Mathematical Statistics
Stockholm University

SIR epidemics and vaccination on random graphs with clustering

Carolina Fransson and Pieter Trapman

Research Report 2018:8

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

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

SIR epidemics and vaccination on random graphs with clustering

Carolina Fransson[*]

Pieter Trapman[†]

February 2018

Abstract

In this research report we consider SIR (Susceptible $\rightarrow$ Infectious $\rightarrow$ Recovered) epidemics on random graphs with clustering. To incorporate group structure of the underlying social network, we use a generalized version of the configuration model in which each node is a member of a specified number of triangles. SIR epidemics on this type of graph have earlier been investigated under the assumption of homogeneous infectivity and also under the assumption of Poisson transmission and recovery rates.

We extend known results from literature by relaxing the assumption of homogeneous infectivity. An important special case of the epidemic model analyzed in this paper is epidemics in continuous time with arbitrary infectious period distribution. We use branching process approximations of the spread of the disease to provide expressions for the basic reproduction number R_0, the probability of a major outbreak and the expected final size. In addition, the impact of random vaccination with a perfect vaccine on the final outcome of the epidemic is investigated. We find that, for this particular model, R_0 equals the perfect vaccine-associated reproduction number.

Generalizations to groups larger than three are discussed briefly.

Keywords: SIR epidemics, Configuration model, Clustering, Branching processes, Vaccination.

[*]Stockholm University, Sweden; carolina.fransson@math.su.se
[†]Stockholm University, Sweden; ptrapman@math.su.se

1　Introduction

One of the most important factors that determine the fate of an outbreak of an infectious disease is the contact pattern of individuals in the population. The frequency and duration of the contacts between individuals typically depend on the nature of their relationship. For this reason, recent interest has focused on the impact of the underlying social network on the spread of the disease. The social network is typically represented by a random graph (Newman et al. 2002), in which the nodes or vertices represent individuals and the edges represent social contacts between the individuals. Two nodes that share an edge are called "neighbors".

A popular choice when generating random graphs with a specified degree distribution is the configuration model (CM). It was introduced by Bollobás (1980) for the special case where the degree distribution is degenerate (i.e. every node of the graph has the same degree) and extended to more general degree distributions by Molloy and Reed (1995, 1998). There is a vast literature on epidemics on configuration model graphs (see e.g. Andersson (1999); Britton et al. (2007); Janson et al. (2014); Barbour and Reinert (2013); Bhamidi et al. (2014)).

An important feature of the configuration model is that, under mild regularity conditions on the degrees, this type of graph is asymptotically unclustered. That is to say, it contains virtually no groups and short circuits. Real world networks do, however, typically exhibit clustering (Newman 2003), and there are a number of graph models that do allow for group structure (Bollobás et al. 2011; Karoński et al. 1999; Newman 2002). Epidemics on graphs with group structure were studied by Trapman (2007); Ball et al. (2009, 2010, 2014); Coupechoux and Lelarge (2015); Britton et al. (2008).

In this paper, we use a generalized version of the configuration model to incorporate clustering of the social network in the analysis of the spread of an infectious disease. The configuration model with clustering (CMC) was independently introduced by Miller (2009) and Newman (2009). It is an extension of the CM in the sense that, for each node u, in addition to the degree of u one also specifies the number of pairs of neighbors of u that are in turn neighbors of each others. In other words, one specifies the number of triangles (with non-overlapping edges) of which u is a member (see section 2.1 for a precise definition of the graph model). This allows for graphs with non-negligible clustering and a specified degree distribution. That is to say, the CMC deviates from the classical Erdős-Rényi graph model (Erdős and Rényi 1959) in two fundamental ways: it allows for for a non-Poissonian degree distributions and is asymptoticly clustered. Epidemics on this type of graph have previously been studied by Miller (2009) and Volz et al. (2011).

Miller (2009) investigated the impact of clustering on the epidemic threshold, formulated as a bond percolation problem. This means that the infectivity of infected individuals is assumed to be homogeneous; an infected individual transmits the disease to each of its neighbors independently with some fixed probability T. Volz et al. (2011) investigated the time evolution and final size of epidemics on CMC graphs under the assumption of exponentially distributed infectious periods during which individuals contact neighbors at a constant rate.

The main contribution of our research is that we extend the results of Miller (2009) and Volz et al. (2011) by allowing for heterogeneous infectivity, i.e. by allowing for some infected individuals to be more contagious than others. Such heterogeneity may, for instance, reflect variability in the infectious period. We provide expressions for the probability of a major outbreak and the final size of an major outbreak. A key tool in our analysis is the approximation of the epidemic seen from a "generation of infection" or "rank" perspective by a multitype Galton Watson branching process. This approximation, which is interesting in its own right, gives rise to the rank based reproduction number R_0 (see e.g. Pellis et al. (2008, 2012)).

The second contribution of this paper concerns vaccination. We investigate the impact of uniform vaccination (i.e. vaccinated individuals are selected uniformly at random) with a perfect vaccine (i.e. a vaccine that provides full and permanent immunity to the disease). We find that it is necessary to vaccinate a fraction $1 - 1/R_0$ of the population in order to prevent a major outbreak of the disease, as in the case of homogeneous mixing. We illustrate our findings with numerical examples.

This paper is structured as follows. In Section 2 we provide the preliminaries for the model. In Section 2.1 we give a more detailed description of how graphs are generated in the CMC and investigate the asymptotic clustering of such graphs and in Section 2.2 the epidemic model is specified. Section 2.3- 2.4 contains an overview of the concept of reproduction numbers and the necessary branching process background. In Section 3, we derive expressions for the probability of a major outbreak and the expected final size under the assumption of an unvaccinated and fully susceptible population, and in Section 4 the analysis is repeated under the assumption of uniform vaccination with a perfect vaccine. We illustrate our findings with numerical examples presented in Section 5 and discuss possible extensions in Section 6.

2 Preliminaries

2.1 The configuration model with clustering

A configuration model with clustering CMC graph is constructed as follows. Let $\{p(k_s, k_\Delta)\}_{k_s, k_\Delta \in \mathbb{N}_0}$ be a prescribed joint degree distribution, where k_s denotes the number of single edges attached to a node, and k_Δ denotes the number of pairs of triangle edges. Throughout, (S, Δ) is assumed to be a generic random vector distributed according to p. Let $\{(S_i, \Delta_i)\}_{i=1}^N$ be a sequence of independent copies of (S, Δ). Analogously to the CM, a graph $G_N = G_N(p)$ of size N is constructed by first assigning the single degree S_i and the triangle degree Δ_i to the node v_i, $i = 1, 2, \ldots, N$. One may think of this step in terms of half-edges; to each node v_i, we attach S_i single half-edges and Δ_i pairs of triangle half-edges. The single half-edges are then matched in pairs and the triangle half-edge pairs in threes by choosing a matching uniformly at random among all possible such matchings. The process of joining half-edges is illustrated in Figure 1. As described in Miller (2009), the matching may be carried out as follows. Two lists of nodes, one single degree list and one triangle degree list are created. A node with joint degree (k_s, k_Δ) appears k_s times in the single list and k_Δ times in the triangle list. The lists are then shuffled uniformly, and the nodes on positions $2m + 1$ and $2m + 2$ in the single degree list and positions $3m + 1$, $3m + 2$ and $3m + 3$ in the triangle degree list are matched, $m \in \mathbb{N}_0$.
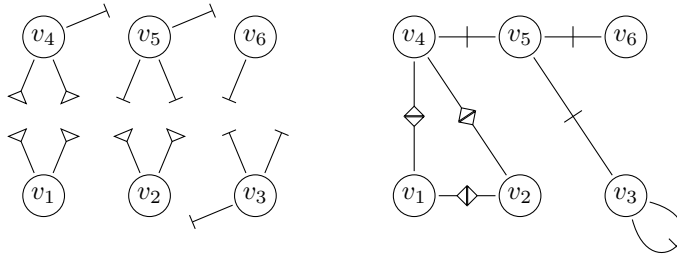

Figure 1: Schematic illustration of the construction of a CMC graph. Triangle half-edges (marked with a triangle) and single half-edges (marked with a perpendicular line) are assigned to the nodes of the graph (left). The half-edges are then matched uniformly at random (right). Note that two of the half-edges attached to v_3 are paired with each other and so form a self-loop.

We define the *total single degree* as

$$D_S^{(N)} := \sum_{i=1}^{N} S_i$$

and the *total triangle degree* as

$$D_\Delta^{(N)} := \sum_{i=1}^{N} \Delta_i.$$

If the total single degree (that is, the length of the single degree list) is not even or if the total triangle edge degree (the length of the triangle degree list) is not a multiple of three we erase a single half-edge and/or one or two triangle half-edge pairs chosen uniformly at random. Similarly, we erase self-loops and merge multiple edges, so that the resulting graph is simple. Under assumption A1 (stated below) on p it holds that the number of single self-loops and single double edges converge in distribution to independent Poisson random variables with finite means (cf. Van der Hofstad (2016, Prop. 7.13)).

For this reason, self-loops and multiple edges are negligible in the limit as $N \to \infty$. In the remainder of this paper, we ignore the small differences in the topology of the graph that arise from erasing multiple edges or self-loops. In addition, we ignore the small differences in effective degree distribution that arise from erasing half-edges so that the number of single and triangle half-edges are multiples of two and three, respectively.

We make the following assumptions on p.

A1) $E(\Delta^2) < \infty$ and $E(S^2) < \infty$.

A2) $P(\max(\Delta, S) \geq 2) > 0$ and $E(\Delta S) > 0$.

Note that the assumption A1 implies $E(\Delta S) < \infty$. Assumption A2 ensures that the mean matrices of the approximating branching processes (presented below) are positively regular (we say that an $r \times r$ matrix M is positively regular if it has finite non-negative entries and for some $n \in \mathbb{N}$ all entries of M^n are strictly positive).

2.1.1 Clustering coefficient of G_N

For any undirected graph we can measure the ammount of clustering in the network using the so-called clustering coefficient, which is defined as follows. Let $G = (V, E)$ be an undirected graph with node set V and edge set E. Define

$$\mathcal{W}_\wedge^G = \{(u, v, w) \in V^3 : (u, v), (v, w) \in E\}$$

the set of all ordered wedges (i.e. directed paths consisting of precisely two edges) of G and

$$\mathcal{W}_\Delta^G = \{(u, v, w) \in V^3 : (u, v), (v, w), (w, u) \in E\} \subset \mathcal{W}_\wedge^G$$

the set of all ordered triangles of G. The *clustering coefficient* $C(G)$ of G is a measure of the degree of clustering of G and is defined as the fraction of the ordered wedges of G that are also triangles:

$$C(G) = \frac{|\mathcal{W}_\Delta^G|}{|\mathcal{W}_\wedge^G|}.$$

As stated in the following proposition, CMC graphs have asymptotically non-zero clustering as $N \to \infty$. An analogous result for fixed degree sequences was presented in Newman (2009). Let $\overset{P}{\longrightarrow}$ denote convergence in probability.

Proposition 1. Let $\{G_N\}_N$ be a sequence of CMC graphs with independent degrees drawn from p. If p satisfies assumption A1 then

$$C(G_N) \overset{P}{\longrightarrow} \frac{E(2\Delta)}{E((2\Delta + S)^2) - E(2\Delta + S)}. \tag{1}$$

The proof is presented in the Appendix.

2.1.2 Downshifted size-biased degrees

The graph G_N may be constructed by joining the half-edges in a random order. In particular, G_N may be constructed as the epidemic progresses; starting with the initial infected case we sequentially match the half-edges along which the disease is transmitted. Since half-edges are chosen uniformly at random in the matching procedure, the probability to choose a specific node is proportional to the number of free half-edges attached to the node in question. That is, if we pair a single half-edge, the probability of choosing a specific node with k_s unpaired single half-edges is proportional to k_s. For this reason, the degree distribution a node explored by joining a single half-edge in the early phase of the epidemic can be approximated by the *single size biased* degree distribution $p^{(s)}$

$$p^{(s)}(k_s, k_\Delta) = \frac{k_s p(k_s, k_\Delta)}{E(S)}. \tag{2}$$

Similarly, the degree distribution of the nodes explored by joining three triangle half-edge pairs in the early phase of the epidemic can be approximated by the *triangle size biased* degree distribution $p^{(\Delta)}$

$$p^{(\Delta)}(k_s, k_\Delta) = \frac{k_\Delta p(k_s, k_\Delta)}{E(\Delta)}. \tag{3}$$

In the epidemic process, we need to account for the fact that an infected individual has at least one non-susceptible neighbor (namely the direct source of its infection). For this reason, we introduce the *downshifted* size biased degree distributions $p_\bullet^{(s)}$ and $p_\bullet^{(\Delta)}$, given by

$$\begin{aligned}
p_\bullet^{(s)}(k_s, k_\Delta) &= p^{(s)}(k_s + 1, k_\Delta) \\
p_\bullet^{(\Delta)}(k_s, k_\Delta) &= p^{(\Delta)}(k_s, k_\Delta + 1).
\end{aligned} \tag{4}$$

Throughout, we will make frequent reference to the following random vectors

$$\begin{aligned}
(S_\bullet^{(s)}, \Delta_\bullet^{(s)}) &\sim p_\bullet^{(s)} \\
(S_\bullet^{(\Delta)}, \Delta_\bullet^{(\Delta)}) &\sim p_\bullet^{(\Delta)}
\end{aligned} \tag{5}$$

and the expected values

$$\begin{aligned}
E(S_\bullet^{(s)}) &= \frac{E(S^2)}{E(S)} - 1 & E(\Delta_\bullet^{(s)}) &= \frac{E(S\Delta)}{E(S)} \\
E(S_\bullet^{(\Delta)}) &= \frac{E(S\Delta)}{E(\Delta)}. & E(\Delta_\bullet^{(\Delta)}) &= \frac{E(\Delta^2)}{E(\Delta)} - 1
\end{aligned} \tag{6}$$

2.2 The epidemic model

We use an SIR model to investigate the dynamics of the spread of the disease. At any given time point, the population is divided into three groups, depending on health status. The groups are susceptible (**S**), infectious (**I**) and recovered (**R**) (see e.g. Britton (2010)). Individuals of the population make contact with other individuals at (possibly random) points in time. If, at some time point, an infectious individual contacts a susceptible individual then the susceptible individual instantaneously becomes infectious. An infectious individual will cease to be contagious after a period of time, which we call the *infectious period* of the individual in question, and is then transferred to the recovered group. Recovered individuals are those that are immune to the disease. Individuals belonging to this group play no further role in the spread of the disease. Because of this last observation, we can

treat individuals that die because of the disease as "recovered". In summary, we allow only the transitions $S \to I$ and $I \to R$. Note that the population is assumed to be closed; we ignore births, deaths and migration.

More specifically, we consider an SIR epidemic in a generation framework on the clustered graph G_N and assume heterogeneity in infectivity. That is, some infected individuals are more contagious than others. Such heterogeneity may, for instance, arise from variability in the infectious period. To this end, let T be a random variable with support in $[0,1]$, and let $\{T_i\}_{i=1}^N$ be a sequence of independent copies of T. Each node v_i of G_N is equipped with a *transmission weight* T_i. If v_i gets infected, then each susceptible neighbor of v_i gets infected by v_i independently in the next generation with probability T_i (conditioned on $\{T_i\}_i$). Node v_i thereafter becomes recovered, playing no further role in the epidemic. An infected node transmits the disease independently of the transmissions from other infected nodes. An infected node does not, however, transmit the disease to its neighbors independently, unless the distribution of T is degenerate. Conditioned on the transmission weights $\{T_i\}_i$ and the structure of G_N, the number of neighbors that an infected node v_i makes (potentially infectious) contact with while infectious has a binomial distribution with parameters d_i and T_i, where d_i is the degree of v_i.

The spread of this epidemic can be fully captured by a directed graph (see e.g. (Pellis et al. 2012)). To construct such directed graph from an undirected CMC graph G_N, we replace each undirected edge of G_N by two parallel directed edges, pointing in the opposite direction. The weight of an edge (v_i, v_j), which represents the (potential) transmission time from v_i to v_j, is taken to be 1 if v_i would make infectious contact with v_j if infected, and ∞ otherwise. The individuals ultimately infected are then the individuals that can be reached from an initial case by following a path consisting of directed edges with finite edge weights.

2.3 Reproduction numbers

A key quantity in the study of epidemics is the basic reproduction number, often denoted by R_0. It is usually defined as the expected number of infected cases caused by a "typical" infected individual in an otherwise susceptible population. For most stochastic epidemic models (including SIR epidemics in homogeneous mixing propulations (Britton 2010), populations with households (Ball et al. 2016) and epidemics on networks (Britton et al. 2007)) it has the threshold property that a major outbreak is possible if and only if $R_0 > 1$. For models where a suitable generation based branching process approximation is available, R_0 is usually defined as the Perron root (the dominant eigenvalue, which exists and is real-valued by assumptions A1

and A2, see for instance Varga (2009, Chapter 2)) of the mean matrix of the approximating Galton Watson branching process. This is the definition used in this article. By standard branching process theory, the interpretation of R_0 as the expected number of cases caused by the typical individual in the early phase of the epidemic and its threshold properties are retained by this definition. The threshold property of R_0 is made precise in Theorem 1 below.

In Section 4, we investigate the spread of an epidemic in a population with vaccination. To this end, in addition to the basic reproduction number R_0, we consider the *perfect vaccine-associated reproduction number* R_V. A vaccine is *perfect* if it provides full and permanent immunity. That is, an individual vaccinated with a perfect vaccine cannot contract the disease. The perfect vaccine-associated reproduction number R_V is defined as (Ball et al. 2016)

$$R_V = \frac{1}{1 - f_{\mathrm{v}}^{(c)}}, \tag{7}$$

where the *critical vaccination coverage* $f_{\mathrm{v}}^{(c)}$ is the fraction of the population that has to be vaccinated with a perfect vaccine in order to reduce R_0 to unity, if the vaccinated individuals are chosen uniformly at random. That is to say, $f_{\mathrm{v}}^{(c)} = 1 - 1/R_V$ is the fraction necessary to vaccinate in order to be guaranteed to prevent a major outbreak (Britton 2010). Note that if $R_0 \leq 1$ then $f_{\mathrm{v}}^{(c)} = 0$.

For many models, including epidemics on graphs generated by the CM (Britton et al. 2007) and the standard stochastic SIR epidemic model (i.e. individuals mix homogeneously, see for instance Britton (2010)), $R_V = R_0$. That is, vaccinating a fraction $1 - 1/R_0$ of the population with a perfect vaccine is sufficient to surely prevent a major outbreak. On the other hand, for the households and households-workplaces model with uniform vaccination, $R_V \geq R_0$ (Ball et al. 2016) with strict inequality possible. In Section 4.1 we show that for the model analyzed in this report, $R_V = R_0$.

2.3.1 Epidemics in continuous time - the rank based approach

As mentioned above, heterogeneity in infectivity might arise from heterogeneity in the infectious period; an important special case of the above described model is epidemics in continuous time with random infectious periods where contacts between individuals take place according to point processes on $\mathbb{R}_{\geq 0}$. Ignoring the real time-dynamics of an epidemic does not impact results that concern the final outcome of the epidemic. This result was first presented by Ludwig (1975), see also Pellis et al. (2008) for a more recent discussion. This leads us to the often more tractable *rank based* approach.

9

In order to define the rank of a vertex, denote the initial case by v_*. The *rank* of a node v in G_N is the distance from v_* to v, if every edge along which the disease would be transmitted is assigned the edge weight 1, and every other edge is assigned the edge weight ∞. That is, the rank of v is the smallest number of directed edges that have to be traversed in order to follow a path of (potential) transmission from v_* to v. We may then analyze the spread of the disease by letting generation n of the epidemic process consist of the individuals of rank n. If, for instance, v_1 is the first node in a triangle consisting of the nodes v_1, v_2, v_3 to be infected, and v_1 infects v_2 and thereafter attempts to infect v_3, then v_3 is attributed to v_1 regardless of whether v_1 or v_2 infected v_3. This is illustrated in Figure 2.

Consider a continuous time epidemic formulated as follows. Suppose that each infected individual remains infectious for a (random) period of time. The infectious periods are distributed as the random variable τ, $\tau \sim F$, and independent (but identically distributed) for different nodes. Suppose further that a node makes contact with each neighbor independently at a Poisson rate β while infected, and that susceptible individuals are fully susceptible, so that each infectious-susceptible contact results in transmission. Without loss of generality we may assume $\beta = 1$, since we may rescale time (and F accordingly). The transmission weight T is then distributed as $1 - e^{-\tau}$, and $E(T) = 1 - \mathcal{L}(1)$ and $E(T(1-T)) = \mathcal{L}(1) - \mathcal{L}(2)$ where $\mathcal{L}(z) = \int_{\mathbb{R}_+} e^{-zx} dF(x)$ is the Laplace transform of the infectious period.
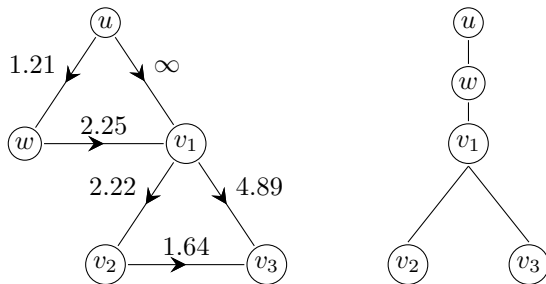

Figure 2: The difference between rank based generations and true generations. Left: The length of the path $v_1 \rightarrow v_3$ (i.e. the transmission time from v_1 to v_3) is 4.89 and exceeds the length $2.22 + 1.64$ of the path $v_1 \rightarrow v_2 \rightarrow v_3$. Therefore, the true path of transmission is $u \rightarrow w \rightarrow v_1 \rightarrow v_2 \rightarrow v_3$. In the rank based approach, however, v_3 is attributed to v_1. Right: The resulting rank generation tree.

2.4 Branching process approximations

To analyze the spread of the disease in the early stages of the epidemic, we employ a multi-type branching process approximation. The graph G_N may be constructed by joining the half-edges in any suitable (possibly random) order. In particular, the graph G_N may be constructed (or explored) as the epidemic progresses; starting with the initial infected case u^* we sequentially match the half-edges along which the disease is transmitted. In the early phase of the epidemic, short cycles (except for the triangles formed by triangle edges) are unlikely to occur. For these reasons, the early spread of the disease is well approximated by a suitably chosen branching process.

Similarly, a branching process approximation can be used to approximate the expected final size of the epidemic (Ball et al. 2009, 2010, 2014). In the graph representation of an epidemic, an individual contracts the disease if and only if there is a path of directed edges with finite edge weights from the initial case to the node representing the individual in question.

Define the susceptibility set $\mathcal{S}(v) = \mathcal{S}_N(v)$ of a node v as the collection of nodes of G_N that can be reached from v by tracing a path of finite length backwards. That is, the individuals that contract the disease are precisely the individuals with susceptibility sets that contain an initial case. Hence, if the initial case is chosen uniformly at random then the probability that a node v contracts the disease is proportional to the size of its susceptibility set $\mathcal{S}(v)$ and this probability can be approximated by exploring $\mathcal{S}(v)$. Figure (3) shows a schematic illustration of a susceptibility set.
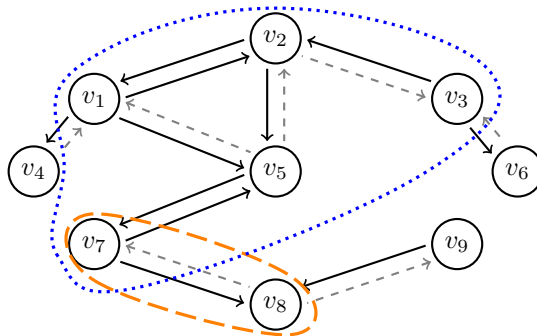

Figure 3: Graph representation of an epidemic in a small ($N = 9$) population. The gray dashed and black solid edges have infinite and finite edge weights (transmission times), respectively. The nodes in the susceptibility set of v_5, $\mathcal{S}(v_5) = \{v_1.v_2, v_3, v_5, v_7\}$, are enclosed by the blue dotted line. The nodes that v_5 would infect if infected, directly or through other nodes, are enclosed by the orange dashed line.

By reversing the direction of the edges of the graph representation of an epidemic, but keeping the weights, the expected final fraction of the population infected in a major outbreak and the probability of a major outbreak are interchanged (Miller 2008), provided that the initial case is chosen uniformly at random. The process so obtained is called the *backward epidemic process* of the node v. If the underlying epidemic model is such that the backward epidemic process can be well approximated by a branching process, then we can use this branching process to compute the asymptotic distribution of the proportion of the population that ultimately escapes infection. This is made precise in the following theorem, due to Ball et al. (2014, Theorem 3.5), who proved the theorem for the related model of random intersection graphs. The statement of Theorem 1 carries over to the forward and backward branching processes considered in this paper. We omit the proof, which is analogous to the proof presented in Ball et al. (2014), see also Ball et al. (2009). Let $\xrightarrow{d}$ denote convergence in distribution.

Theorem 1. Let q and q_b be the extinction probabilities of the forward and backward approximating branching processes respectively, and let S_N be the proportion of the population that ultimately escapes an epidemic in a population of size N. Then

$$S_N \xrightarrow{d} S$$

as $N \to \infty$ where $P(S = 1) = 1 - P(S = q_b) = q$.

In other words, in the limit of large population sizes, the epidemic "takes off" with probability $1 - q$, and if this happens a fraction $1 - q_b$ of the population is ultimately infected (with probability converging to 1 as $N \to \infty$). Note that since R_0 is defined as the Perron root of the mean matrix of the forward branching process, $q < 1$ if and only if $R_0 > 1$.

3 An epidemic in a fully susceptible population

We now have the tools to analyze the spread of an infectious disease on a graph generated by the CMC. In the present section, the population is assumed to be fully susceptible to the disease, apart from the initially infectious individuals.

3.1 Forward process

Before analyzing the forward process, we need to set some terminology. For a given triangle u, v, w, where u is the first individual to be infected in the triangle u, v, w, we refer to v and w as *twins*. We approximate the spread of the disease during the early phase by a multi-type branching process consisting of the following three types (except for the initial case):

Type 1: A node infected along a triangle whose twin is infected at the same time step or earlier

Type 2: A node infected along a triangle edge that is not of type 1

Type 3: A node infected along a single edge

Figure 4 shows three examples of possible paths of transmission within a triangle giving rise to type 1 and 2 individuals in the approximating branching process.
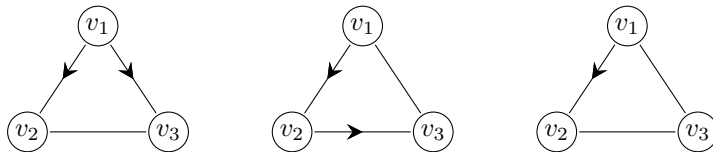

Figure 4: Three examples of possible paths of transmission in a triangle v_1, v_2, v_3, where v_1 is the first node to be infected. Left: v_1 infects both v_2 and v_3. Both v_2 and v_3 are represented by type 1 individuals in the approximating branching process. Center: v_1 infects v_2 and v_2 infects v_3. Then v_3 and v_2 are represented by type 1 and type 2 individuals, respectively. Right: v_1 infects v_2. Then v_2 is represented by a type 2 individual.

Denote by

$$M_f = (m_{ij})_{i,j=1}^3 = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix}$$

the mean matrix of the above described branching process. Suppose that v_1 is the first individual to be infected in the triangle v_1, v_2, v_3. The probability that v_1 transmits the disease both to v_2 and v_3 is $E(T^2)$. Similarly, the probability that v_1 transmits the disease to either v_2 or v_3, but not to both, is $2E(T(1-T))$.

Thus, by linearity of expectation and because the distribution of the susceptible neighbors of infected nodes in the early phase of the epidemic is given by the downshifted degree distributions in (4), we obtain

$$M_f = \begin{pmatrix} 2E(T^2)E(\Delta_\bullet^{(\Delta)}) & 2E(T(1-T))E(\Delta_\bullet^{(\Delta)}) & E(T)E(S_\bullet^{(\Delta)}) \\ 2E(T^2)E(\Delta_\bullet^{(\Delta)}) + E(T) & 2E(T(1-T))E(\Delta_\bullet^{(\Delta)}) & E(T)E(S_\bullet^{(\Delta)}) \\ 2E(T^2)E(\Delta_\bullet^{(S)}) & 2E(T(1-T))E(\Delta_\bullet^{(S)}) & E(T)E(S_\bullet^{(S)}) \end{pmatrix} \cdot$$

(8)

(Recall that the random variables $\Delta_\bullet^{(\Delta)}$, $\Delta_\bullet^{(s)}$, $S_\bullet^{(\Delta)}$ and $S_\bullet^{(s)}$ defined in (5) have the downshifted size biased distributions). Note that all entries of M_f are finite and that S and Δ both have finite second moments by assumption A1.

If M_f is positively regular (see the last paragraph in Section 2.1) then R_0 is given by the Perron root of M_f. With little effort, one can use the expected values provided in (6) to show that necessary and sufficient conditions for M_f to be positively regular are that assumptions A1-A2 hold and that $0 < E(T) < 1$. If some of these conditions are not satisfied, we may analyze the spread of the disease by reducing the number of types of the approximating forward branching process.

3.1.1 Probability of a major outbreak

For two s-dimensional vectors $\bar{a} = (a_1, \ldots, a_s)^\mathsf{T}$ and $\bar{b} = (b_1, \ldots, b_s)^\mathsf{T}$, we define
$$\bar{a}^{\bar{b}} := a_1^{b_1} \cdot \ldots \cdot a_s^{b_s}.$$
Let $f : [0,1]^3 \to \mathbb{R}^3$ be the probability generating function of the offspring distribution of the three types in the approximating branching process. That is, for $\bar{z} = (z_1, z_2, z_3)^\mathsf{T} \in [0,1]^3$ the ith component of $f(\bar{z})$ is given by

$$f(\bar{z})_i = E\left(\bar{z}^{\bar{\xi}_i}\right) \tag{9}$$

where $\bar{\xi}_i = (\xi_{i,1}, \xi_{i,2}, \xi_{i,3})$ is distributed as the offspring of a type i individual, $i = 1, 2, 3$.

Similarly, let $f_* : [0,1]^3 \to \mathbb{R}$ be the probability generating function of the offspring distribution of the initial case. If $\bar{\xi} = (\xi_{*,1}, \xi_{*,2}, \xi_{*,3})^\mathsf{T}$ is distributed as the offspring of the initial case, then f_* is given by

$$f_*(\bar{z}) = E\left(\bar{z}^{\bar{\xi}}\right).$$

For $i = 1, 2, 3$, let $(S^{(i)}, \Delta^{(i)})$ be the joint degree of a type i case with offspring $(\xi_{i,1}, \xi_{i,2}, \xi_{i,3})$ and transmission weight T. That is,

$$(S^{(1)}, \Delta^{(1)}) \stackrel{d}{=} (S^{(2)}, \Delta^{(2)}) \stackrel{d}{=} (S^{(\Delta)}, \Delta^{(\Delta)})$$

and

$$(S^{(3)}, \Delta^{(3)}) \stackrel{d}{=} (S^{(s)}, \Delta^{(s)}).$$

Here $\stackrel{d}{=}$ denotes equality in distribution. By conditional independence we have

$$E(z_1^{\xi_{i,1}} z_2^{\xi_{i,2}} z_3^{\xi_{i,3}}) = E\left(E(z_3^{\xi_{i,3}} | T, S^{(i)}, \Delta^{(i)}) E(z_1^{\xi_{i,1}} z_2^{\xi_{i,2}} | T, S^{(i)}, \Delta^{(i)}) \right).$$

Conditioned on the transmission weight T and the single degree $S^{(1)}$, $\xi_{1,3}$ has a binomial distribution with parameters $S^{(1)}$ and T. Thus

$$E(z_3^{\xi_{1,3}} | T, S^{(1)}, \Delta^{(1)}) = \sum_{k_0 + k_1 = S^{(1)}} \binom{S^{(1)}}{k_1} (Tz_3)^{k_1} (1 - T)^{k_0}$$

$$= (Tz_3 + 1 - T)^{S^{(1)}}.$$

Similarly

$$E(z_1^{\xi_{1,1}} z_2^{\xi_{1,2}} | T, S^{(1)}, \Delta^{(1)})$$

$$= \sum_{k_0 + k_1 + k_2 = \Delta^{(1)} - 1} \binom{\Delta^{(1)} - 1}{k_0, k_1, k_2} (1 - T)^{2k_0} (2(1 - T)Tz_2)^{k_1} (Tz_1)^{2k_2}$$

$$= ((1 - T)^2 + 2T(1 - T)z_2 + T^2 z_1^2)^{\Delta^{(1)} - 1}.$$

Thus

$$E(z_1^{\xi_{1,1}} z_2^{\xi_{1,2}} z_3^{\xi_{1,3}})$$

$$= E((Tz_3 + 1 - T)^{S_\bullet^{(\Delta)}} ((1 - T)^2 + 2T(1 - T)z_2 + T^2 z_1^2)^{\Delta_\bullet^{(\Delta)}}) \qquad (10)$$

where $(\Delta_\bullet^{(\Delta)}, S_\bullet^{(\Delta)})$ is independent of T.

Since the conditional offspring distribution of a type 2 individual is identical to the offspring distribution of a type 1 individual except that a type 2 individual may give birth to one additional type 1 individual with probability T, we have

$$E(z_1^{\xi_{2,1}} z_2^{\xi_{2,2}} z_3^{\xi_{2,3}})$$

$$= E((Tz_3 + 1 - T)^{S_\bullet^{(\Delta)}} ((1 - T)^2 + 2T(1 - T)z_2 + T^2 z_1^2)^{\Delta_\bullet^{(\Delta)}} (Tz_1 + 1 - T)). \qquad (11)$$

Similarly,

$$
\begin{aligned}
E&(z_1^{\xi_{3,1}} z_2^{\xi_{3,2}} z_3^{\xi_{3,3}})\\
&= E((Tz_3 + 1 - T)^{S_\bullet^{(s)}}((1-T)^2 + 2T(1-T)z_2 + T^2 z_1^2)^{\Delta_\bullet^{(s)}}).
\end{aligned}
\tag{12}
$$

Substituting (10)-(12) into (9) gives an expression for f.

By standard branching process theory, if $R_0 > 0$ the extinction probability of a process descending from a type i individual, $i = 1, 2, 3$, is given by q_i, where $\bar{q} = (q_1, q_2, q_3)^{\mathsf{T}}$ is the unique solution of $\bar{q} = f(\bar{q})$ in $[0,1)^3$. We also have

$$
\bar{q} = \lim_{n \to \infty} f^{\circ n}(\bar{0}),
\tag{13}
$$

where $f^{\circ n}$ is the composition of f with itself n times.

Since the approximating branching process dies out if and only if each of the processes started by the children of the initial case die out, the probability of extinction is given by $f^*(\bar{q})$. After some calculations, analogous to the calculations that led to (10)-(12), we find that the probability of extinction is given by

$$
f_*(\bar{q}) = E\left((Tq_3 + 1 - T)^S((1-T)^2 + 2T(1-T)q_2 + T^2 q_1^2)^\Delta\right)
$$

where (S, Δ) is independent of T. We conclude that, by Theorem 1, the probability of a major outbreak is given by $1 - f^*(\bar{q})$, where $\bar{q}$ is the limit in (13).

3.2 Backward process

Let w be a given node of G_N, chosen uniformly at random. We use a backward branching process to approximate the probability that w contracts the disease, which by an exchangeability argument equals the expected final size of a major outbreak. The offspring of an individual v in the backward process are the individuals that would potentially have infected v, if they were infected themselves.

The members of the susceptibility set are divided into the following two groups. This gives rise to a two-type approximating backward branching process.

Type 1: The vertex is included in the susceptibility set by virtue of potential transmission along a single edge

Type 2: The vertex is included in the susceptibility set by virtue of potential transmission along a triangle edge

We assign kinship as follows. The children of type 1 of an individual v_1 are the individuals included in the susceptibility set due to potential transmission along a single edge. The children of type 2 of v_1 are the individuals included in the susceptibility set due to potential transmission of the disease to v_1, within a triangle of which v_1 is a member. We note that, given a triangle v_1, v_2, v_3 where v_1 is the primary case, both v_2 and v_3 will be members of the susceptibility set of v_1 by virtue of transmissions within the triangle if at least one of the following events happens:

E_1) v_2 and v_3 both "infects" v_1

E_2) v_2 infects v_1 and v_3 "infects" v_2

E_3) v_3 infects v_1 and v_2 "infects" v_3

Here "infects" is conditional on the "infector" being infected during the epidemic.

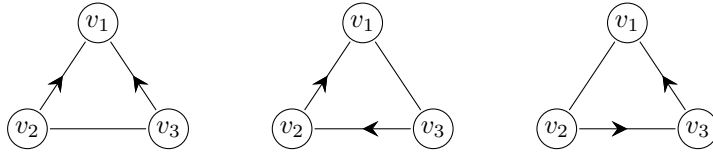The events E_1-E_3 are illustrated in Figure (5).

Figure 5: The individuals v_2 and v_3 are both in the susceptibility set $\mathfrak{S}(v_1)$ of v_1 by virtue of transmission within the triangle v_1, v_2, v_3 if and only if at least one of the events E_1 (left), E_2 (center) or E_3 (right) happens.

Standard calculations give that the probability of the union of the events E_1-E_3 is given by $p_2 = 3E(T)^2 - 2E(T)E(T^2)$. Similarly, the probability that neither v_1 nor v_2 will be members of the susceptibility set of v by transmissions within the triangle is given by $p_0 = (1 - E(T))^2$. For later use, denote $1 - p_0 - p_2$ by p_1.

3.2.1 Expected final size of a major outbreak

Let b be the probability generating function of the offspring distribution of the two types of the approximating backward branching process. Furthermore, let b_* be the probability generating function of the offspring distribution of the ancestor w. Analogously to the forward branching process, the probability that the bloodline started by a type i, $i = 1, 2$, individual will become extinct is given by q_i^b, where $\bar{q}_b = (q_1^b, q_2^b)^\mathsf{T}$ is the unique solution of $\bar{q}_b = b(\bar{q}_b)$ in $[0, 1)^2$ (recall $R_0 > 1$). The probability of extinction is given by $b_*(\bar{q}_b)$.

Proceeding in the same manner as in Section 3.1.1 yields

$$b(z_1, z_2)_1 = E\left((E(T)z_1 + 1 - E(T))^{S_\bullet^{(s)}}(p_0 + p_1z_2 + p_2z_2^2)^{\Delta_\bullet^{(s)}}\right)$$

where p_0, p_1 and p_2 are as in Section 3.2. Similarly

$$b(z_1, z_2)_2 = E\left((E(T)z_1 + 1 - E(T))^{S_\bullet^{(\Delta)}}(p_0 + p_1z_2 + p_2z_2^2)^{\Delta_\bullet^{(\Delta)}}\right),$$

and the probability of ultimate extinction of the backward process is given by

$$b_*(\bar{q}_b) = E\left((E(T)q_1^b + 1 - E(T))^S(p_0 + p_1q_2^b + p_2(q_2^b)^2)^\Delta\right).$$

We conclude that the expected final size of a major outbreak is given by $1 - b_*(\bar{q}_b)$.

4 Vaccination

4.1 Random vaccination with a perfect vaccine

Assume that a fraction $f_v < 1$ of the population is vaccinated, and that the vaccinated individuals are chosen uniformly at random (without replacement) from the population. The vaccine is perfect, in the sense that a vaccinated individual gains full and lasting immunity to the disease. If the population size N is large, we may use a slightly different model, where each individual is vaccinated with probability f_v, independently of the vaccination status of other individuals. By the law of large numbers, for our purposes the models are equivalent in the limit as the population size $N \to \infty$.

As before, we may approximate the early phase of the epidemic by a multi-type branching process. The individuals of the approximating branching process are now of the following three types.

Type 1: Infected along a triangle edge and has a twin that is known not to be susceptible

Type 2: Infected along a triangle edge and has a twin that might be susceptible

Type 3: Infected along a single edge

To clarify the types, assume that in the early phase of the epidemic v_1 is the primary case in the triangle v_1, v_2, v_3. If v_1 attempts to transmit the disease both to v_2 and v_3 and succeeds (that is, none of v_2 and v_3 are vaccinated) then both v_2 and v_3 are represented by type 1 individuals in the approximating branching process. This happens with probability

$$E(T^2)(1 - f_{\mathrm{v}})^2. \tag{14}$$

If v_1 attempts to transmit the disease both to v_2 and v_3, but only succeeds to transmit the disease to v_3 (that is, v_2 is vaccinated and v_3 is not vaccinated), then in the approximating branching process the individual representing v_1 gives birth to one type 1 individual (representing v_3) within the triangle v_1, v_2, v_3. This happens with probability

$$E(T^2) f_{\mathrm{v}} (1 - f_{\mathrm{v}}). \tag{15}$$

If v_1 attempts to transmit the disease only to v_2 and succeeds (that is, v_2 is not vaccinated) then in the approximating branching process, the individual representing v_1 gives birth to one type 2 individual (representing v_2) within the triangle v_1, v_2, v_3. This happens with probability

$$E(T(1 - T))(1 - f_{\mathrm{v}}). \tag{16}$$
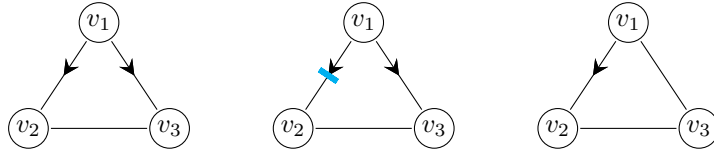
The above described events are illustrated in Figure 6.

Figure 6: Three examples of transmission dynamics within a triangle v_1, v_2, v_3. An attempted transmission of the disease is represented by an arrow, an attempted transmission to a vaccinated individual is represented by an arrow and a blue bar. Left: v_1 attempts to transmit the disease both to v_2 and v_3, and succeeds. Both v_2 and v_3 are represented by type 1 individuals in the approximating branching process. Center: v_1 attempts to transmit the disease both to v_2 and v_3, the transmission to v_2 is blocked since v_2 is vaccinated. Then v_3 is represented by a type 1 individual. Right: v_1 succeeds to transmit the disease to v_2, but does not attempt to infect v_3. Then v_2 is represented by a type 2 individual.

Denote the mean matrix of the approximating branching process by $M_f^{(\mathrm{v})} = (m_{i,j}^{(\mathrm{v})})_{i,j=1}^3$. Using the expressions in (14) and (15) gives the expected number of type 1 individuals produced by a type 1 individual

$$m_{1,1}^{(v)} = \left(2(1-f_v)^2 E(T^2) + 2(1-f_v)f_v E(T^2)\right) E\left(\Delta_\bullet^{(\Delta)}\right)$$
$$= (1-f_v)2E(T^2)E(\Delta_\bullet^{(\Delta)})$$
$$= (1-f_v)m_{1,1}$$

where $m_{1,1}$ is an element of the mean matrix M_f of the forward branching process presented in (8).

Proceeding in the same fashion, we obtain the elements of the mean matrix $M_f^{(v)} = (m_{i,j}^{(v)})_{i,j=1}^3$ of the branching process with random vaccination. It turns out that

$$M_f^{(v)} = (1-f_v)M_f.$$

It is readily verified that the Perron root of $M_f^{(v)}$ is

$$r_f^{(v)} = (1-f_v)r_f, \tag{17}$$

where r_f is the Perron root of M_f. Setting $r_f^{(v)}$ to 1 in (17) and solving for f_v yields the critical vaccination coverage $f_v^{(c)} = 1 - 1/r_f$.

We conclude that, for this particular graph model, equality holds between the basic reproduction number R_0 and the perfect vaccine-associated reproduction number R_V as defined in (7).

4.1.1 Probability of a major outbreak

Let h be the probability generating function of the offspring distribution of the three types in our model including vaccination. As in Section 3.1.1, we use the probability generating function to approximate the probability of extinction of the epidemic. To this end, let $(\zeta_{i,1}, \zeta_{i,2}, \zeta_{i,3})$ be distributed as the offspring of a type i individual with transmission weight T, $i = 1, 2, 3$, and let $(S^{(i)}, \Delta^{(i)})$ be distributed as the joint degree of this individual. That is,

$$(S^{(1)}, \Delta^{(1)}) \overset{d}{=} (S^{(2)}, \Delta^{(2)}) \overset{d}{=} (S_\circ^{(\Delta)}, \Delta_\circ^{(\Delta)})$$

and

$$(S^{(3)}, \Delta^{(3)}) \overset{d}{=} (S_\circ^{(s)}, \Delta_\circ^{(s)}).$$

Note that $(S^{(i)}, \Delta^{(i)})$ and T are independent.

By conditional independence

$$E\left(z_1^{\zeta_{1,1}} z_2^{\zeta_{1,2}} z_3^{\zeta_{1,3}}\right) = E\left(E\left(z_3^{\zeta_{1,3}} | S^{(1)}, \Delta^{(1)}, T\right) E\left(z_1^{\zeta_{1,1}} z_2^{\zeta_{1,2}} | S^{(1)}, \Delta^{(1)}, T\right)\right)$$

for $\bar{z} = (z_1, z_2, z_3)^\mathsf{T} \in [0, 1]^3$.

Conditioned on the transmission weight T and the joint degree $(S^{(1)}, \Delta^{(1)})$, the number of attempted transmissions from a type 1 individual along single edges has a binomial distribution with parameters $S^{(1)}$ and T, and each attempted transmission succeeds with probability $(1 - f_\mathrm{v})$. Thus,

$$
\begin{aligned}
E\left(z_3^{\zeta_{1,3}} | S^{(1)}, \Delta^{(1)}, T\right) &= \sum_{k_0 + k_1 = S^{(1)}} \binom{S^{(1)}}{k_1} z_3^{k_1} \left(T(1 - f_\mathrm{v})\right)^{k_1} \left((1 - T) + T f_\mathrm{v}\right)^{k_0} \\
&= \left(T(1 - f_\mathrm{v}) z_3 + 1 - T + T f_\mathrm{v}\right)^{S^{(1)}}.
\end{aligned}
\tag{18}
$$

Similarly, for a type 1 individual w with triangle degree $\Delta^{(1)}$, by conditioning on the number of attempted transmissions (in k_i of the $\Delta^{(1)} - 1$ triangles that are not yet affected by the disease, w attempts to transmit the disease to i individuals, $i = 0, 1, 2$) and the vaccination status of the individuals contacted by w we obtain

$$E(z_1^{\zeta_{1,1}} z_2^{\zeta_{1,2}} | S^{(1)}, \Delta^{(1)}, T)$$

$$= \sum_{k_0+k_1+k_2=\Delta^{(1)}-1} \binom{\Delta^{(1)}-1}{k_0, k_1, k_2} (1-T)^{2k_0} \big(2T(1-T)\big)^{k_1} T^{2k_2}$$

$$\left(\sum_{\tilde{k}_0+\tilde{k}_1+\tilde{k}_2=k_2} \binom{k_2}{\tilde{k}_0, \tilde{k}_1, \tilde{k}_2} \big((1-f_{\rm v})z_1\big)^{2\tilde{k}_2} \big(2f_{\rm v}(1-f_{\rm v})z_1\big)^{\tilde{k}_1} f_{\rm v}^{2\tilde{k}_0} \right)$$

$$\left(\sum_{k_0'+k_1'=k_1} \binom{k_1}{k_0', k_1'} (1-f_{\rm v})^{k_1'} z_2^{k_1'} f_{\rm v}^{k_0'} \right)$$

$$= \sum_{k_0+k_1+k_2=\Delta^{(1)}-1} \binom{\Delta^{(1)}-1}{k_0, k_1, k_2} (1-T)^{2k_0} \big(2T(1-T)\big)^{k_1} T^{2k_2}$$

$$\Big(\big((1-f_{\rm v})z_1\big)^2 + 2f_{\rm v}(1-f_{\rm v})z_1 + f_{\rm v}^2 \Big)^{k_2}$$

$$\Big((1-f_{\rm v})z_2 + f_{\rm v} \Big)^{k_1}$$

$$= \Big[(1-T)^2 + 2T(1-T)\big[(1-f_{\rm v})z_2 + f_{\rm v}\big]$$

$$+ T^2 \big[\big((1-f_{\rm v})z_1\big)^2 + 2f_{\rm v}(1-f_{\rm v})z_1 + f_{\rm v}^2 \big] \Big]^{\Delta^{(1)}-1}.$$

$$(19)$$

Combining (18) and (19) yields

$$E\left(z_1^{\zeta_{1,1}} z_2^{\zeta_{1,2}} z_3^{\zeta_{1,3}} \right) = E\Bigg[\Big(T(1-f_{\rm v})z_3 + 1 - T + Tf_{\rm v} \Big)^{S_\bullet^{(\Delta)}}$$

$$\Big((1-T)^2 + 2T(1-T)\big((1-f_{\rm v})z_2 + f_{\rm v}\big)$$

$$+ T^2 \big(\big((1-f_{\rm v})z_1\big)^2 + 2f_{\rm v}(1-f_{\rm v})z_1 + f_{\rm v}^2 \big) \Big)^{\Delta_\bullet^{(\Delta)}} \Bigg].$$

$$(20)$$

By noting that the offspring distribution of a type 2 individual is identical to the offspring distribution of a type 1 individual, except that a type 2 may

give birth to one additional type 1 individual with probability $T(1 - f_{\mathrm{v}})$ we obtain

$$
\begin{aligned}
E\left(z_1^{\zeta_{2,1}} z_2^{\zeta_{2,2}} z_3^{\zeta_{2,3}}\right) = E\Bigg[& \left(T(1 - f_{\mathrm{v}})z_3 + 1 - T + Tf_{\mathrm{v}}\right)^{S_\bullet^{(\Delta)}} \\
& \left((1 - T)^2 + 2T(1 - T)\left((1 - f_{\mathrm{v}})z_2 + f_{\mathrm{v}}\right)\right. \\
& \left.+ T^2\left(\left((1 - f_{\mathrm{v}})z_1\right)^2 + 2f_{\mathrm{v}}(1 - f_{\mathrm{v}})z_1 + f_{\mathrm{v}}^2\right)\right)^{\Delta_\bullet^{(\Delta)}} \\
& \left(z_1 T(1 - f_{\mathrm{v}}) + 1 - T(1 - f_{\mathrm{v}})\right)\Bigg].
\end{aligned}
$$

(21)

Similarly,

$$
\begin{aligned}
E\left(z_1^{\zeta_{3,1}} z_2^{\zeta_{3,2}} z_3^{\zeta_{3,3}}\right) = E\Bigg[& \left(T(1 - f_{\mathrm{v}})z_3 + 1 - T + Tf_{\mathrm{v}}\right)^{S_\bullet^{(s)}} \\
& \left((1 - T)^2 + 2T(1 - T)\left((1 - f_{\mathrm{v}})z_2 + f_{\mathrm{v}}\right)\right. \\
& \left.+ T^2\left(\left((1 - f_{\mathrm{v}})z_1\right)^2 + 2f_{\mathrm{v}}(1 - f_{\mathrm{v}})z_1 + f_{\mathrm{v}}^2\right)\right)^{\Delta_\bullet^{(s)}}\Bigg].
\end{aligned}
$$

(22)

Combining these results yields the probability generating function h of the offspring distribution of a type 1, 2, 3 individual respectively. That is, $h(\bar{z})_1$ is given by (20), $h(\bar{z})_2$ is given by (21) and $h(\bar{z})_3$ is given by (22).

The probability generating function h^* of the initial case is given by

$$
\begin{aligned}
h^*(\bar{z}) = & E(z_1^{\zeta_{*,1}} z_2^{\zeta_{*,2}} z_3^{\zeta_{*,3}}) \\
= & E\Bigg[\left(T(1 - f_{\mathrm{v}})z_3 + 1 - T + Tf_{\mathrm{v}}\right)^{S} \\
& \left((1 - T)^2 + 2T(1 - T)\left((1 - f_{\mathrm{v}})z_2 + f_{\mathrm{v}}\right)\right. \\
& \left.+ T^2\left(\left((1 - f_{\mathrm{v}})z_1\right)^2 + 2f_{\mathrm{v}}(1 - f_{\mathrm{v}})z_1 + f_{\mathrm{v}}^2\right)\right)^{\Delta}\Bigg].
\end{aligned}
$$

(23)

for $\bar{z} = (z_1, z_2, z_3)^{\mathsf{T}} \in [0, 1]^3$, where (S, Δ) is distributed as the joint degree of the initial case and independent of T. The probability of extinction of the approximating branching process is given by $h^*(\bar{q}^{(\mathrm{v})})$, where $\bar{q}^{(\mathrm{v})}$ is given by the point in $[0, 1]^3$ closest to the origin that satisfies $\bar{q}^{(\mathrm{v})} = h(\bar{q}^{(\mathrm{v})})$. Thus, by Theorem 1 the probability of a major outbreak is $1 - h_*(\bar{q}^{(\mathrm{v})})$.

4.1.2 The backward process

We now turn our attention to the backward process and final size of an epidemic in a population where a fraction f_v is vaccinated with a perfect vaccine. To this end, we introduce the following three types, where individuals are classified by their vaccination status and the type of the edge along which they would transmit the disease if infected.

Type 1: Transmits along triangle edge, no information on vaccination status is available

Type 2: Transmits along triangle edge and is known not to be vaccinated since it is successfully infected by its twin

Type 3: Transmits along single edge, no information on vaccination status is available

To clarify the types a bit more, let v_1, v_2, v_3 be a given triangle. At least one of v_2 and v_3 belongs to the susceptibility set of v_1 by virtue of potential transmissions within the triangle if some the following events, illustrated in Figure 7, happens. Note that all cases infected by virtue of transmission within the triangle v_1, v_2, v_3 are attributed to v_1.

E_1) v_2 attempts to infect v_1 and v_3 attempts to infect v_2, both succeed, and v_3 does not attempt to infect v_1. Or the same thing might happen, with v_2 and v_3 interchanged. This results in one type 1 and one type 2 individual in the approximating branching process. If v_1 is represented by a type 1 or 3 individual this happens with probability
$$2\big(1 - f_\mathrm{v}\big)^2 E(T) E\big(T(1 - T)\big),$$
if v_1 is represented by an individual of type 2 this happens with probability
$$2(1 - f_\mathrm{v}) E(T) E\big(T(1 - T)\big).$$

E_2) Only one of v_2 and v_3 attempts to infect v_1, and succeeds. The other node does not attempt to infect any node within the triangle. This results in one type 1 offspring. If v_1 is represented by an individual of type 1 or 3 this happens with probability
$$2(1 - f_\mathrm{v}) E(T) E\big(T(1 - T)\big),$$

24

if v_1 is represented by an individual of type 2 this happens with probability

$$2E(T)E\big(T(1-T)\big).$$

$E_3)$ v_2 and v_3 both attempt to infect v_1 and succeeds. This results in two type 1 individuals born in the approximating branching process. If v_1 is represented by an individual of type 1 or 3 this happens with probability

$$(1-f_\mathrm{v})E(T^2),$$

if v_1 is represented by an individual of type 2 this happens with probability

$$E(T^2).$$

$E_4)$ v_2 attempts to infect v_1 and succeeds. The other node, v_3, attempts to infect v_2, but fails due to v_2 being vaccinated. The individual v_3 does not attempt to infect v_1. In this scenario, v_2 belongs to the susceptibility set of v_1. However, we do not include v_2 is the approximating branching process. This does not have any impact on the result of our analysis, since we are only interested in the probability of extinction of the backward process and v_2 does not produce any offspring in this process.
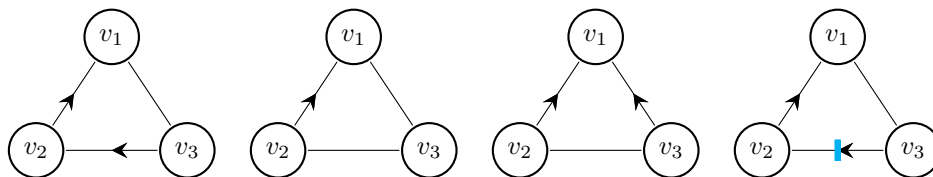

Figure 7: At least one of v_2 and v_3 will belong to the susceptibility set of v_1 by virtue of potential transmissions within the triangle if some of the following types of scenarios (left to right in the picture) occur: E_1, E_2, E_3, E_4. An attempted transmission of the disease is represented by an arrow, an attempted transmission to a vaccinated individual is represented by an arrow and a blue bar.

4.1.3 Expected final size

Let $b^{(v)}$ and $b_*^{(v)}$ be the probability generating function of the offspring distribution of the three types of the approximating backward branching process and of the ancestor, respectively. Furthermore, let $\bar{\zeta}_i = (\zeta_{i,1}^b, \zeta_{i,2}^b, \zeta_{i,3}^b)$ be distributed as the offspring of a type i, $i = 1, 2, 3$, individual and denote by E_s the conditional expectation given that the parent of $\zeta_{i,1}^b, \zeta_{i,2}^b, \zeta_{i,3}^b$ is susceptible. Let further $\bar{\zeta}_* = (\zeta_{*,1}^b, \zeta_{*,2}^b, \zeta_{*,3}^b)$ be distributed as the offspring of the ancestor. Denote the extinction probability of a process descending from a type i individual by q_i^b, $i = 1, 2, 3$ and let $\bar{q}^b = (q_1^b, q_2^b, q_3^b)^\mathsf{T}$.

To find an expression for $b^{(v)}$, we first note that for $\bar{z} = (z_1, z_2, z_3)^\mathsf{T}$

$$
E\left(\bar{z}^{\bar{\zeta}_3}\right) = f_v + (1 - f_v)E_s\left(E_s\left(z_3^{\zeta_{3,3}^b}|S^{(3)}, \Delta^{(3)}\right) E_s\left(z_1^{\zeta_{3,1}^b} z_2^{\zeta_{3,2}^b}|S^{(3)}, \Delta^{(3)}\right)\right)
\tag{24}
$$

where, as before, $(S^{(i)}, \Delta^{(i)})$ is distributed as the joint degree of a type i individual, $i = 1, 2, 3$.

Now

$$
\begin{aligned}
E_s\left(z_3^{\zeta_{3,3}^b}|S^{(3)}, \Delta^{(3)}\right) &= \sum_{k_0+k_1=S^{(3)}-1} \binom{S^{(3)} - 1}{k_0, k_1} z_3^{k_1} E(T)^{k_1} E(1 - T)^{k_0} \\
&= \left(E(T)z_3 + 1 - E(T)\right)^{S^{(3)}-1}.
\end{aligned}
\tag{25}
$$

By conditioning on the number of triangles k_2 in which an event of type E_3 occurs, the number of triangles k_1^a in which an event of type E_1 occurs, the number of triangles k_1^b in which an event of type E_4 occurs and the number

of triangles k_1^c in which an event of type E_2 occurs we obtain

$$E_s(z_1^{\zeta_{3,1}} z_2^{\zeta_{3,2}} | S^{(3)}, \Delta^{(3)})$$

$$= \sum_{k_0 + k_1^a + k_1^b + k_1^c + k_2 = \Delta^{(3)}} \binom{\Delta^{(3)}}{k_0, k_1^a, k_1^b, k_1^c, k_2} E(1-T)^{2k_0}$$

$$\left(2E(T)E\big(T(1-T)\big)(1-f_{\mathrm{v}}) \right)^{k_1^a}$$

$$\left(2E(T)E\big(T(1-T)\big)f_{\mathrm{v}} \right)^{k_1^b} \left(2E(T)E\big((1-T)^2\big) \right)^{k_1^c}$$

$$E(T)^{2k_2} z_2^{k_1^a} z_1^{k_1^a + k_1^c + 2k_2}$$

$$= \Big(\big(E(1-T)\big)^2 + 2E\big(T\big)E\big(T(1-T)\big)(1-f_{\mathrm{v}})z_2 z_1 + 2E(T)E\big(T(1-T)\big)f_{\mathrm{v}}$$

$$+ 2E(T)E\big((1-T)^2\big)z_1 + E(T)^2 z_1^2 \Big)^{\Delta^{(3)}}.$$

$$(26)$$

Inserting the right hand sides of (25) and (26) in (24) gives

$$E(z_1^{\zeta_{3,1}} z_2^{\zeta_{3,2}} z_3^{\zeta_{3,3}})$$

$$= f_{\mathrm{v}} + (1-f_{\mathrm{v}})E\left[\Big(E(T)z_3 + 1 - E(T) \Big)^{S_\bullet^{(s)}} \right.$$

$$\Big(\big(E(1-T)\big)^2 + 2E(T)E\big(T(1-T)\big)(1-f_{\mathrm{v}})z_1 z_2$$

$$+ 2E(T)E\big(T(1-T)\big)f_{\mathrm{v}}$$

$$\left. + 2E(T)E\big((1-T)^2\big)z_1 + E(T)^2 z_1^2 \Big)^{\Delta_\bullet^{(s)}} \right].$$

$$(27)$$

27

Similarly

$$E(z_1^{\zeta_{2,1}} z_2^{\zeta_{2,2}} z_3^{\zeta_{2,3}}) = E\Bigg[\bigg(E(T)z_3 + 1 - E(T)\bigg)^{S_\bullet^{(\Delta)}}$$

$$\bigg(\Big(E(1-T)\Big)^2 + 2E(T)E\big(T(1-T)\big)(1-f_{\mathrm{v}})z_1 z_2$$

$$+ 2E(T)E\big(T(1-T)\big)f_{\mathrm{v}}$$

$$+ 2E(T)E\big((1-T)^2\big)z_1 + E(T)^2 z_1^2\bigg)^{\Delta_\bullet^{(\Delta)}}\Bigg].$$
$$(28)$$

and

$$E(z_1^{\zeta_{1,1}} z_2^{\zeta_{1,2}} z_3^{\zeta_{1,3}}) = f_{\mathrm{v}} + (1 - f_{\mathrm{v}})E(z_1^{\zeta_{2,1}} z_2^{\zeta_{2,2}} z_3^{\zeta_{2,3}}). \qquad (29)$$

Combining these results yields the probability generating function of the offspring distribution of the three types; $b^{(\mathrm{v})}(\bar{z})_3$ is given by (27) and $b^{(\mathrm{v})}(\bar{z})_2$ is given by (28). By replacing $(S_\bullet^{(s)}, \Delta_\bullet^{(s)})$ in the right hand side of (27) by $(S_\bullet^{(\Delta)}, \Delta_\bullet^{(\Delta)})$ we obtain $b^{(\mathrm{v})}(\bar{z})_1$.

Also by replacing $(S_\bullet^{(s)}, \Delta_\bullet^{(s)})$ in the right hand side of (27), but now by (S, Δ) we obtain the probability generating function $b_*^{(\mathrm{v})}(\bar{z})$ of the offspring of the initial case. The expected final size of the epidemic, conditioned on that a major outbreak occurs, is given by

$$1 - b_*^{(\mathrm{v})}(\bar{q}^b).$$

5   Numerical example

Under very general assumptions, increasing the heterogeneity in infectiousness leads to a decrease in the the probability of a major outbreak, the expected final size and R_0 (Kuulasmaa 1982; Meester and Trapman 2011; Miller 2008), see also Ball (1985); Kenah and Robins (2007); Miller (2007). In particular, for a fixed (marginal) transmission probability $E(T)$, the probability of a major outbreak and the expected final size are maximized if $T = E(T)$ with probability 1 and minimized if $P(T = 1) = E(T) = 1 - P(T = 0)$. Similarly, for given $E(T)$, R_0 is maximized if $T = E(T)$ with probability 1 and minimized if $P(T = 1) = E(T) = 1 - P(T = 0)$.

We illustrate this with the following example. Consider the three degree distributions

1. $p(2,1) = 1$

2. $p(4,0) = 0.95 = 1 - p(2,1)$

3. $p(0,2) = 0.95 = 1 - p(2,1)$.

That is, in all three degree distributions the total degree is 4 with probability 1. In addition, distribution 1 corresponds to a network where every node is member of exactly one triangle. Distribution 2 corresponds to a network where a node is not a member of any triangle with probability 0.95, while with probability 0.05 a node is member of one triangle. Finally, distribution 3 corresponds to a network where a node is a member of two triangles with probability 0.95, while with probability 0.05 a node is member of one triangle.

Furthermore, let T have distribution $\text{Beta}(\alpha, \alpha)$ for some $\alpha > 0$. That is, T has density, $C_\alpha x^{\alpha-1}(1-x)^{\alpha-1}$, on the interval $(0,1)$, where C_α is a normalizing constant. Then $E(T) = 1/2$ and we can tune the heterogeneity of the infectivity of infected individuals by varying α. In particular

$$E(T^2) = \frac{1}{2}\left(1 - \frac{1}{2 + \alpha^{-1}}\right).$$

Note that $\alpha \to \infty$ corresponds to T being uniform on $(0,1)$, while $\alpha = 0$ corresponds to $P(T=0) = P(T=1) = 1/2$. Figure 8 shows the probability that a major outbreak does not occur, the expected final size, R_0 and the critical vaccination coverage $f_v^{(c)}$ as functions of α or $E(T^2)$.

As can be seen in Figure 8, ignoring actual heterogeneity of infectivity in this case leads to an overestimation of the probability of a major outbreak (8a-8b). This effect is particularly evident in the presence of high clustering; the steeper slope of the curve corresponding to distribution 3 (8b) and the relatively low probability of a major outbreak when α is small can be explained by the fact that the approximating forward branching process is close to being critical when α is small. Figure 8c-8d shows that heterogeneity of infectivity has virtually no impact on the expected final size of a major outbreak and R_0 in the near absence of clustering. In the presence of clustering, on the other hand, ignoring heterogeneity of infectivity leads to an underestimation of the expected final size and a substantial overestimation of the critical vaccination coverage $f_v^{(c)}$. Note that R_0 and $f_v^{(c)}$ depend on the distribution of T only through the first and second moment of T.
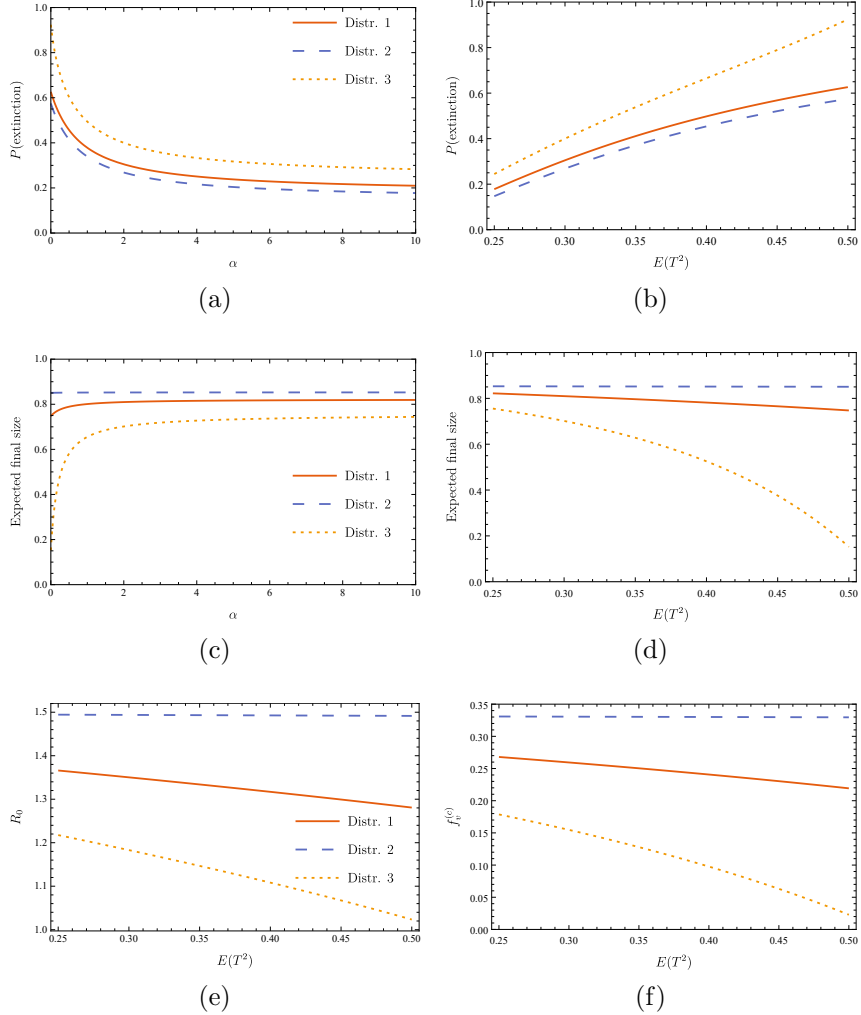
Figure 8: The impact of heterogeneity in infectivity for the three degree distributions. (a) The probability that a major outbreak does not occur as a function of α. (b) The probability that a major outbreak does not occur as a function of $E(T^2)$. (c) The expected final size of a major outbreak as a function of α. (d) The expected final size of a major outbreak as a function of $E(T^2)$. (e) The basic reproduction number R_0 as a function of $E(T^2)$. (f)The critical vaccination coverage $f_v^{(c)}$ as a function of $E(T^2)$.

6   Discussion

In this paper, we have incorporated clustering in the spread of an infectious disease by allowing for groups of size three with non-overlapping edges. It

is, in principle, straightforward to extend the methods used in this paper to larger group sizes. The CMC may, for instance, be generalized to larger group sizes as follows. Let $K = \{k_1, \ldots, k_r\} \subset \mathbb{N}_{\geq 2}$ be the set of possible group sizes. In the matching procedure, each node is equipped with an r-dimensional degree in $\mathbb{N}_0^r$. The ith component (the k_i-degree) of a degree specifies the number of groups of size k_i to which the node in question belongs. Analogously to the construction of a CMC graph, groups are then formed by creating one list for each group size; a node with k_i-degree d_i appears precisely d_i times in the list corresponding to groups of size k_i. The lists are then shuffled and half-edges of nodes in positions $k+1, \ldots, k+k_i$ in the k_i-list are joined. The structure of a graph so obtained would be characterized by fully connected cliques, and similar to that of a random intersection graph (Ball et al. 2014). One possible approach to investigate epidemics on such graphs would be to approximate the spread of the disease by a multitype Galton Watson process where groups (or cliques) are represented by the particles of the branching process. The types of the approximating branching process would then be vectors in $\mathbb{N}^2$ of the form (m, n), where m represents the size of the clique and n represents the number of members of the clique that the primary case of the clique attempts to infect. Another possible approach would be to use an infinite type branching process in the spirit of Ball et al. (2014). We believe that the result would be analogous to the results obtained in Ball et al. (2014).

Appendix: Proof of proposition 1

Let $\bar{d} = \{(S_i, \Delta_i)\}_{i \in \mathbb{N}}$ be a given (i.e. non-random) degree sequence that satisfies the following regularity assumptions.

A1) $\frac{\sum_{i=1}^N \mathbb{1}(S_i = k_1, \Delta_i = k_2)}{N} \to p(k_1, k_2)$ for any $k_1, k_2 \in \mathbb{Z}_{\geq 0}$.

A2) $\frac{\sum_{i=1}^N \Delta_i^2}{N} \to E(\Delta^2)$ and $\frac{\sum_{i=1}^N S_i^2}{N} \to E(S^2)$

where (S, Δ) has distribution p, which is assumed to satisfy A1-A2 in Section 2.1. Let further $\overline{G} = \{G_N\}_{N \in \mathbb{N}}$ be a sequence of graphs generated by the CMC, where the degree sequence of G_N is given by $\bar{d}_N = \{(S_i, \Delta_i)\}_{i=1}^N$ and denote $D_S^{(N)} = \sum_{i=1}^N S_i$.

Under the assumptions A1-A2 the expected number of self-loops and the expected number of multiple edges are borth of order $O(1)$ (cf. Van der Hofstad (2016, prop. 7.11)). Denote by A_N the number of wedges of G_N

that are "deleted" when merging multiple edges and erasing self-loops, that is

$$A_N = \sum_{i=1}^{N} \binom{S_i + 2\Delta_i}{2} 2 - |\mathcal{W}_\wedge^{G_N}| = \sum_{i=1}^{N} (S_i + 2\Delta_i)(S_i + 2\Delta_i - 1) - |\mathcal{W}_\wedge^{G_N}|,$$

then $E(A_N) = O(1)$.

From the definition of A_N, we deduce that the total number of ordered triangles of G_N is bounded from below by $|\mathcal{W}_\Delta^{G_N}| \geq \sum_{i=1}^{N} 2\Delta_i - A_N$ and the total number of ordered wedges is bounded from above by

$$|\mathcal{W}_\wedge^{G_N}| \leq \sum_{i=1}^{N} \binom{S_i + 2\Delta_i}{2} 2 = \sum_{i=1}^{N} (S_i + 2\Delta_i)(S_i + 2\Delta_i - 1).$$

Therefore, by the definition of $C(G_N)$ and the assumptions above

$$C(G_N) \geq \frac{\left(\frac{\sum_{i=1}^{N} 2\Delta_i}{N}\right) - A_N}{\left(\frac{\sum_{i=1}^{N}(S_i+2\Delta_i)(S_i+2\Delta_i-1)}{N}\right)} \xrightarrow{P} \frac{E(2\Delta)}{E((2\Delta + S)^2) - E(2\Delta + S)} \quad (30)$$

as $N \to \infty$.

This lower bound is tight in the limit as the number of nodes $N \to \infty$. Indeed, denote by $\mathcal{W}_s^{G_N}$ the set of ordered triangles of G_N that consists solely of single edges, i.e.

$$\mathcal{W}_s^{G_N} = \{(u, v, w) \in V_N^3 : \ (u, v), (u, w) \text{ and } (v, w) \text{ are single edges}\},$$

where V_N is the node set of G_N. Now, whenever $D_S^{(N)} \geq 6$

$$E\left(|\mathcal{W}_s^{G_N}|\right) \leq \sum_i \left(\binom{S_i}{2} \sum_j \frac{S_j}{D_S^{(N)} - 2} \left(\sum_l \frac{S_l}{D_S^{(N)} - 3} \left(\frac{(S_j - 1)(S_l - 1)}{D_S^{(N)} - 5}\right)\right)\right) \quad (31)$$

where the sums run over the integers $1, \ldots, N$.

Dividing by N in (31) and letting N approach infinity gives $E(|\mathcal{W}_S^{G_N}|)/N \to 0$ as $N \to \infty$. Thus $|\mathcal{W}_S^{G_N}|/N \to 0$ in probability. Repeating this procedure for triangles formed by a combination of triangle and single edges gives

$$C(G_N) \xrightarrow{P} \frac{E(2\Delta)}{E((2\Delta + S)^2) - E(2\Delta + S)}. \quad (32)$$

The assertion now follows by bounded convergence and the law of large numbers.

Acknowledgements

References

Håkan Andersson. Epidemic models and social networks. *Mathematical Scientist*, 24(2):128–147, 12 1999.

Frank Ball. Deterministic and stochastic epidemics with several kinds of susceptibles. *Advances in Applied Probability*, 17(1):1â22, 1985.

Frank Ball, David Sirl, and Pieter Trapman. Threshold behaviour and final outcome of an epidemic on a random network with household structure. *Advances in Applied Probability*, 41(3):765–796, September 2009.

Frank Ball, David Sirl, and Pieter Trapman. Analysis of a stochastic SIR epidemic on a random network incorporating household structure. *Mathematical Biosciences*, 224(2):53 – 73, 2010.

Frank Ball, David Sirl, and Pieter Trapman. Epidemics on random intersection graphs. *The Annals of Applied Probability*, 24(3):1081–1128, 2014.

Frank Ball, Lorenzo Pellis, and Pieter Trapman. Reproduction numbers for epidemic models with households and other social structures II: Comparisons and implications for vaccination. *Mathematical Biosciences*, 274:108 – 139, 2016.

Andrew Barbour and Gesine Reinert. Approximating the epidemic curve. *Electron. J. Probab.*, 18:30 pp., 2013.

Shankar Bhamidi, Remco van der Hofstad, and Júlia Komjáthy. The front of the epidemic spread and first passage percolation. *Journal of Applied Probability*, 51(A):101â121, 2014.

Béla Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1(4):311 – 316, 1980.

Béla Bollobás, Svante Janson, and Oliver Riordan. Sparse random graphs with clustering. *Random Structures & Algorithms*, 38(3):269–323, 2011.

Tom Britton. Stochastic epidemic models: A survey. *Mathematical Biosciences*, 225(1):24 – 35, 2010.

Tom Britton, Svante Janson, and Anders Martin-Löf. Graphs with specified degree distributions, simple epidemics, and local vaccination strategies. *Advances in Applied Probability*, 39(4):922–948, 2007.

Tom Britton, Maria Deijfen, Andreas N. Lagerås, and Mathias Lindholm. Epidemics on random graphs with tunable clustering. *J. Appl. Probab.*, 45 (3):743–756, 09 2008.

Emilie Coupechoux and Marc Lelarge. Contagions in random networks with overlapping communities. *Advances in Applied Probability*, 47(4):973â988, 2015.

Paul Erdős and Alfréd Rényi. On random graphs i. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959 1959.

Svante Janson, Malwina Luczak, and Peter Windridge. Law of large numbers for the SIR epidemic on a random graph with given degrees. *Random Structures & Algorithms*, 45(4):726–763, 2014.

Michał Karoński, Edward Scheinerman, and Karen Singer-cohen. On random intersection graphs: The subgraph problem. *Combinatorics, Probability and Computing*, 8(1-2):131â159, 1999.

Eben Kenah and James M. Robins. Second look at the spread of epidemics on networks. *Phys. Rev. E*, 76:036113, Sep 2007.

Kari Kuulasmaa. The spatial general epidemic and locally dependent random graphs. *Journal of Applied Probability*, 19(4):745–758, 1982.

Donald Ludwig. Final size distribution for epidemics. *Mathematical Biosciences*, 23(1):33 – 46, 1975.

Ronald Meester and Pieter Trapman. Bounding the size and probability of epidemics on networks. *Advances in Applied Probability*, 43(2):335–347, 2011.

Joel Miller. Epidemic size and probability in populations with heterogeneous infectivity and susceptibility. *Physical Review E*, 76:010101(R), 2007.

Joel Miller. Bounding the size and probability of epidemics on networks. *Journal of Applied Probability*, 45(2):498–512, 2008.

Joel Miller. Percolation and epidemics in random clustered networks. *Phys. Rev. E*, 80:020901, Aug 2009.

Michael Molloy and Bruce Reed. A critical point for random graphs with a given degree sequence. *Random Struct. Algorithms*, 6(2-3):161–180, March 1995.

Michael Molloy and Bruce Reed. The size of the giant component of a random graph with a given degree sequence. *Combinatorics, Probability and Computing*, 7(3):295â305, 1998.

Mark Newman. Spread of epidemic disease on networks. *Physical review. E*, 66:016128, 08 2002.

Mark Newman. Properties of highly clustered networks. *Phys. Rev. E*, 68: 026121, Aug 2003.

Mark Newman. Random graphs with clustering. *Phys. Rev. Lett.*, 103: 058701, Jul 2009.

Mark Newman, D. Watts, and S. Strogatz. Random graph models of social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(3):2566–2572, 2002.

Remco Van der Hofstad. *Random graphs and complex networks*, volume 43. Cambridge University Press, 2016.

Lorenzo Pellis, Neil M. Ferguson, and Christophe Fraser. The relationship between real-time and discrete-generation models of epidemic spread. *Mathematical Biosciences*, 216(1):63 – 70, 2008.

Lorenzo Pellis, Frank Ball, and Pieter Trapman. Reproduction numbers for epidemic models with households and other social structures. I. definition and calculation of R_0. *Mathematical Biosciences*, 235(1):85–97, 2012.

Pieter Trapman. On analytical approaches to epidemics on networks. *Theoretical Population Biology*, 71(2):160 – 173, 2007.

Richard S. Varga. *Matrix Iterative Analysis*. Springer, 2009. ISBN 9783642051562 3642051561.

Erik Volz, Joel Miller, Alison Galvani, and Lauren Meyers. Effects of heterogeneous and clustered contact patterns on infectious disease dynamics. *PLoS computational biology*, 7:e1002042, 06 2011.