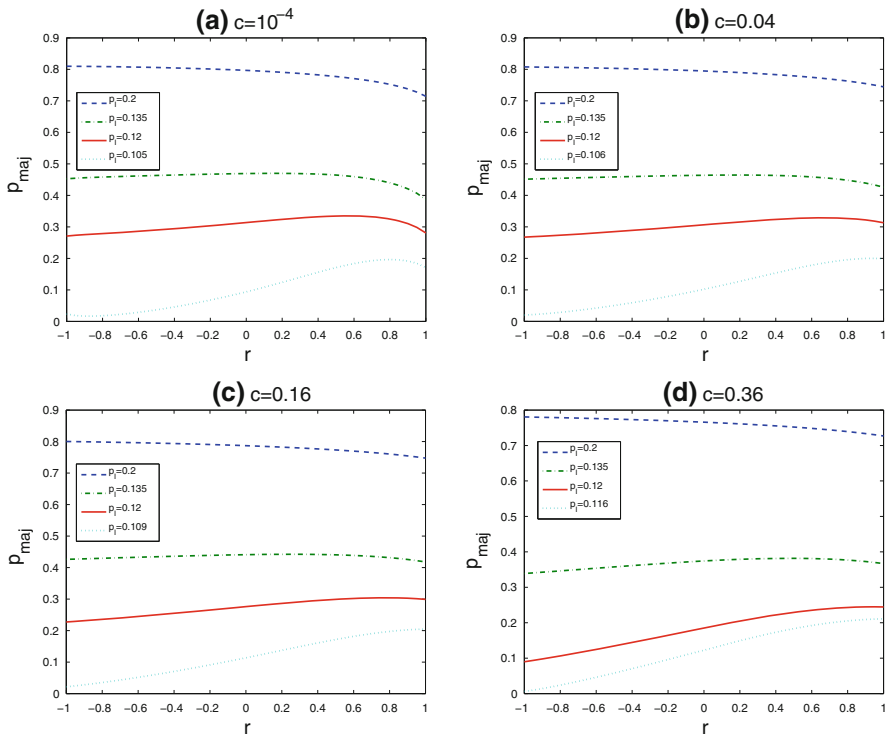
Fig. 5 Investigation of the dependence of p_{maj} on r, p_I and c. Plot of p_{maj} versus r for varying values of p_I. Here $G \sim \text{Poi}(10-\mu)$ and $H \sim \text{Poi}^+(\mu)$, with μ taking the values, in order, 0.1, 2, 4, 6; corresponding to clustering coefficients 10^{-4}, 0.04, 0.16, 0.36. 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

degree correlation ρ is monotonically increasing in r, so the qualitative behaviour is also valid as a function of ρ. Figure 5 illustrates this behaviour in the case where $G \sim \text{Poi}(10 - \mu)$, $H \sim \text{Poi}^+(\mu)$, so $D \sim \text{Poi}(10)$, 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 $p_I = 0.105$ line in plot (a) of Fig. 5 is explored in further detail in Fig. 6. Note, however, that the model parameters that 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 $p_{maj}(= z)$ for different values of r (i.e. degree correlation) are observed near criticality in all of the plots in Fig. 5; even though the non-monotonicity is only observed in plot (a).
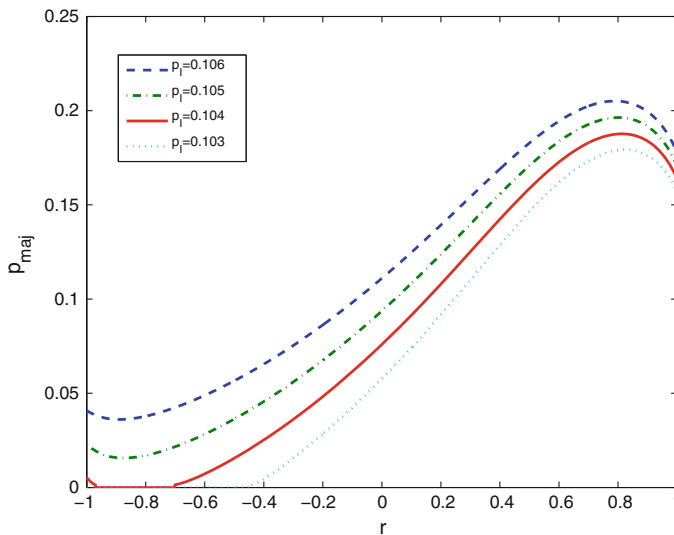
Fig. 6 Detail of the dependence of p_{maj} on r and p_I near criticality, for small c. These plots are of p_{maj} versus r for near-critical values of p_I, when $G \sim \text{Poi}(9.9)$, $H \sim \text{Poi}^+(0.1)$ and $n_Q = 10$

Finally, Fig. 7 illustrates the effect on $p_{maj}(= z)$ 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 Fig. 2) by varying μ and using (9). 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 7 shows how $p_{maj}(= z)$ varies with c for both the unrewired and rewired models. Note that, as one might expect, $p_{maj}(= z)$ decreases with c for both models; indeed this is proved formally for the rewired model in Sect. 5.2. Note also that $p_{maj}(= z)$ 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 derive asymptotic features of the model. 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 threshold parameter R_* and the probability p_{maj} and relative final 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
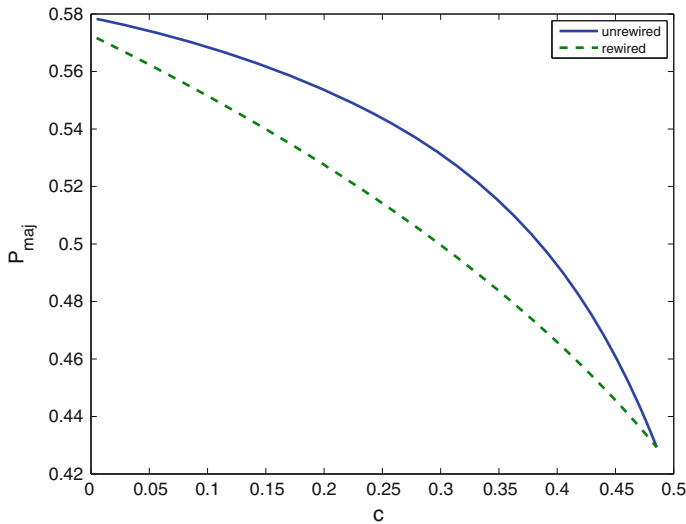
Fig. 7 Demonstration of different epidemic properties of models with the same network properties (D, c, ρ) but constructed using different models parameters (H, G, r, p_{RW}). The plots are of $p_{\text{maj}} (= z)$ versus c when $D \sim \text{Poi}(10)$, $\rho = 0.2$, $n_Q = 10$ and $p_I = 0.15$; each value of c being achieved using both our unrewired and rewired network models

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 Eq. (1), and the facts that ρ is increasing in r (see Proposition 3 and its following comment) and that, under rewiring, c decreases in p_{RW} (see Proposition 4). The relationship between c and H is more complicated, but, loosely speaking, having larger households increases c (see the discussion at the end of Sect. 3.2). 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).

A limitation of the paper is that tuning of model parameters to achieve a specified (D, c, ρ) and numerical experiments are both considered only for a Poisson degree distribution (with the minor exception of some preliminary investigations involving a heavy tailed global degree distribution mentioned at the end of the third paragraph of Sect. 6). Further research involving other choices of degree distribution is clearly worthwhile. Indeed some degree distributions D are not achievable by our model (see comment after Proposition 1). The model of Trapman (2007) (see also Coupechoux and Lelarge 2012) provides a method of constructing a network with any specified degree distribution D (subject to a mild moment condition) and any

clustering coefficient in the range $[0, C_D^{\max}]$, where the upper limit $C_D^{\max}$ depends on the degree distribution D; see Eq. (7) of Trapman (2007), though note that this assumes $P(D = 0) = 0$, and Proposition 7 of Coupechoux and Lelarge (2012). This model, in which degree correlation ρ is not controlled, is similar to our model with $r = 0$ (or $n_Q = 1$), the difference being that individuals residing in households of size $h > 1$ have precisely one global neighbour. Thus our method of pairing stubs can be applied to that model also, although the range of achievable clustering coefficients is smaller. For example, if $D \sim \text{Poi}(\gamma)$ then the maximum achievable clustering coefficient is $C_D^{\max} = 1 - 2\gamma^{-2}(\gamma - 1 + \exp(-\gamma))$, compared to 1 using our model.

A rather different algorithm for constructing a network with given degree distribution and clustering coefficient is that of Volz (2004). This algorithm starts by allocating stubs to nodes using a random sample from D, as in the configuration model, and then forms the network on a component basis. Each component starts from a node selected randomly from nodes not currently in the network and is grown by a complicated iterative procedure, akin to a branching process, that uses a Metropolis–Hastings Markov chain Monte Carlo method. In contrast to our model, degree correlation is not tuned and the complexity of the algorithm means that it is generally not possible to calculate asymptotic properties of epidemics defined on the network, which instead have to be estimated empirically using simulations. However, Volz's algorithm aims to produce a network that is sampled uniformly at random from the space of all networks on a set of n nodes having specified degree distribution D and clustering coefficient c. (Volz notes that it is difficult to prove that his algorithm yields a network that is truly unbiased in this sense but that, as $n \to \infty$, it does produce networks that satisfy many properties of such an unbiased network.) Thus, although more calculation is possible with our model, the networks it produces come from a different distribution defined on the above space.

It is important to observe that, as illustrated in Fig. 7, there may be distinct network models having the same local network features D, ρ and c but 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.

In the current work, households are used purely as a device for incorporating clustering into a network model. Household structure is a key feature of human populations, that has significant impact on disease dynamics and control, and applicability of our model to real-world populations is an interesting area of research that is not addressed in this paper. The unrewired network model is a model of social connectivity that includes households (cf. Ball et al. 2010), but note that rewiring may change substantially the distribution of household size. However, in both the unrewired and rewired models, clustering is present essentially only within households. In real-world applications we may also want to introduce clustering in the network of global contacts. One way of achieving this would be to first use the present model to give possible between-household

connections and then use a second partitioning of the population to give actual households.

Appendix A: Clustering coefficient for the unrewired network

In this appendix we outline how the proofs in Appendix C of Ball et al. (2010) may be adapted to prove Theorem 1. To facilitate connection with the proofs in Ball et al. (2010), it is convenient to index our network model by the number of households, m say, rather than by the number of individuals n. More specifically, let $H_1, H_2, \ldots$ and $G_1, G_2, \ldots$ be independent and identically distributed copies of H and G, respectively. Then a realisation, $\mathcal{G}^{(m)}$ say, of our network model on m households is constructed in the obvious fashion using $H_1, H_2, \ldots, H_m$ and $G_1, G_2, \ldots, G_{N(m)}$, where $N(m) = \sum_{i=1}^{m} H_i$. Let $\hat{c}^{(m)}$ denote the proportion of ordered triplets in $\mathcal{G}^{(m)}$ that are ordered triangles. We show that, under the conditions of Theorem 1 (i.e. that σ_H^2 and σ_G^2 are both finite and n_Q is fixed and independent of n, or equivalently m), $\hat{c}^{(m)} \overset{\text{a.s.}}{\to} c$ as $m \to \infty$, where c is given by (3); from which $c^{(n)} \overset{\text{a.s.}}{\to} c$ as $n \to \infty$ follows easily since the number of individuals per household converges almost surely to $\mu_H \in [1, \infty)$ as $m \to \infty$.

A key step in the proofs in Ball et al. (2010) is the construction from $\mathcal{G}^{(m)}$ of an auxiliary graph $\hat{\mathcal{G}}^{(m)}$, having vertices $1, 2, \ldots, m$, in which all vertices in the same household, i say, in $\mathcal{G}^{(m)}$ are replaced by the single vertex i in $\hat{\mathcal{G}}^{(m)}$ and any global edge between individuals in households i and j in $\mathcal{G}^{(m)}$ is replaced by an edge between vertices i and j in $\hat{\mathcal{G}}^{(m)}$. Thus the degree of a typical vertex in $\hat{\mathcal{G}}^{(m)}$ is distributed as $\sum_{i=1}^{H_1} G_i$. Moreover, a realisation of $\hat{\mathcal{G}}^{(m)}$ can be obtained by suitable pairing of stubs, with stubs in $\hat{\mathcal{G}}^{(m)}$ being in one-to-one correspondence with stubs in $\mathcal{G}^{(m)}$ and also sharing the same X-labels.

Observe that (1) a triangle consisting of two household edges and one global edge in $\mathcal{G}^{(m)}$ has a corresponding self-loop in $\hat{\mathcal{G}}^{(m)}$, (2) a triangle consisting of one household edge and two global edges in $\mathcal{G}^{(m)}$ has a corresponding pair of parallel edges in $\hat{\mathcal{G}}^{(m)}$ and (3) a triangle consiting of three global edges in $\mathcal{G}^{(m)}$ has a corresponding triangle in $\hat{\mathcal{G}}^{(m)}$. Thus the number of ordered triangles not wholly within households in $\mathcal{G}^{(m)}$ is no more than the sum of the numbers of self-loops, parallel edges and ordered triangles in $\hat{\mathcal{G}}^{(m)}$.

Let L_m be the total number of stubs in $\hat{\mathcal{G}}^{(m)}$. If $r = 0$, when the network $\mathcal{G}^{(m)}$ reduces to the model of Ball et al. (2010), these stubs are joined uniformly at random to create $\hat{\mathcal{G}}^{(m)}$. Let $s_1, s_2, \ldots, s_{2k}$ denote $2k$ specified stubs and E_k be the event that s_1 is paired with s_2, s_3 is paired with $s_4, \ldots$ and s_{2k-1} is paired with s_{2k}. Then the conditional probability that E_k occurs given L_m is $L_m^{[k]}$, where $a^{[k]} = [(a-1)(a-3) \ldots (a-2k+1)]^{-1}$. Suppose that $r \neq 0$, let $L_m(0)$ be the total number of stubs in $\hat{\mathcal{G}}^{(m)}$ that have X-label 0 and, for $i = 1, 2, \ldots, n_Q$, let $L_m(i)$ be the total number of stubs in $\hat{\mathcal{G}}^{(m)}$ that have

X-label 1 and belong to the ith quantile. Let $\hat{L}_m = \min\{L_m(0), L_m(1), \ldots, L_m(n_Q)\}$ if $|r| \neq 1$ and $\hat{L}_m = \min\{L_m(1), L_m(2), \ldots, L_m(n_Q)\}$ if $|r| = 1$. The conditional probability that the above event E_k occurs, given $L_m(0), L_m(1), \ldots, L_m(n_Q)$, is at most $\hat{L}_m^{[k]}$. (Note that the conditional probability may be zero, for example if a pair of stubs have different X-values.) By the strong law of large numbers, $m^{-1}L_m(0) \overset{\text{a.s.}}{\to} (1 - |r|)\mu_H\mu_G$ and $m^{-1}L_m(i) \overset{\text{a.s.}}{\to} |r|\mu_H\mu_G/n_Q (i = 1, 2, \ldots, n_Q)$ as $m \to \infty$. It follows that there exists a constant $\hat{c} > 0$ such that, almost surely, $\hat{L}_m > \hat{c}m$ for all but finitely many m.

It is readily verified that the proof of Lemma 2 in Ball et al. (2010, Appendix C) carries over to the present model. The only material change is that the quantities $L_m^{[k]}$ in that proof are replaced by the corresponding $\hat{L}_m^{[k]}$ and equality in the final displayed equation in the left hand column of page 70 of Ball et al. (2010) is replaced by inequality ($\leq$). Theorem 5 of Ball et al. (2010) then follows, whence the number of ordered triangles in $\hat{\mathcal{G}}^{(m)}$ per vertex converges almost surely to 0 as $m \to \infty$. As noted in the remark near the end of Appendix C of Ball et al. (2010), similar arguments show that the number of self-loops and parallel edges in $\hat{\mathcal{G}}^{(m)}$ per vertex also converge almost surely to 0 as $m \to \infty$, whence so does the number of triangles not wholly in households per household in $\mathcal{G}^{(m)}$. Theorem 1 of the present paper then follows when H has finite third moment.

Suppose that now that H has infinite third but finite second moment. Recall the notation $N_{\text{T}}^{(n)}$ and $N_{\triangle}^{(n)}$ from Sect. 3.2 and let $N_{\text{T},\tilde{\text{H}}}^{(n)}$ be the number of ordered triplets that contain at least one global edge in the network, which, as in Sect. 3.2, is indexed by its number of nodes n. Now

$$c^{(n)} = 1 - \frac{N_{\text{T}}^{(n)} - N_{\triangle}^{(n)}}{N_{\text{T}}^{(n)}} \geq 1 - \frac{N_{\text{T},\tilde{\text{H}}}^{(n)}}{N_{\text{T}}^{(n)}},$$

since any triplet in a household is necessarily a triangle. Using the strong law of large numbers, $n^{-1}N_{\text{T}}^{(n)} \overset{\text{a.s.}}{\to} \infty$ and $n^{-1}N_{\text{T},\tilde{\text{H}}}^{(n)} \overset{\text{a.s.}}{\to} \mu_H^{-1}\text{E}[H\{2G(H-1) + G(G-1)\}]$ as $n \to \infty$ (see the final paragraph of Appendix C of Ball et al. (2010)), whence $c^{(n)} \overset{\text{a.s.}}{\to} 1$ as $n \to \infty$, since $\mu_H^{-1}\text{E}[H\{2G(H-1) + G(G-1)\}]$ is finite. This completes the proof of Theorem 1.

Appendix B: Clustering coefficient for the rewired network

In this appendix we show that Theorem 1 holds for the rewired network model, under the stronger condition that $\text{E}[H^3]$ is finite. Observe that the rewiring process described in Sect. 3.4 does not change (1) the total number of ordered triplets or (2) the total number of ordered triangles involving only global edges. Also, any triangle involving at least one household edge of an unbroken household is unchanged by rewiring. For the rewired network, let $\tilde{N}_{\triangle,\text{U}}^{(n)}$ denote the total number of ordered triangles that are wholly within unbroken households and $\tilde{N}_{\triangle,\text{B}}^{(n)}$ denote the total number of ordered triangles that include at least one broken household edge. By the strong law

of large numbers, $n^{-1}\tilde{N}_{\triangle,\text{U}}^{(n)} \overset{\text{a.s.}}{\to} \mu_H^{-1}(1 - p_{RW})\text{E}[H(H-1)(H-2)]$ as $n \to \infty$ and $c^{(n)}(G, H, r, p_{RW}) \overset{\text{a.s.}}{\to} c(G, H, r, p_{RW})$ as $n \to \infty$ then follows if $n^{-1}\tilde{N}_{\triangle,\text{B}}^{(n)} \overset{\text{a.s.}}{\to} 0$ as $n \to \infty$.

Recall that under rewiring the red stubs arising from broken households are paired only with such stubs labelled with the same household size. If the support of H is finite then a similar argument to that used above for the number of ordered triangles in $\hat{\mathcal{G}}^{(m)}$ yields that $n^{-1}\tilde{N}_{\triangle,\text{B}}^{(n)} \overset{\text{a.s.}}{\to} 0$ as $n \to \infty$. However, this breaks down when H has infinite support, since the term corresponding to $\hat{L}_m$ becomes the minimum of an infinite set. For $h = 1, 2, \ldots$, the random variable $\tilde{N}_{\triangle,\text{B}}^{(n)}$ can be expressed as $\tilde{N}_{\triangle,\text{B}}^{(n)} = \tilde{N}_{\triangle,\text{B},\leq h}^{(n)} + \tilde{N}_{\triangle,\text{B},>h}^{(n)}$, where

$$\tilde{N}_{\triangle,\text{B},\leq h}^{(n)} = \sum_{h'=1}^{h} \tilde{N}_{\triangle,\text{B},h'}^{(n)}, \qquad \tilde{N}_{\triangle,\text{B},>h}^{(n)} = \sum_{h'=h+1}^{\infty} \tilde{N}_{\triangle,\text{B},h'}^{(n)},$$

and $\tilde{N}_{\triangle,\text{B},h'}^{(n)}$ is the number of ordered triangles that include at least one size-h' broken household edge; note that a triangle cannot contain broken household edges of differing sizes.

By a similar argument to that in the finite support case, $n^{-1}\tilde{N}_{\triangle,\text{B},\leq h}^{(n)} \overset{\text{a.s.}}{\to} 0$ as $n \to \infty$. Now $\tilde{N}_{\triangle,\text{B},>h}^{(n)} \leq \tilde{N}_{\text{T,B},>h}^{(n)}$, where $\tilde{N}_{\text{T,B},>h}^{(n)}$ is the number of ordered triplets that contain at least one broken household edge having size $> h$. By the strong law of large numbers,

$$n^{-1}\tilde{N}_{\text{T,B},>h}^{(n)} \overset{\text{a.s.}}{\to} \mu_H^{-1} p_{RW}\text{E}\left[H(H-1)(H-2+2G)1_{\{H>h\}}\right] \quad \text{as} \quad n \to \infty, \quad (38)$$

where $1_{\{H>h\}}$ is the indicator function of the event $\{H > h\}$. Now $\text{E}[H(H-1)(H-2+2G)]$ is finite, since $\text{E}[H^3]$ and $\text{E}[G]$ are both finite, so the limit in (38) may be made arbitrarily close to zero by choosing h sufficiently large, whence $n^{-1}\tilde{N}_{\triangle,\text{B}}^{(n)} \overset{\text{a.s.}}{\to} 0$ as $n \to \infty$ and $c^{(n)}(G, H, r, p_{RW}) \overset{\text{a.s.}}{\to} c(G, H, r, p_{RW})$ as $n \to \infty$ follows.

Appendix C: Derivation of degree correlation ρ

In this appendix we derive the formula for the degree correlation ρ for our model given in Proposition 3. 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 $\rho = \text{corr}(X_L, X_R)$, 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 $\text{P}(I_G = 1) = p_G = 1 - \text{P}(I_G = 0)$. We determine first the probability p_G that E is a global edge.

Let $N_G^{(n)}$ and $N_H^{(n)}$ denote respectively the number of global and household edges in the network. Then $\mu_{N_G^{(n)}} = \frac{n}{2}\mu_G$, since each stub contributes to half an edge, and $\mu_{N_H^{(n)}} = \frac{n}{2}\sum_{h=1}^{\infty}(h-1)P(\tilde{H} = h) = \frac{n}{2}(\mu_{\tilde{H}} - 1)$, since the household size of an individual chosen uniformly 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 (4).

Note that

$$\text{cov}(X_L, X_R) = \text{E}[\text{cov}(X_L, X_R | I_G)] + \text{cov}(\text{E}[X_L | I_G], \text{E}[X_R | I_G]). \quad (39)$$

We calculate the two quantities on the right hand side of (39) 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,

$$\text{cov}(X_L, X_R | I_G = 0) = \sigma_{\hat{H}}^2. \quad (40)$$

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 $i, j = 1, 2, \ldots, n_Q$,

$$P(Q_L = i, Q_R = j) = \begin{cases} \frac{1-r}{n_Q^2} + \delta_{i,j} \frac{r}{n_Q} & \text{if } r \geq 0, \\ \frac{1-|r|}{n_Q^2} + \delta_{i, n_Q + 1 - j} \frac{|r|}{n_Q} & \text{if } r < 0. \end{cases} \quad (41)$$

Now,

$$\text{cov}(X_L, X_R | I_G = 1) = \text{E}[\text{cov}(X_L, X_R | I_G = 1, Q_L, Q_R)] \\ + \text{cov}(\text{E}[X_L | I_G = 1, Q_L], \text{E}[X_R | I_G = 1, Q_R]). \quad (42)$$

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

$$\text{cov}(X_L, X_R | I_G = 1, Q_L, Q_R) = 0. \quad (43)$$

Further, for $i = 1, 2, \ldots, n_Q$, $\text{E}[X_L | I_G = 1, Q_L = i] = \text{E}[X_R | I_G = 1, Q_R = i] = \mu_{\tilde{D}}^{(i)}$ (see Eq. (6)). Using the distribution (41) and noting that $\mu_{\tilde{D}} = n_Q^{-1} \sum_{i=1}^{n_Q} \mu_{\tilde{D}}^{(i)}$ yields

$$\text{cov}(\text{E}[X_L | I_G = 1, Q_L], \text{E}[X_R | I_G = 1, Q_R]) = g_{\tilde{D}, n_Q}(r), \quad (44)$$

where $g_{\tilde{D}, n_Q}(r)$ is defined at (5).

Note that $P(I_G = 1) = p_G = 1 - P(I_G = 0)$. Then, Eqs. (40), (42), (43) and (44) yield

$$\text{E}[\text{cov}(X_L, X_R | I_G)] = (1 - p_G)\sigma_{\hat{H}}^2 + p_G g_{\tilde{D}, n_Q}(r). \quad (45)$$

We turn now to the second quantity on the right hand side of (39). Note that $\text{E}[X_L | I_G] = \text{E}[X_R | I_G]$, so $\text{cov}(\text{E}[X_L | I_G], \text{E}[X_R | I_G]) = \text{var}(\text{E}[X_L | I_G])$. Suppose that $I_G = 0$. Then, in the above notation, $X_L = H_E - 1 + G_L$, where $G_L \overset{d}{=} G$. Thus,

$$\mathrm{E}[X_L|I_G = 0] = \mu_{\hat{H}-1} + \mu_G. \tag{46}$$

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,

$$\mathrm{E}[X_L|I_G = 1] = \mu_{\tilde{H}-1} + \mu_{\tilde{G}}. \tag{47}$$

Recalling that $\mathrm{P}(I_G = 1) = p_G = 1 - \mathrm{P}(I_G = 0)$ and that $\mu_{\tilde{G}} = \mathrm{E}[G^2]/\mu_G$, Eqs. (46) and (47) yield

$$\mathrm{cov}(\mathrm{E}[X_L|I_G], \mathrm{E}[X_R|I_G]) = p_G(1 - p_G)\left(\mu_{\hat{H}} - \mu_{\tilde{H}} - \frac{\sigma_G^2}{\mu_G}\right)^2. \tag{48}$$

Combining Eqs. (39), (45) and (48) gives

$$\mathrm{cov}(X_L, X_R) = (1 - p_G)\sigma_{\hat{H}}^2 + p_G g_{\tilde{D},n_Q}(r) + p_G(1 - p_G)\left(\mu_{\hat{H}} - \mu_{\tilde{H}} - \frac{\sigma_G^2}{\mu_G}\right)^2. \tag{49}$$

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

$$\mathrm{var}(X_L) = \mathrm{E}[\mathrm{var}(X_L|I_G)] + \mathrm{var}(\mathrm{E}[X_L|I_G]). \tag{50}$$

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 $\mathrm{var}(X_L|I_G = 0) = \sigma_{\hat{H}}^2 + \sigma_G^2$; and if $I_G = 1$ then $X_L \overset{d}{=} \tilde{H} - 1 + \tilde{G}$, where $\tilde{H}$ and $\tilde{G}$ are independent, so $\mathrm{var}(X_L|I_G = 1) = \sigma_{\tilde{H}}^2 + \sigma_{\tilde{G}}^2$. Hence,

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

which on substituting into (50), 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 (48) yields

$$\mathrm{var}(X_L) = (1 - p_G)\left(\sigma_{\hat{H}}^2 + \sigma_G^2\right) + p_G\left(\sigma_{\tilde{H}}^2 + \sigma_{\tilde{G}}^2\right) + p_G(1 - p_G)\left(\mu_{\hat{H}} - \mu_{\tilde{H}} - \frac{\sigma_G^2}{\mu_G}\right)^2. \tag{51}$$

The expression (7) for the degree correlation ρ, given in Sect. 3.3, follows from Eqs. (49) and (51), since $\mathrm{var}(X_L) = \mathrm{var}(X_R)$.

References

Andersson H (1997) Epidemics in a population with social structures. Math Biosci 140:79–84
Andersson H (1998) Limit theorems for a random graph epidemic model. Ann Appl Prob 8:1331–1349
Andersson H (1999) Epidemic models and social networks. Math Sci 24(2):128–147

Andersson H, Britton T (2000) Stochastic epidemic models and their statistical analysis. Springer lecture notes in statistics, vol 151. Springer, New York

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

Ball FG (1983) The threshold behaviour of epidemic models. J Appl Prob 20:227–241

Ball FG (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 FG (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 FG, Lyne OD (2001) Stochastic multitype SIR epidemics among a population partitioned into households. Adv Appl Prob 33:99–123

Ball FG, Mollison D, Scalia-Tomba G (1997) Epidemics with two levels of mixing. Ann Appl Prob 7:46–89

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

Ball FG, O'Neill PD (1999) The distribution of general final state random variables for stochastic epidemic models. J Appl Prob 36:473–491

Ball FG, Sirl DJ (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 FG, Sirl DJ, 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 FG, Sirl DJ, Trapman P (2010) Analysis of a stochastic SIR epidemic on a random network incorporating household structure. Math Biosci 224: 53–73 (See also erratum (2010), ibid. 225:81.)

Ball FG, Sirl DJ, Trapman P (2012) Epidemics on random intersection graphs (submitted). arXiv:1011.4242

Britton T, Lindholm M, Turova T (2011) A dynamic network in a dynamic population: asymptotic properties. J Appl Prob 48:1163–1178

Britton T, Nordvik MK, 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, Lagerås AN (2008) Epidemics on random graphs with tunable clustering. J Appl Prob 45:743–756

Coupechoux E, Lelarge M (2012) How clustering affects epidemics in random, networks. arXiv:1202.4974

Diekmann O (1978) Thresholds nad traveling waves for the geographical spread of infection. J Math Biol 6:109–130

Diekmann O, Heesterbeek JAP (2000) Mathematical epidemiology of infectious diseases. John Wiley, Chichester

Diekmann O, Heesterbeek JAP, Metz JAJ (1990) On the definition and computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. J Math Biol 28:365–382

Diekmann O, de Jong MCM, Metz JAJ (1998) A deterministic epidemic model taking account of repeated contacts between the same individuals. J Appl Prob 35:448–462

Durrett R (2006) Random graph dynamics. Cambridge University Press, Cambridge

Erdős P, Rényi A (1959) On random graphs. Publicationes Mathematicae 6:290–297

Feller W (1971) An introduction to probability theory and its applications, vol II, 2nd edn. Wiley, New York

Gleeson JP (2009) Bond percolation on a class of clustered random networks. Phys Rev E 80:036107

Gleeson JP, Melnik S, Hackett A (2010) How clustering affects the bond percolation threshold in complex networks. Phys Rev E 81:066114

van der Hofstad R, Litvak N (2012) Degree–degree correlations in random graphs with heavy-tailed degrees. arXiv:1202.3071v3

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

Janson S (2009) The probability that a random multigraph is simple. Combinatorics Prob Comput 18:205–225

Karrer B, Newman MEJ (2010) Random graphs containing arbitrary distributions of subgraphs. Phys Rev E 82:066118

Ma J, van den Driessche P, Willeboordse FH (2012) Effective degree household network disease model. J Math Biol. doi:10.1007/s00285-011-0502-9

May RM, Anderson RM (1987) Transmission dynamics of HIV infections. Nature 326:137–142

Miller JC (2009) Percolation and epidemics in random clustered networks. Phys Rev E 80:020901(R)

Mode CJ (1971) Multitype branching processes. Theory and applications. In: Modern analytic and computational methods in science and mathematics, vol 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, Reed B (1995) A critical point for random graphs with a given degree sequence. Rand Struct Alg 6:161–179

Newman MEJ, Strogatz SH, Watts DJ (2001) Random graphs with arbitrary degree distributions and their applications. Phys Rev E 64:026118

Newman MEJ (2002a) Assortative mixing in networks. Phys Rev Lett 89:208701

Newman MEJ (2002b) Spread of epidemic disease on networks. Phys Rev E 66:016128

Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45:167–256

Newman MEJ (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

Volz E (2004) Random networks with tunable degree distribution and clustering. Phys Rev E 70:056115

Warde WD, Katti SK (1971) Infinite divisibility of discrete distributions, II. Ann Math Stat 42:1088–1090

Watts SC, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393:440–442