Stochastic SIR epidemics in structured populations

CIMPA school on Probilistic Models in Epidemiology

David Sirl, University of Nottingham, UK

December 2015

(With thanks to Frank Ball)

Motivation

- Mathematically tractable epidemic models are valuable tools for understanding, predicting, mitigating, planning, . . . in the context of infectious diseases.
- Classical models include many assumptions of homogeneity, most of which are unrealistic.
- In this lecture series we focus on ways of reflecting population structure by differentiating between 'local' and 'global' contacts.
- Broadly, this means adding another infection mechanism, or layer of structure, to the standard homogeneously mixing stochastic SIR model.

Contents

Section 1: Review (& extension) of standard SIR epidemic

- Approach to analysis
- Exact results for the final outcome
- Branching process approximation for early stages of an outbreak
- Threshold theorem and probability of a major outbreak
- Law of Large Numbers and Central Limit Theorem for the final size of a major outbreak

Standard stochastic SIR epidemic

- Population of n individuals.
- Each is Susceptible, Infectious or Removed.
- Start with 1 infectious (chosen uniformly at random) and the remaining $n-1$ susceptible.
- Infectious individual remains so for a random time distributed as I, a random variable with arbitrary distribution which we specify via its MGF/LST $\phi(\theta) = \mathbb{E}[\mathrm{e}^{-\theta I}]$.
- Through its infectious period an infective makes contacts with each other individual in the population at the points of a Poisson process of rate λ. If an individual so contacted is susceptible it becomes infectious, otherwise nothing happens.
- Epidemic ceases when no infectious individuals remain.
- All infectious periods, contact processes are mutually independent.

Analysis

- First object of interest is the *final size* Z, the number of initial susceptibles that are ultimately removed (i.e. were infected during the epidemic).
- Analysis in the limit as $n \to \infty$.
- $Z^{(n)}$ is either $O(1)$ or $O(n)$: minor or major outbreak.
- Therehold theorem.
- Major outbreak probability.
- Behaviour of $Z^{(n)}$ conditional on a major outbreak.
- Key tool: Branching process approximation of early stages.

- We will obtain analogues of these results for models with additional population structure.

Tools

The following results/ideas from the standard homogeneously mixing SIR epidemic will be crucial to our analysis of these models with additional population structure:

- Properties of SIR epidemics in small groups: joint generating function of the size and severity.
- As above, including outside infection.
- Branching process approximations, we will use a discrete-time 'generation' based approach rather than the 'real-time' approach.
- Formula for the expected final size.

Homogeneously mixing SIR epidemic

- Here we modify the setup and notation slightly:
  - Assume that there are m initial infectives and n initial susceptibles, so that the population size is $m + n$.
- The rest remains the same:
  - Infectious period distributed as I, with $\phi(\theta) = \mathbb{E}[\mathrm{e}^{-\theta I}]$.
  - Contacts during the infectious period at per-pair rate λ.
  - Independence.
- We analyse the final size and the severity of this process $E_{n,m}(\lambda, I) = ((X(t), Y(t)),\ t \geq 0)$. Define
  - Extinction time $T = \inf\{t > 0 \mid Y(t) = 0\}$,
  - Final size $Z = X(0) - X(T)$,
  - Severity $A = \sum_{i=-(m-1)}^{n} \mathbb{1}_{\{i \text{ infected}\}} I_i$.

Generating functions of size and severity

Theorem (see Picard and Lefèvre (1990)). Let

$$\psi_{n,m}(s,\theta) = \mathbb{E}[s^{n-Z}\,\mathrm{e}^{-\theta A}]$$

be the joint generating function of the number of survivors and severity of $E_{n,m}(\lambda, I)$. Then

$$\psi_{n,m}(s,\theta) = \sum_{k=0}^{n} \frac{n!}{(n-k)!}\,\phi(\theta+\lambda k)^{n+m-k}\,G_k(s \mid \mathcal{U}).$$

Here $\mathcal{U} = (u_0, u_1, \dots)$ has $u_k = \phi(\theta+\lambda k)$ and $G_k(s \mid \mathcal{U})$ is the k-th Gontcharoff polynomial (with parameter sequence $\mathcal{U}$), defined by $\sum_{i=0}^{k} \frac{k!}{(k-i)!}\,u_i^{k-i}\,G_i(x \mid \mathcal{U}) = x^k \quad (k = 0, 1, \dots)$.

Generating functions of size and severity

Corollary 1 Set $s = 1$ to obtain the MGF of the severity A of $E_{n,m}(\lambda, I)$.

Corollary 2 Let

$$f_{n,m}(s) = \mathbb{E}[s^{n-Z}]$$

be the PGF of the ultimate number of susceptibles $n - Z$ in $E_{n,m}(\lambda, I)$. Then, setting $\theta = 0$ in Theorem 1 yields

$$f_{n,m}(s) = \sum_{k=0}^{n} \frac{n!}{(n-k)!} \phi(\lambda k)^{n+m-k} G_k(s|\mathcal{V}),$$

where $\mathcal{V} = (v_0, v_1, \dots)$ with $v_k = \phi(\lambda k)$.

Final size of $E_{n,m}(\lambda, I)$

Formulae for the mean final size and the distribution of the final size follows from Corollary 2. Starting with $f_{n,m}(s)$:

- Differentiating once and setting $s = 1$ yields

$$\mathbb{E}[Z] = n - \mathbb{E}[n - Z] = n - \sum_{k=1}^{n} \frac{n!}{(n-k)!} \phi(\lambda k)^{n+m-k} G_{k-1}(1 \mid E^1 \mathcal{V}),$$

  where $E^k \mathcal{V} = (v_k, v_{k+1}, \dots)$.

- Differentiating $n - z$ times and setting $s = 0$ yields

$$\mathbb{P}(Z = z) = \frac{1}{(n-z)!} \sum_{k=n-z}^{n} \frac{n!}{(n-k)!} \phi(\lambda k)^{n+m-k} G_{k-n+z}(0 \mid E^{n-z} \mathcal{V}).$$

SIR model with outside infection

- Consider the model $E_{n,m}(\lambda, I)$ with the additional feature that susceptibles may be infected from outside the population.
- Specifically, each susceptible avoids outside infection independently with probability π.
- Individuals infected from outside the population infect susceptibles within the population as in $E_{n,m}(\lambda, I)$.
- Denote model by $\tilde{E}_{n,m}(\lambda, I, \pi)$.
- Let $\tilde{Z}$ and $\tilde{A}$ denote the size and severity of $\tilde{E}_{n,m}(\lambda, I, \pi)$, and let

$$\tilde{\psi}_{n,m}(s, \theta) = \mathbb{E}[s^{n-\tilde{Z}} \mathrm{e}^{-\theta \tilde{A}}].$$

(Addy et al. (1991))

Size and severity of $\tilde{E}_{n,m}(\lambda, I, \pi)$

Theorem (Ball et al. (1997)) For $n, m = 0, 1, \cdots$,

$$\tilde{\psi}_{n,m}(s, \theta) = \sum_{k=0}^{n} \frac{n!}{(n-k)!} \phi(\theta + \lambda k)^{n+m-k} \pi^k G_k(s \mid \mathcal{U}),$$

where $\mathcal{U} = u_0, u_1, \ldots$ with $u_k = \phi(\theta + \lambda k)$.

Size and severity of $\tilde{E}_{n,m}(\lambda, I, \pi)$

- Expressions for the MGF of the severity $\tilde{A}_n$ and the PGF of the number of survivors $n - \tilde{Z}_n$ follow as before. Also,

$$\mathbb{E}[\tilde{Z}_n] = n - \sum_{k=1}^{n} \frac{n!}{(n-k)!} \phi(\lambda k)^{n+m-k} \pi^k G_{k-1}(1 \mid \mathcal{W}),$$

  where $\mathcal{W} = (w_0, w_1, \dots)$ with $w_k = \phi(\lambda(k+1))$.
- For fixed m, let $\tilde{P}_i^n = \mathbb{P}(\tilde{Z}_n = i)$ $(i = 0, 1, \dots, n)$. Then

$$\sum_{i=0}^{j} \frac{\binom{n-i}{j-i} \tilde{P}_i^n}{\phi(\lambda(n-j))^{m+i} \pi^{n-j}} = \binom{n}{j} \qquad (j = 0, 1, \dots, n).$$

Branching process approximations

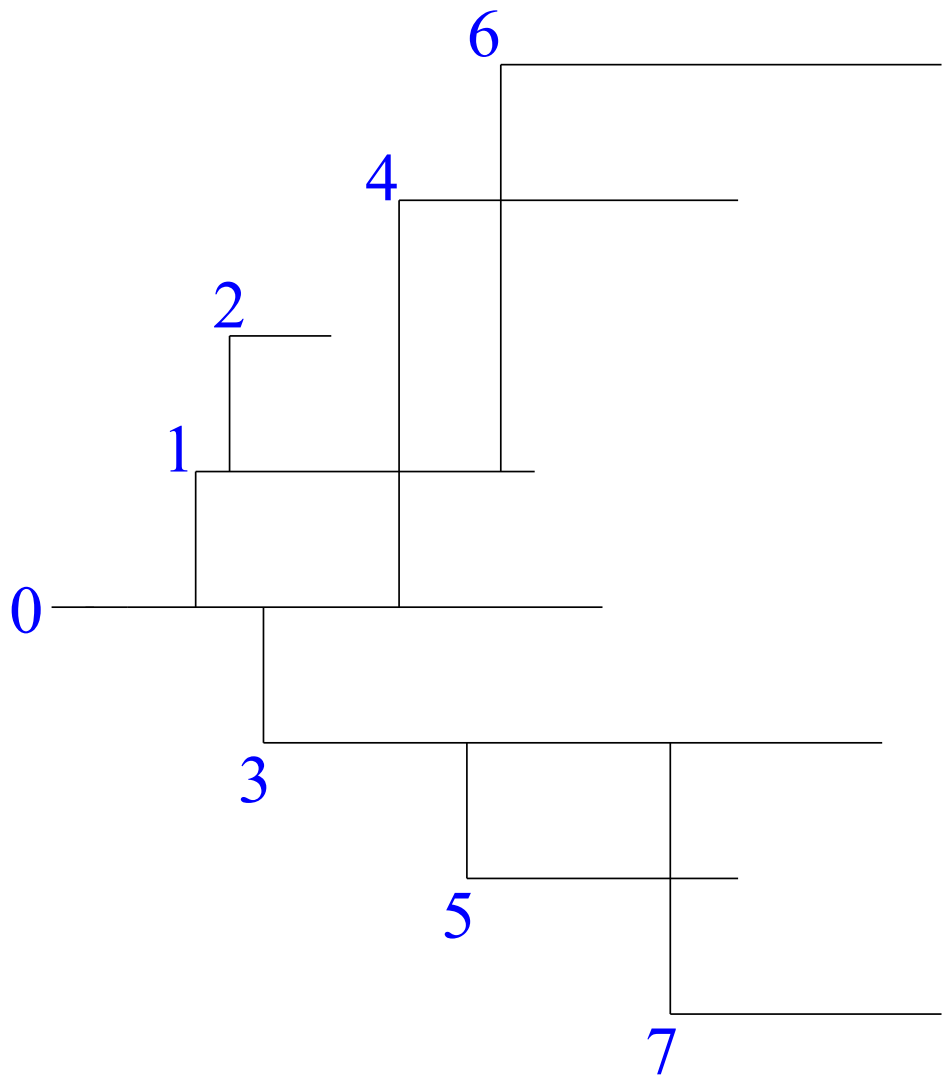
- Depending upon the aim of our analysis, there are two (main) different approaches to branching process approximation of the number of infectives in the early stages of an epidemic:
  - Real-time; Crump-Mode-Jagers branching process.
  - Generation based; embedded (Bienaymé-)Galton-Watson process.
- Here we motivate the latter and briefly touch on the connection between the two.

General branching process

- In a general Crump–Mode–Jagers (CMJ) branching process individuals have IID *life histories* $\mathcal{H} = (I, \xi)$, where I denotes a typical individual's *age at death* and ξ is a point process of ages at which she reproduces. [Note that $\xi((I, \infty)) = 0$.]
- Thus if an individual with life history $\mathcal{H} = (I, \xi)$ is born at time b and $0 < \tau_1 \leq \tau_2 \leq \cdots \leq I$ denote the points of ξ then she has one child at each time $b + \tau_1, b + \tau_2, \ldots$.
- The life histories are pieced together in the obvious fashion to form the population process.
- Such a process approximates $Y(t)$, the (real-time) evolution of the number of infectives.

(Haccou *et al.* (2005).)

General branching process



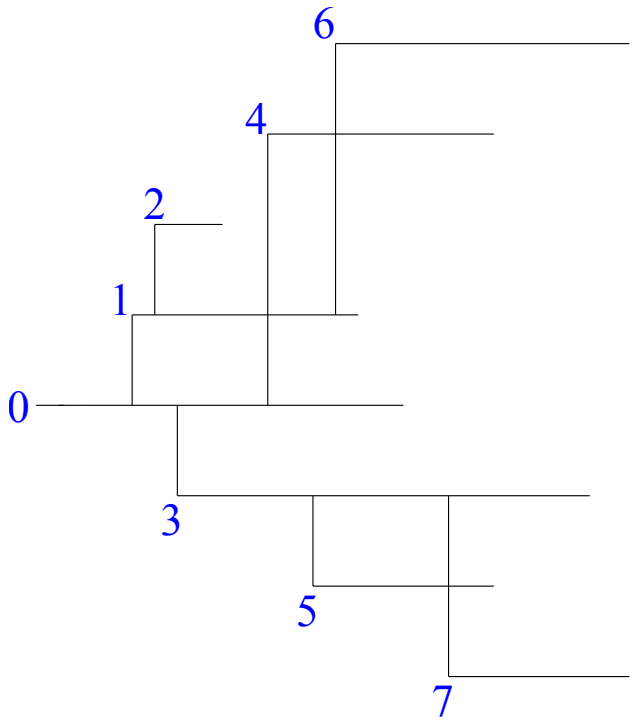
Branching processes and final size

- If we are interested in analysing the *final size* properties of an SIR epidemic process then we do not need all of this information.
- All that matters is who has infectious contact with who.
- The times of the contacts do not affect the final size.

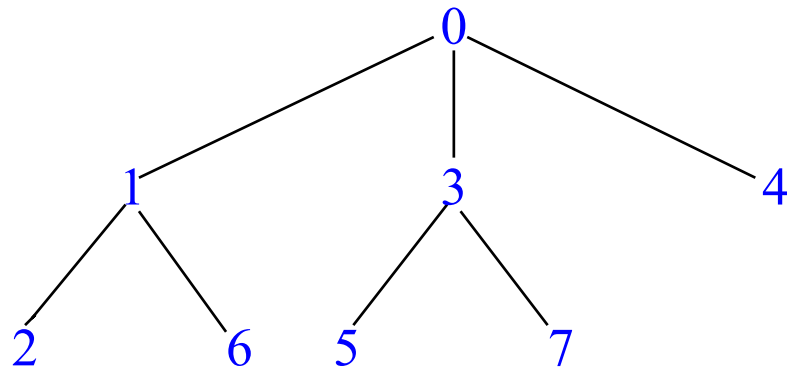
(Ludwig (1975))

Embedded Galton-Watson process

CMJ process



Embedded GW process



Embedded Galton-Watson process

- Suppose that there are m initial individuals. These comprise generation 0. For $k = 0, 1, 2, \ldots$, generation $k+1$ consists of the daughters of generation-k individuals.
- Let $R = \xi((0, \infty))$ be a random variable describing the number of offspring of a typical individual.
- For $k = 0, 1, \ldots$, let Y_k denote the size of generation k. Then $Y_0 = m$ and, for $k = 1, 2, \ldots$,

$$Y_k = \begin{cases} R_{k-1,1} + R_{k-1,2} + \cdots + R_{k-1,Y_{k-1}} & \text{if } Y_{k-1} > 0, \\ 0 & \text{if } Y_{k-1} = 0, \end{cases}$$

where $R_{k,i} \overset{\text{iid}}{\sim} R$.

Extinction

- Suppose $\mathbb{P}(R = 1) < 1$. Then a Galton-Watson process ultimately either goes extinct or grows unboundedly.
- Let

$$f(s) = \mathbb{E}\left[s^R\right] = \sum_{k=0}^{\infty} \mathbb{P}(R = k)s^k \quad (0 \leq s \leq 1)$$

  be the PGF of R and let π be the probability that the GW process goes extinct given that there is one ancestor.
- Then π is the smallest non-negative solution of $f(s) = s$.
- Let $R_0 = \mathbb{E}[R]$. Then $\pi < 1 \iff R_0 > 1$.
- If there are m ancestors, the extinction probability is π^m.
- If $\mathbb{P}(I < \infty) = 1$, a CMJ process goes extinct if and only if its embedded GW process of generation sizes does so.

Total progeny

- Let $Z = Y_1 + Y_2 + \dots$ denote the total progeny of the embedded Galton-Watson process $\{Y_k : k = 0, 1, \dots\}$, not including the m ancestors. (Note that Z is also the total progeny of the corresponding CMJ branching process.)
- Then

$$\mathbb{P}(Z = k) = \frac{m}{m+k}\mathbb{P}(R_1 + R_2 + \cdots + R_{m+k} = k) \qquad (k = 0, 1, \dots),$$

  where $R_1, R_2, \dots$ are IID copies of R.
- Note that

$$\sum_{k=0}^{\infty} \mathbb{P}(Z = k) = \begin{cases} 1 & \text{if } R_0 \leq 1, \\ \pi^m < 1 & \text{if } R_0 > 1. \end{cases}$$

Asymptotic final size properties of $E_{n,m}(\lambda, I)$

- As $n \to \infty$, $Z^{(n)} \xrightarrow{\mathrm{D}} Z$, where Z is the total progeny of a suitable branching process.
- If $P(Z = \infty) > 0$ then $Z^{(n)}/n \xrightarrow{\mathrm{D}} Z'$, where

$$1 - \mathbb{P}(Z' = 0) = \mathbb{P}(Z' = z) = p_{\mathrm{maj}}$$

  and $1 - p_{\mathrm{maj}}$ and $1 - z$ are branching process extinction probabilities.
- (CLT for size of major outbreaks.)
- Asymptotic expected relative final size z satisfies

$$1 - z = \exp(-z\frac{\lambda}{N}N\mu_I) = \exp(-z\lambda\mu_I).$$

Summary

- Exact results for small populations.
- To study final size we may ignore time.
- Threshold theorem.
- Law of large numbers (and central limit theorem) for final size of major outbreaks.

Going forward

This course is about (stochastic SIR) structured population epidemic models.

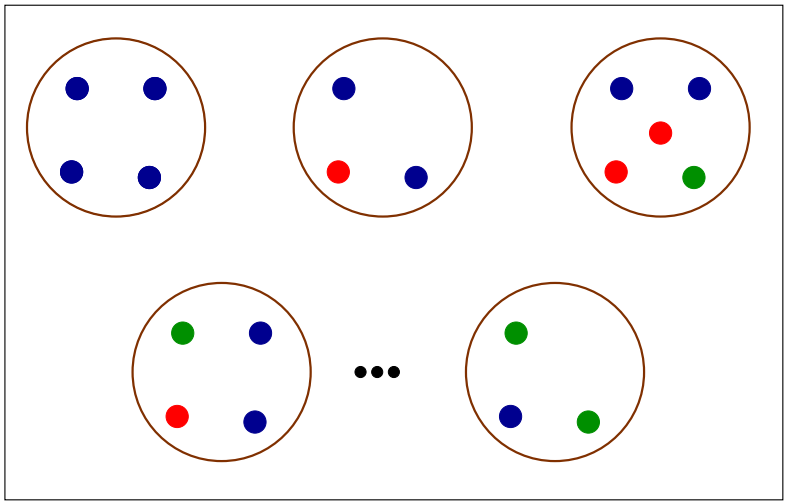
We look in detail at

- Households models
  (Ball, Mollison & Scalia-Tomba (1997), Ball & Lyne (2006)).
- General two-level-mixing model
  (Ball & Neal (2002, 2008)).
- Network and households model
  (Ball, Sirl & Trapman (2009, 2010)).

Why study households models?

- Household structure is a key departure from homogeneous mixing for human populations and can have significant impact on disease dynamics.
- There are outbreak control measures associated with households and similar structures (e.g. schools and workplaces).
- Epidemic data are often collected at the household level.
- Households models are mathematically reasonably tractable and consequently are generally easier to interpret than complex simulation models.

Households SIR epidemic model



m_n households of size n $(n = 1, 2, \dots)$

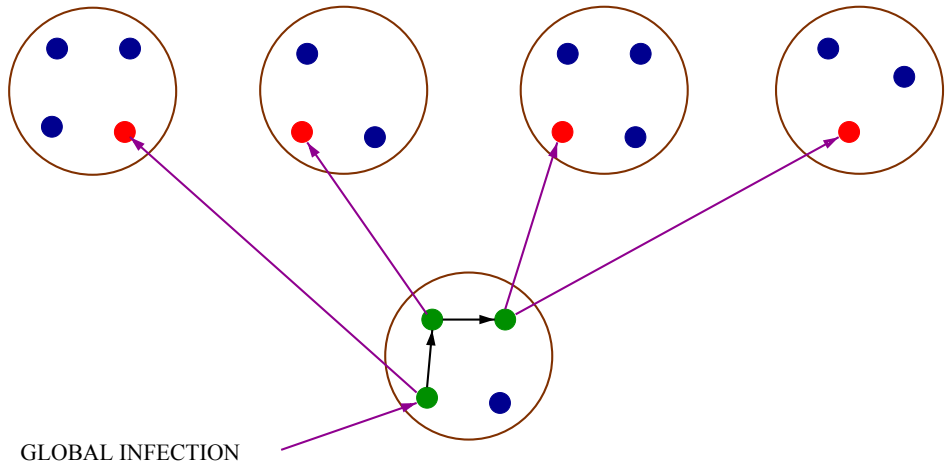
total of $m = \sum_{n=1}^{\infty} m_n$ households

and $N = \sum_{n=1}^{\infty} n m_n < \infty$ individuals

- SIR (susceptible $\rightarrow$ infective $\rightarrow$ removed)
- Infectious period $\sim I$, an arbitrary but specified distribution
- Infection rates (individual to individual)
  - local (within-household) λ_L
  - global (between-household) λ_G / N
- Latent period

(Bartoszyński (1972), Becker and Dietz (1995), Ball *et al.* (1997))

Threshold parameter R_*



- R_* = mean number of global contacts emanating from a typical single-household epidemic. Letting

$$\tilde{\alpha}_n = \frac{n m_n}{N} = \mathbb{P}(\text{randomly chosen person lives in a household of size } n),$$

$$\mu_n(\lambda_L) = \text{mean size of size-}n\text{ household epidemic with 1 initial infective},$$

$$R_* = \sum_{n=1}^{\infty} \tilde{\alpha}_n \mu_n(\lambda_L) \lambda_G \mathbb{E}[I].$$

- Therefore $p_{\mathrm{maj}} > 0 \iff R_* > 1$.

Probility of a major outbreak

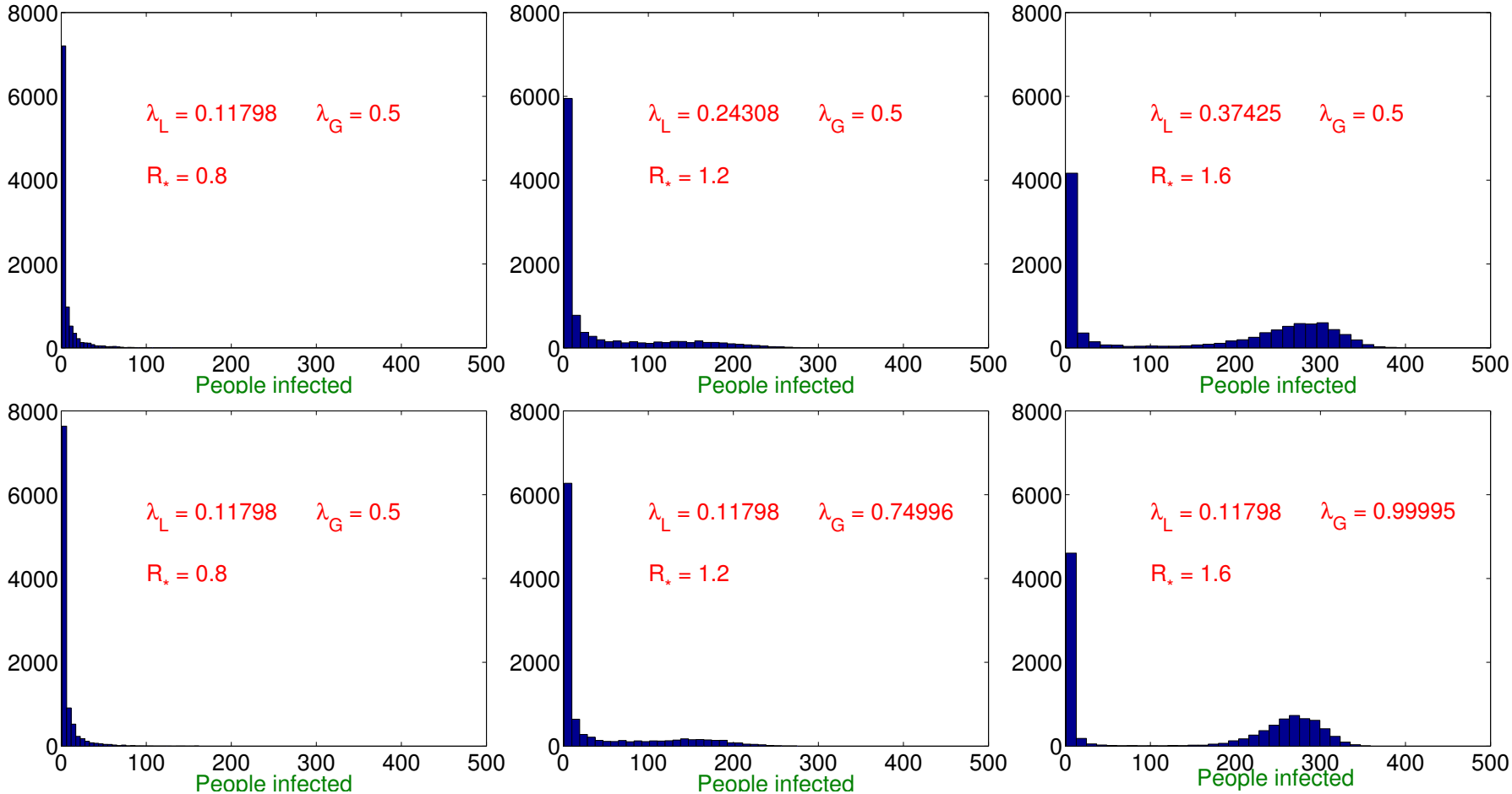
- Approximate early stages of an epidemic by a branching process of infected households.
- Number of global contacts emanating from a single size-n household epidemic, R_n say, follows a Poisson distribution with random mean $\lambda_G A_{n-1}$, where A_{n-1} is the severity of a single-household epidemic with initially 1 infective and $n-1$ susceptibles. Thus, recalling notation from Section 1,

$$\mathbb{E}\left[s^{R_n}\right] = \mathbb{E}\left[\mathbb{E}\left[s^{R_n} \mid A_{n-1}\right]\right] = \mathbb{E}\left[\mathrm{e}^{-\lambda_G A_{n-1}(1-s)}\right] = \psi_{n-1,1}(1, \lambda_G(1-s)).$$

- If the epidemic is started by an individual chosen uniformly at random from the population becoming infected then $p_{\mathrm{maj}} = 1 - \sigma$, where σ is the smallest non-negative solution of $f(s) = s$ and

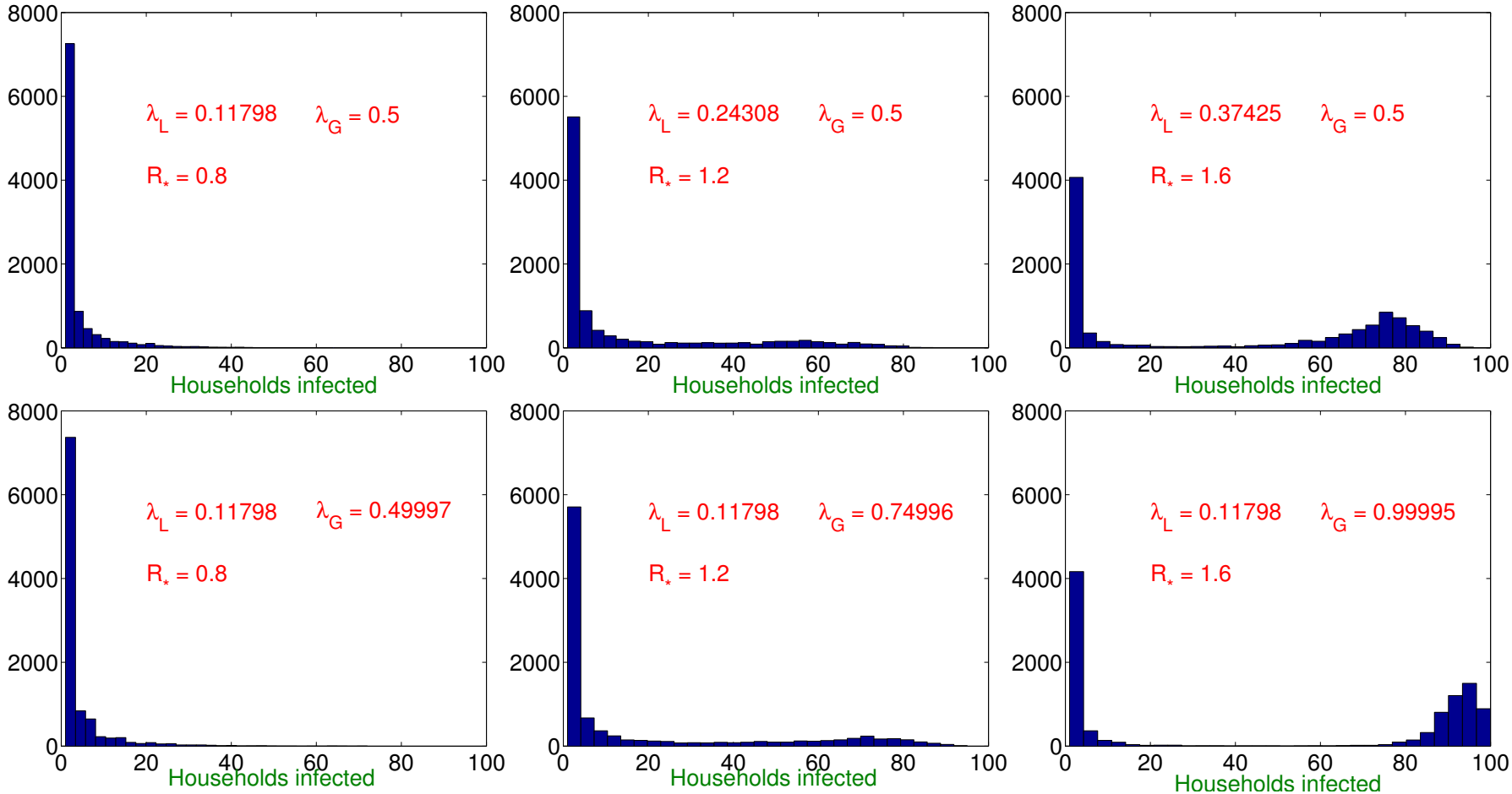
$$f(s) = \sum_{n=1}^{\infty} \tilde{\alpha}_n \phi_{n-1,1}(1, \lambda_G(1-s)).$$

Number of people infected



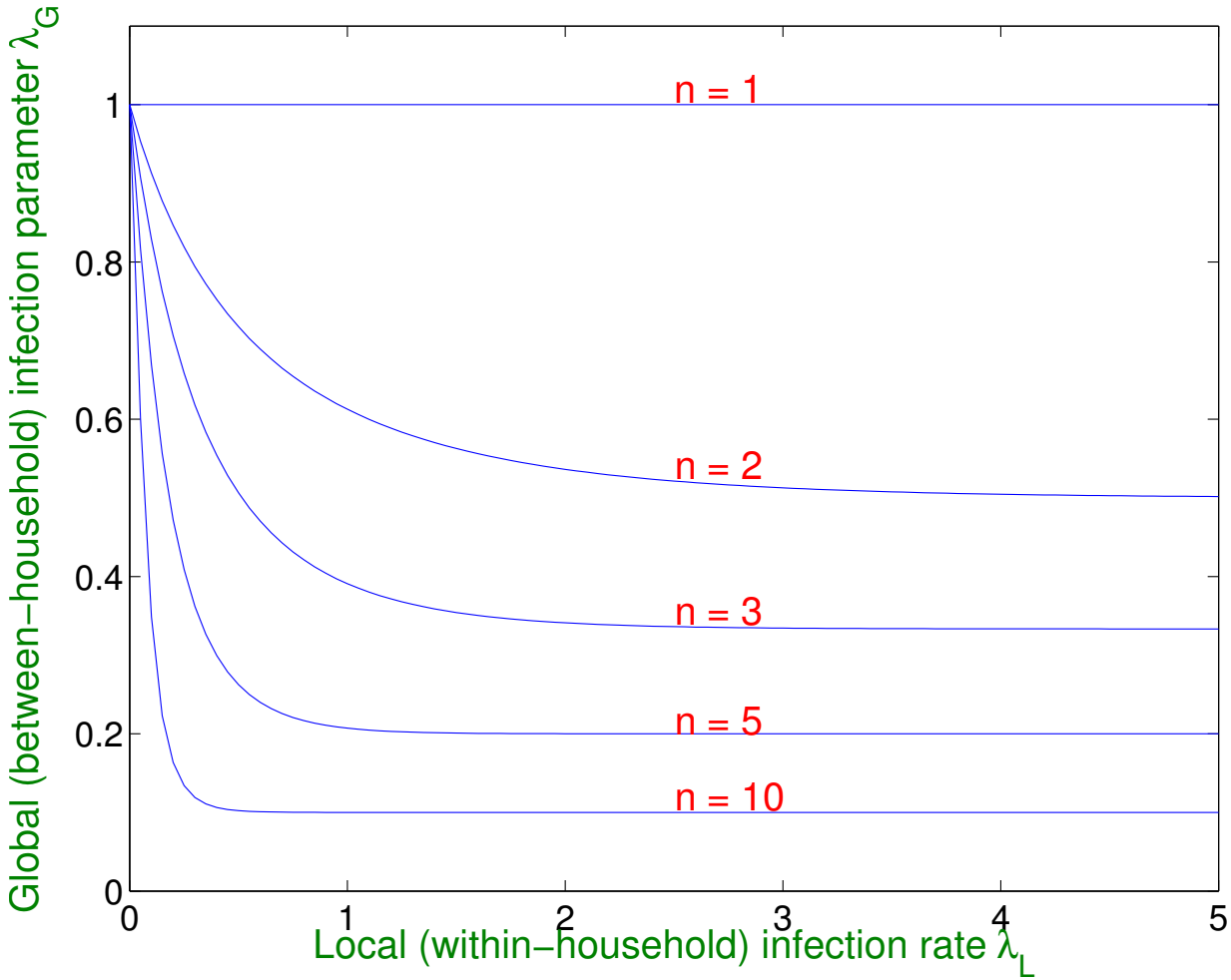
Number of people infected in each set of 10,000 simulations, $I \equiv 1$, population consisting of 100 households of size 5

Number of households infected



Number of households infected in each set of 10,000 simulations, $I \equiv 1$, population consisting of 100 households of size 5

Critical values of (λ_L, λ_G)



Critical values of (λ_L, λ_G) so that $R_* = 1$, when $I \equiv 1$ and all households have size n.

SIR model with outside infection

- Recall the model $\tilde{E}_{n,m}(\lambda_L, I, \pi)$, in which individuals avoid outside infection independently with probability π, and let $\tilde{Z}_{n,m}$ denote the final size of $\tilde{E}_{n,m}(\lambda_L, I, \pi)$.
- For $n = 1, 2, \ldots,$ let

$$\tilde{\mu}_n(\lambda_L, \pi) = \mathbb{E}[\tilde{Z}_{n,0}]$$

  be the expected final size of such an epidemic in an initially fully susceptible household of size n.
- An expression for $\tilde{\mu}_n(\lambda_L, \pi)$ in terms of Gontcharoff polynomials is given in Section 1.

Final outcome of major outbreak

Suppose m is large and there are few initial infectives. Set
$z =$ expected proportion of the population infected by the epidemic and
$\pi =$ probability that a typical individual avoids *global* infection.

- Then z and π satisfy the following equations:

$$\pi = \exp\left(-\frac{\lambda_G}{N} N z \mu_I\right) = \exp(-\lambda_G z \mu_I), \tag{1}$$

$$z = \sum_{n=1}^{\infty} \tilde{\alpha}_n \tilde{\mu}_n(\lambda_L, \pi)/n. \tag{2}$$

- If $R_* \leq 1$ then $z = 0$ is the only solution of (1)–(2) in $[0, 1]$.
- If $R_* > 1$ then there is a unique second solution $z^* \in (0, 1)$, giving the mean relative size of major outbreak.
- Final outcome in an initially fully-susceptible household having size n is distributed according to final outcome of $\tilde{E}_{n,0}(I, \lambda_L, \pi^*)$, where $\pi^* = \exp(-\lambda_G z^* \mu_I)$.

Vaccination

- For $n = 1, 2, \cdots$ and $v = 0, 1, \cdots, n$, let
  x_{nv} = proportion of size-n households that have v members vaccinated,
  μ_{nv} = mean number of global contacts emanating from a single-household epidemic in a household in state (n, v), initiated by an individual chosen uniformly at random being contacted globally.
- Post-vaccination
$$R_v = \sum_{n=1}^{\infty} \tilde{\alpha}_n \sum_{v=0}^{n} x_{nv} \mu_{nv}$$
- Vaccination coverage
$$c = \sum_{n=1}^{\infty} \tilde{\alpha}_n \sum_{v=0}^{n} \frac{v}{n} x_{nv}$$
- Determination of optimal vaccination scheme (e.g. to reduce R_v to 1 with minimum vaccination coverage) is a linear programming problem, whose solution can be constructed explicitly.

(Becker and Starczak (1997), Ball and Lyne (2002, 2006))

Calculation of μ_{nv}

- Recall
  x_{nv} = proportion of size-n households that have v members vaccinated,
  μ_{nv} = mean number of global contacts emanating from a single-household epidemic in a household in state (n, v), initiated by an individual chosen uniformly at random being contacted globally.
- μ_{nv} depends on model for vaccine action.
- For an *all-or-nothing* model, in which vaccinees are rendered completely immune independently with probability ε, otherwise the vaccine has no effect

$$\mu_{nv} = \sum_{k=0}^{v} \underbrace{\binom{v}{k} \varepsilon^k (1-\varepsilon)^{v-k}}_{(1)} \underbrace{\frac{n-k}{n}}_{(2)} \underbrace{\mu_{n-k}(\lambda_L)}_{(3)} \lambda_G \mu_I$$

(1) $\mathbb{P}(k$ vaccinations are successful)
(2) $\mathbb{P}$(globally contacted individual is susceptible)
(3) Mean size of single-household epidemic

Vaccine response model

- Vaccine response described by a random vector (A, B).

  $A =$ relative susceptibility compared to an unvaccinated individual
  [force of infection acting on that individual at time t reduced from λ_t to $A\lambda_t$

  $B =$ relative infectivity should vaccinee become infected
  [total force of infection exerted by that individual reduced from $\int_0^\infty \lambda'_s \, ds$ to $B \int_0^\infty \lambda'_s \, ds$]

- All-or-nothing $\mathbb{P}(A = 0, B = 0) = 1 - P(A = 1, B = 1) = \varepsilon$

  Non-random $\mathbb{P}(A = a, B = b) = 1$

  Leaky non-random with $a = 1 - \varepsilon, b = 1$

- Vaccine efficacy: $VE_{\mathrm{SI}} = 1 - \mathbb{E}[AB]$ $(= \varepsilon)$

(Becker and Starczak (1998))

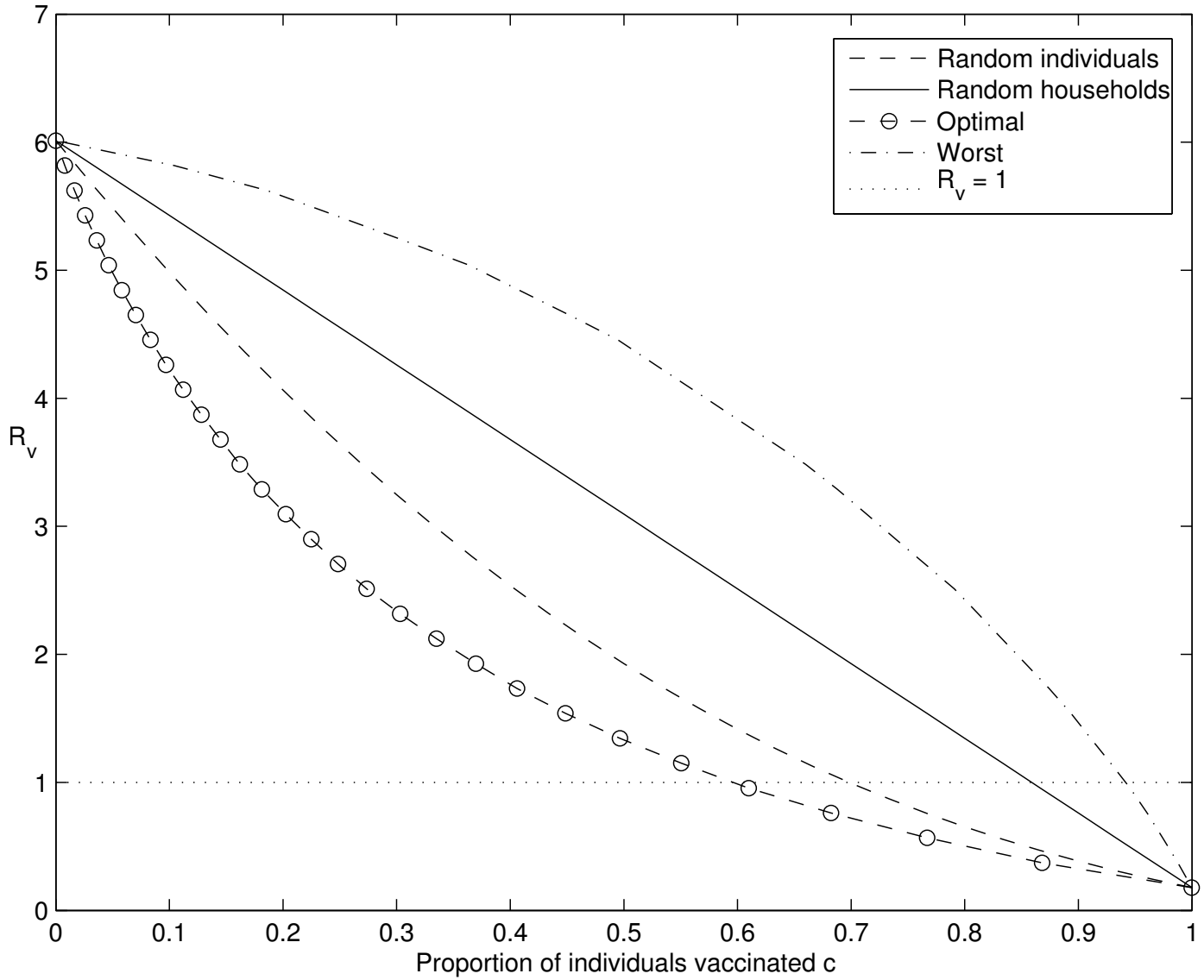
Variola Minor, Sao Paulo, 1956

- Data comprise final numbers infected in each of 338 households. Household size varied from 1 to 12 (mean 4.56)
- Each individual labelled vaccinated or unvaccinated

  773 unvaccinated — 425 infected (58%)
  809 vaccinated — 85 infected (11%)
- Fit households SIR model with non-random vaccine response, assuming infectious period $I \equiv 1$, using pseudolikelihood method of Ball and Lyne (2014) to obtain the estimates

$$\hat{\lambda}_L = 0.3821,\ \hat{\lambda}_G = 1.4159,\ \hat{a} = 0.1182,\ \hat{b} = 0.8712.$$

Comparison of vaccination strategies



Optimal vaccination schemes

- Let
  $h_{nv} = m_n x_{nv}$ = number of households of size n with v members vaccinated.
- Recalling $\tilde{\alpha}_n = n m_n / N$,

$$R_v = \sum_{n=1}^{\infty} \tilde{\alpha}_n \sum_{v=0}^{n} x_{nv} \mu_{nv}$$
$$= \sum_{n=1}^{\infty} \sum_{v=0}^{n} h_{nv} M_{n,v},$$

  where $M_{n,v} = n\mu_{nv}/N$.
- Consider the vaccine *gain*
  $G_{n,v} = M_{n,v} - M_{n,v+1}$, the reduction in R_v from vaccinating one further member of a (n, v)-household.

Vaccine gain matrix

n	$v=0$	1	2	3
1	0.5625^{10}			
2	1.5000^{6}	0.7969^{9}		
3	2.6506^{3}	1.7876^{5}	1.0446^{8}	
4	3.8165^{1}	2.9420^{2}	2.0762^{4}	1.3025^{7}

Vaccine gain matrix ($G_{n,v}$) for a population consisting of 100 households of each size 1, 2, 3 and 4, when $I \sim \mathrm{Exp}(1)$, $\lambda_L = 5$ and $\lambda_G = 0.75$, for an all-or-nothing vaccine with $\varepsilon = 0.75$.

Vaccine gain matrices

All-or-nothing, $\varepsilon = 0.75$

n	$v = 0$	1	2	3
1	0.5625^{10}			
2	1.5000^{6}	0.7969^{9}		
3	2.6506^{3}	1.7876^{5}	1.0446^{8}	
4	3.8165^{1}	2.9420^{2}	2.0762^{4}	1.3025^{7}

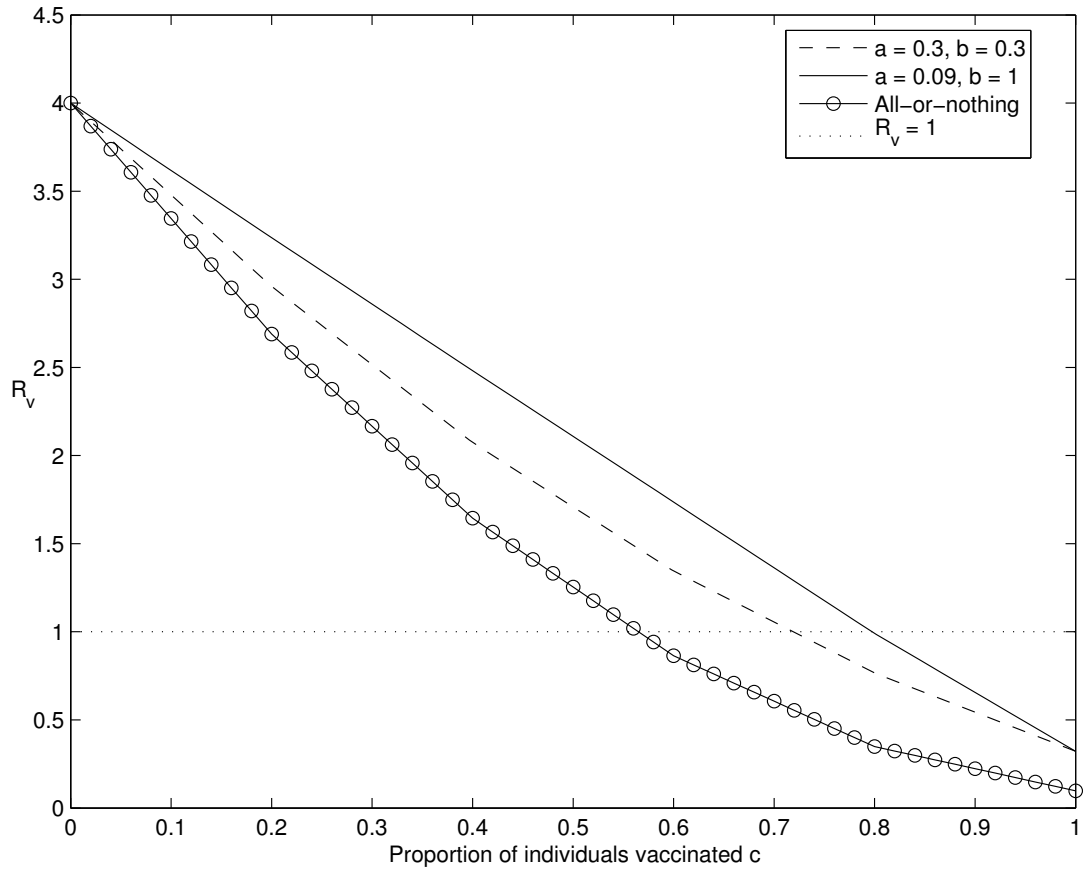
Non-random, $a = b = 0.5$

n	$v = 0$	1	2	3
1	0.5625^{10}			
2	1.2768^{7}	0.8899^{9}		
3	2.0582^{4}	1.6686^{6}	1.2706^{8}	
4	2.8224^{1}	2.4464^{2}	2.0640^{3}	1.6714^{5}

Leaky, $a = 0.25$

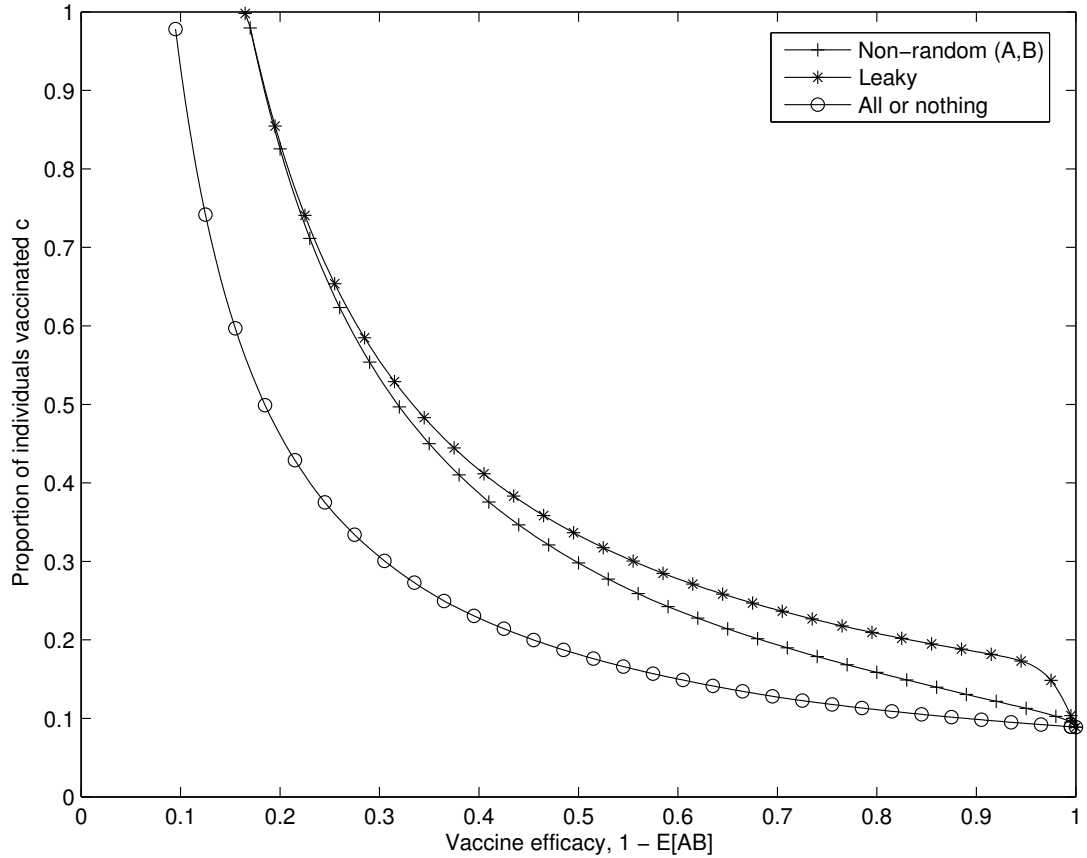
n	$v = 0$	1	2	3
1	0.5625^{10}			
2	1.2396^{8}	0.9271^{9}		
3	1.8857^{5}	1.7141^{6}	1.3976^{7}	
4	2.4365^{1}	2.3821^{2}	2.2418^{3}	1.9437^{4}

Effect of vaccine action model on R_v



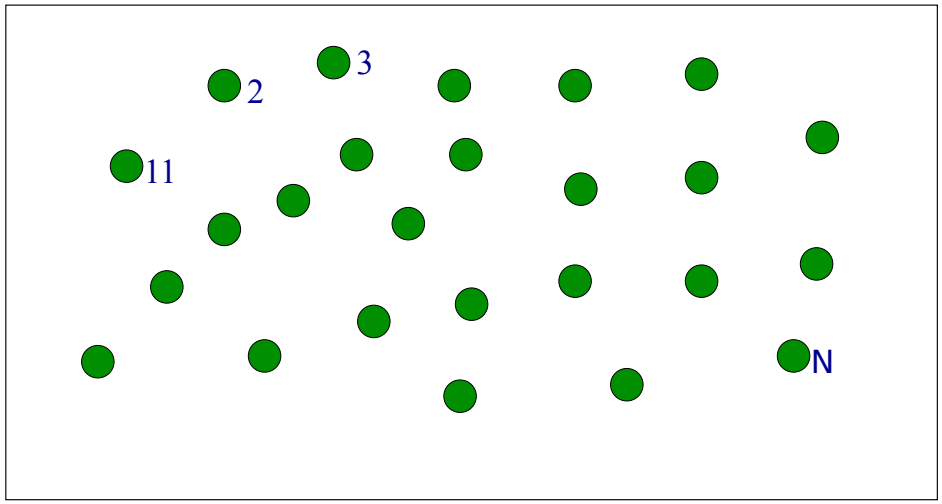
Effect of different vaccine actions calibrated by $VE_{\mathrm{SI}} = 1 - \mathbb{E}[AB] = 0.91$, for a population of households of size 5, with $I \equiv 1$, $\lambda_L = 8$ and $\lambda_G = 0.8$.

Effect of vaccine action model on c_v



Effect of different vaccine actions on critical vaccination coverage c_v to achieve $R_v = 1$ for a population of households of size 8, with $I \equiv 1$, $\lambda_L = 10$, $\lambda_G = 0.15$. In non-random case $a = b = \sqrt{\mathbb{E}[AB]}$.

Section 3: General two-level mixing epidemic model



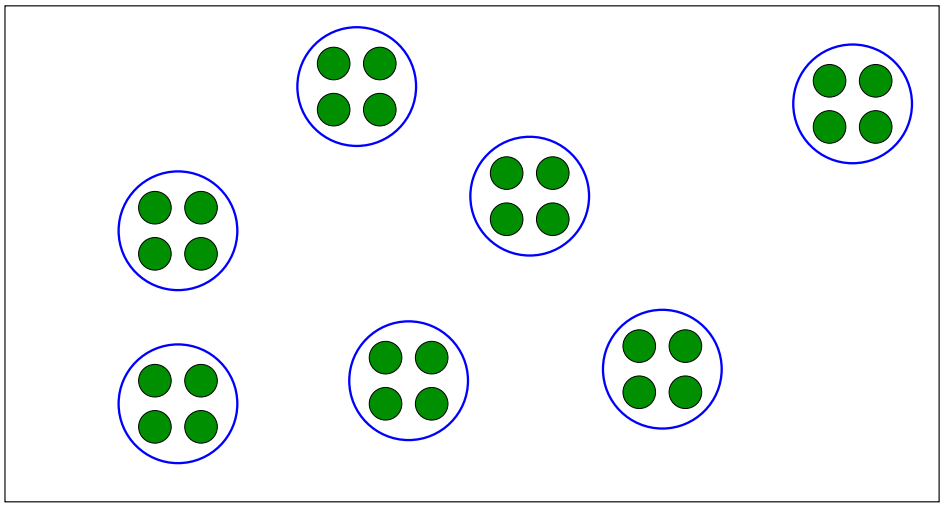
Population

$\mathcal{N} = \{1, 2, \cdots, N\}$

- SIR (susceptible $\rightarrow$ infective $\rightarrow$ removed).
- Infectious periods $I_1, I_2, \ldots, I_N \overset{\text{i.i.d.}}{\sim} I$ (arbitrary but specified).
- Infection rates (individual $\rightarrow$ individual).
  - local λ_{ij}^L,
  - global λ_G/N.
- ($\lambda_{ij}^L \equiv 0$ yields homogeneous mixing.)

(Ball and Neal (2002))

Households model



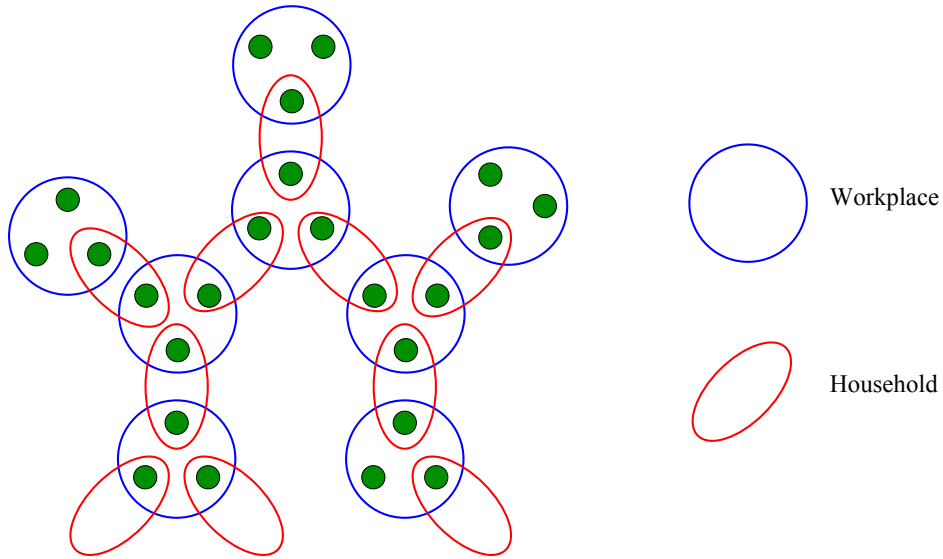
m households, each of size n

$N = mn$

- $\lambda_{ij}^{L} = \begin{cases} \lambda_L & \text{if } i \text{ and } j \text{ belong to the same household} \\ 0 & \text{otherwise} \end{cases}$
- Unequal-sized households.

(Bartoszyński (1972), Becker and Dietz (1995), Ball *et al.* (1997))

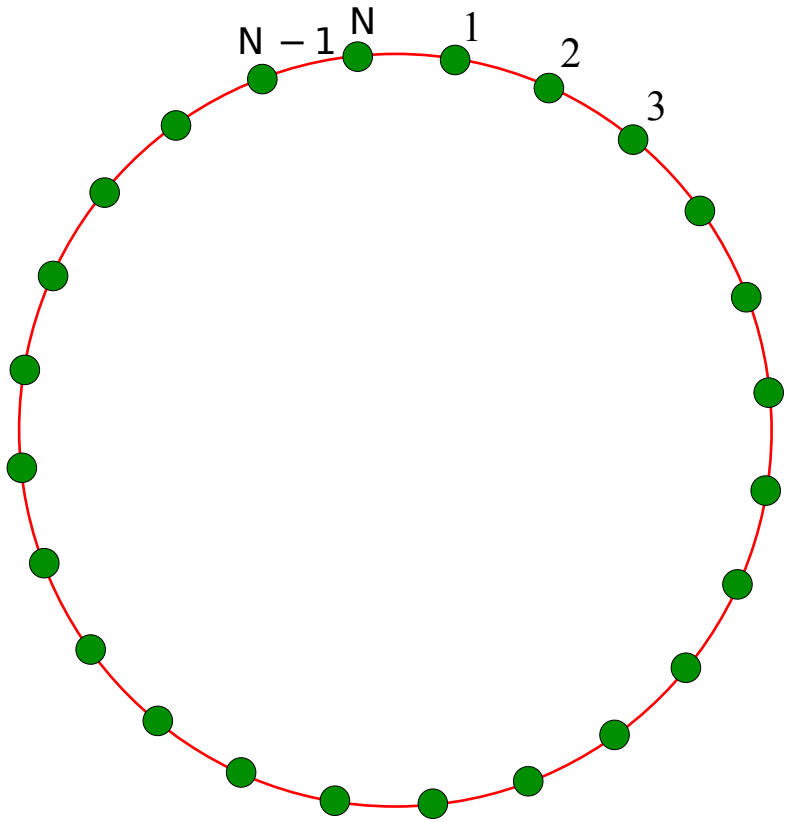
Overlapping groups model



- m_α households, each of size n_α, m_β workplaces, each of size n_β, so $N = m_\alpha n_\alpha = m_\beta n_\beta$.
- $\lambda_{ij}^L = \begin{cases} \lambda_\alpha^L & \text{if } i \text{ and } j \text{ belong to the same household,} \\ \lambda_\beta^L & \text{if } i \text{ and } j \text{ belong to the same workplace,} \\ 0 & \text{otherwise.} \end{cases}$

(Ball and Neal (2002), cf. Andersson (1999); Ball *et al.* (2014))

Great circle model



Basic form:

$$\lambda_{ij}^{L} = \begin{cases} \lambda_L & \text{if } i \text{ and } j \text{ are neighbours,} \\ 0 & \text{otherwise.} \end{cases}$$

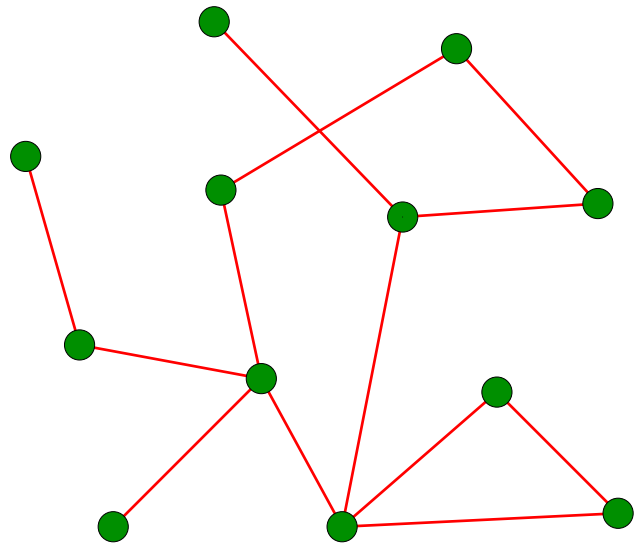
'Small-world' networks

More general contact distribution:

$$\lambda_{ij}^{L} = \lambda_L \, v(i - j \pmod N)$$

(Ball *et al.* (1997), Ball and Neal (2002, 2003))

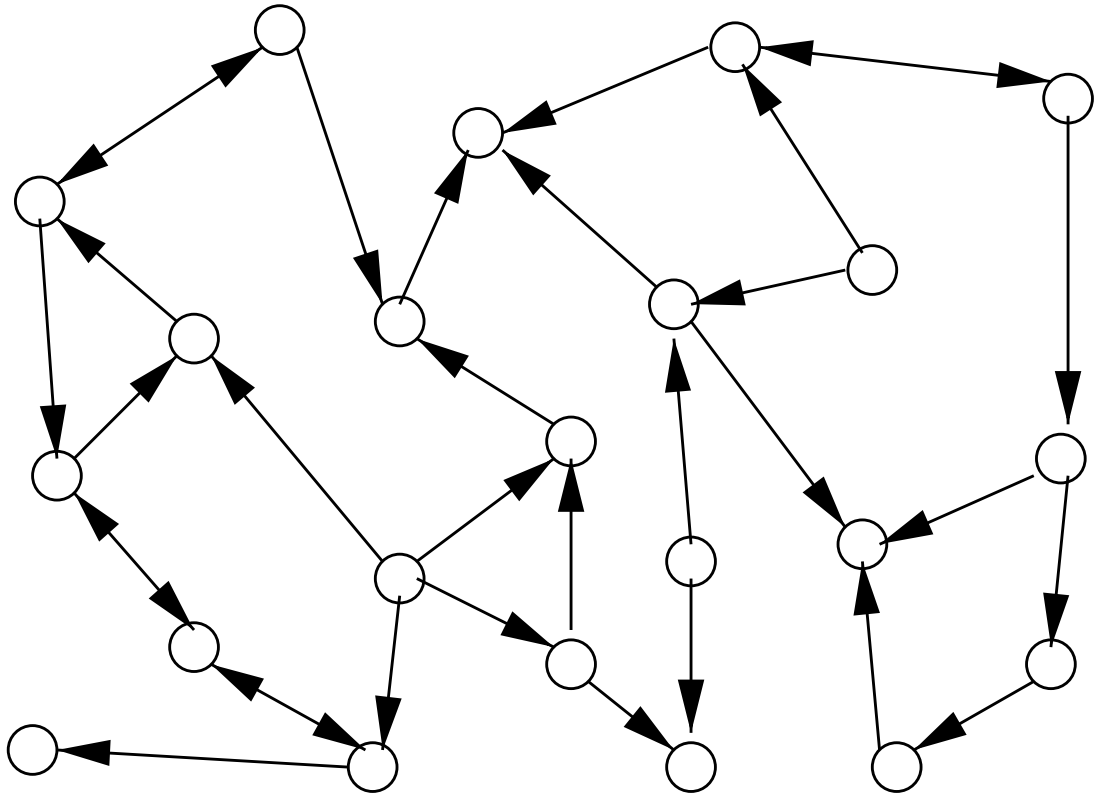
Networks with casual contacts



- 'independent' random graph of possible local contacts with specified degree distribution $p_k = \mathbb{P}(D = k) \quad (k = 0, 1, \dots)$
- $\lambda_{ij}^{L} = \begin{cases} \lambda_L & \text{if } i \text{ and } j \text{ are neighbours} \\ 0 & \text{otherwise} \end{cases}$

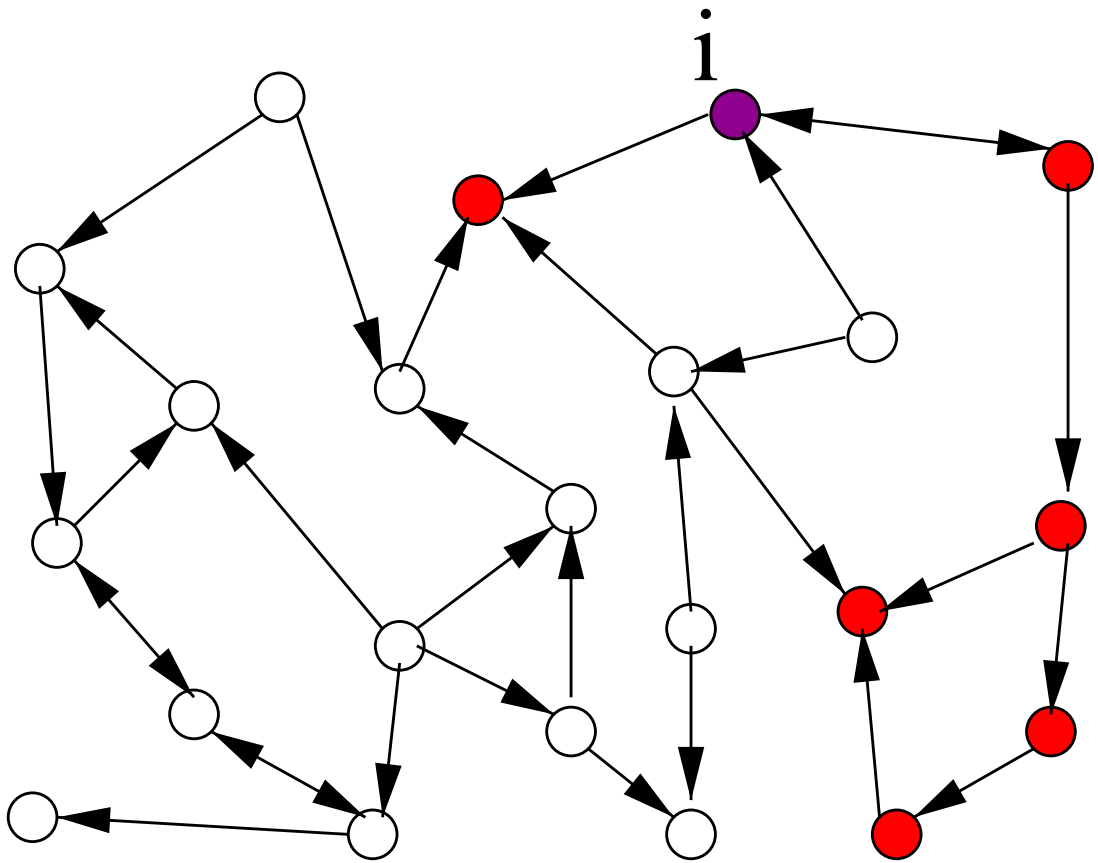
(Diekmann *et al.* (1998), Ball and Neal (2002, 2008), Kiss et al. (2006), Newman (2002))

Digraph of local infectious contacts



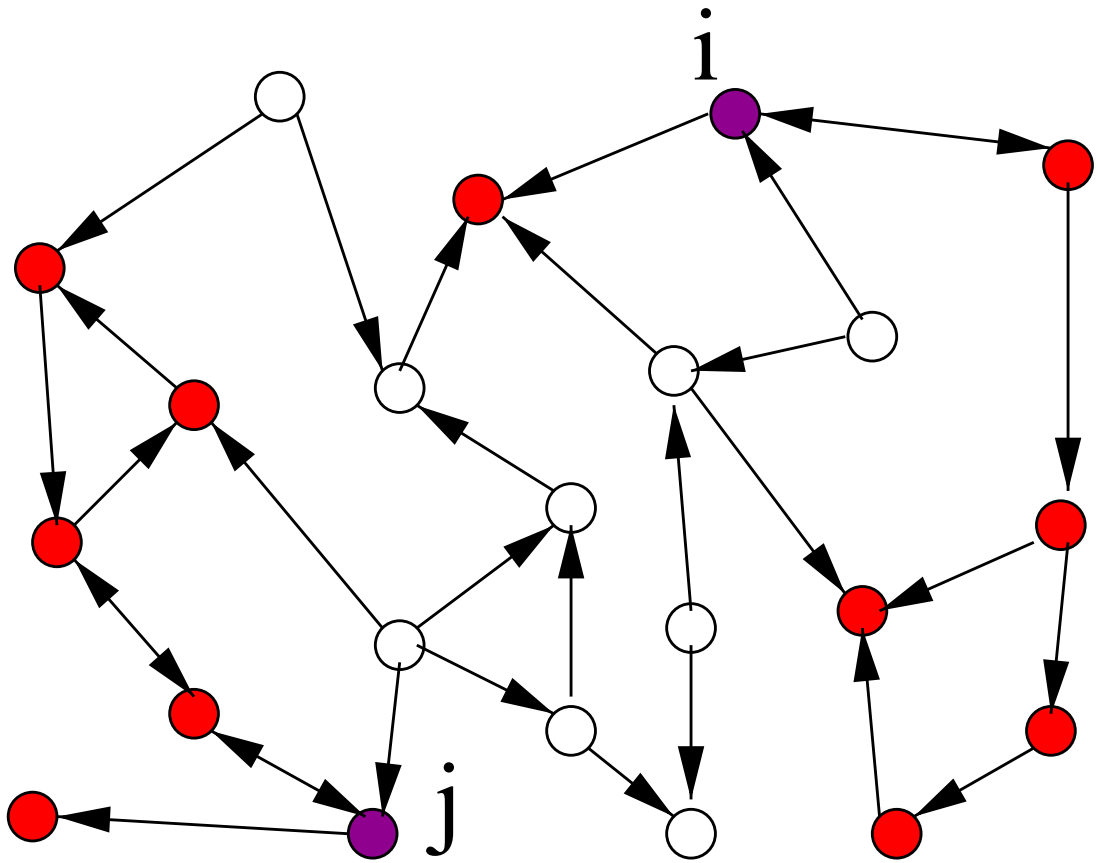
- $i \to j$ if and only if i, if infected, contacts j locally.
- Conditional on the infectious periods $I_1, I_2, \ldots, I_N$, $\mathbb{P}(i \to j) = 1 - \mathrm{e}^{-\lambda_{ij}^L I_i}$ independently for distinct (i, j).

Local infectious clump $\mathcal{C}_i^N$



- Define $\mathcal{C}_i^N = \{j \in \mathcal{N} : i \rightsquigarrow j\}$, where $i \rightsquigarrow j$ if and only if there exists a chain of directed arcs from i to j.
- Set $C_i^N = |\mathcal{C}_i^N|$.

Local infectious clumps



- $\mathcal{C}_i^N = \{j \in \mathcal{N} : i \rightsquigarrow j\}$; $C_i^N = |\mathcal{C}_i^N|$.
- In early stages, clumps don't overlap if N is large (unless local epidemic is supercritical).

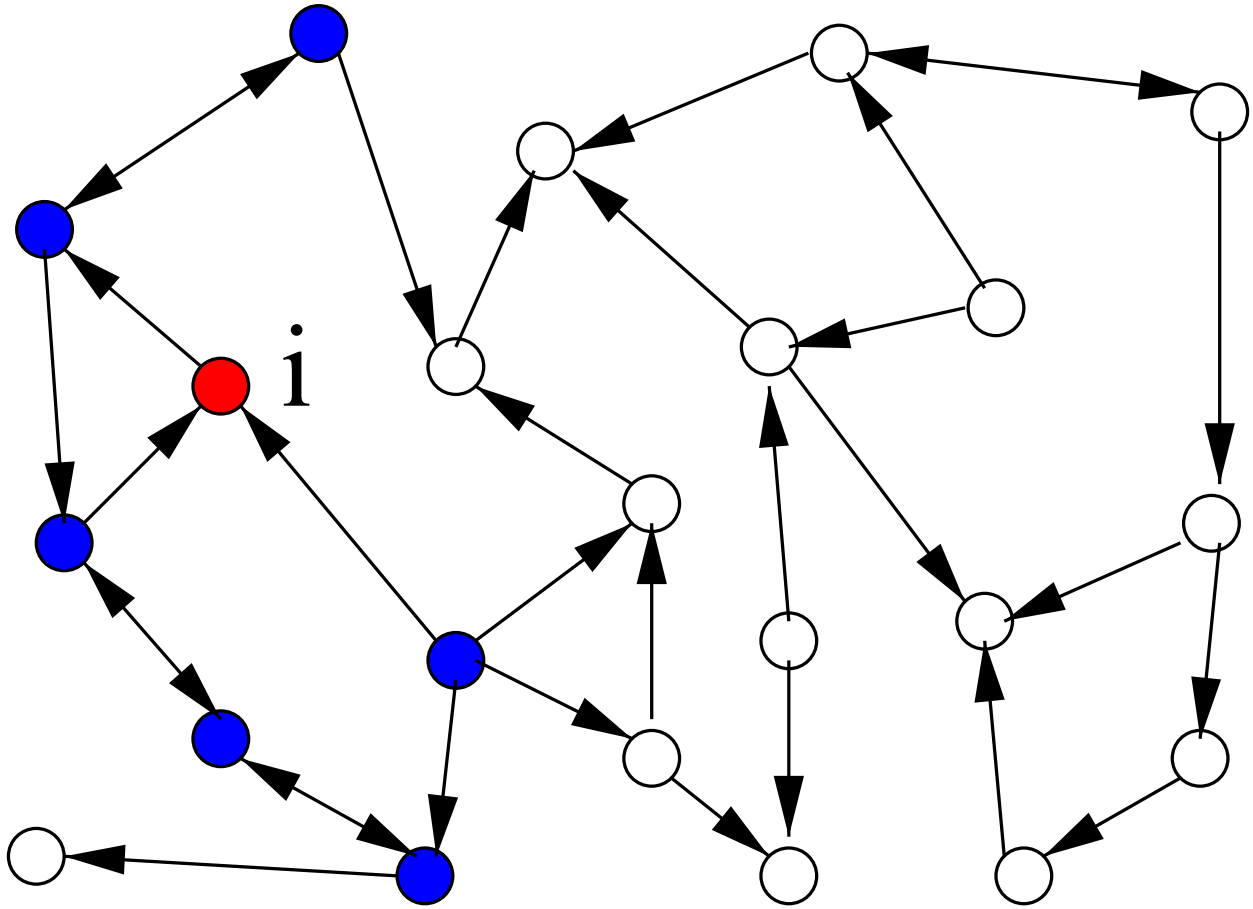
Threshold parameter R_*

- As $N \to \infty$, process of *infected clumps* tends to a branching process having offspring random variable $R \sim \mathrm{Poi}(\lambda_G A)$, where $A = \sum_{j \in \mathcal{C}} I_j$.
- Major outbreak occurs if and only if this branching process does not go extinct.
- We therefore have a threshold parameter

$$R_* = \mathbb{E}[R] = \lambda_G \mathbb{E}[A] = \lambda_G \mathbb{E}\left[\sum_{j \in \mathcal{N}} I_j 1_{\{j \in \mathcal{C}\}}\right]$$
$$= \lambda_G \sum_{j=1}^{N} \mathbb{E}[I_j] \mathbb{P}(j \in \mathcal{C}) = \lambda_G \mu_I \mathbb{E}[C].$$

- Thus $p_{\mathrm{maj}} > 0 \iff R_* > 1$.

Local susceptibility set $\mathcal{S}_i^N$



- $\mathcal{S}_i^N = \{j \in \mathcal{N} : j \rightsquigarrow i\}$; $S_i^N = |\mathcal{S}_i^N|$.
- i is ultimately infected $\iff$ $\mathcal{S}_i^N$ is contacted globally.

Final outcome of major outbreak

Suppose N is large and there are few initial infectives. Set
$z =$ expected proportion of the population infected by the epidemic and
$\pi =$ probability that a typical individual avoids *global* infection.

- Then z and π satisfy the following equations:

$$\pi = \exp\left(-\frac{\lambda_G}{N} N z \mu_I\right) = \exp(-\lambda_G z \mu_I), \tag{3}$$

$$\begin{aligned} 1 - z &= \mathbb{P}(\text{typical susceptible avoids infection by epidemic}) \\ &= \mathbb{P}(\text{typical local susceptibility set avoids global infection}) \\ &= \sum_{k=1}^{\infty} \mathbb{P}(S = k)\pi^k = f_S(\pi) = f_S(\mathrm{e}^{-\lambda_G z \mu_I}). \end{aligned} \tag{4}$$

- If $R_* \leq 1$ then $z = 0$ is the only solution of (3)–(4) in $[0, 1]$.
- If $R_* > 1$ then there is a unique second solution $z^* \in (0, 1)$, giving mean 'size' of major outbreak.
- Proof of this (and CLT) available using Scalia-Tomba (1985) embedding technique. (Local digraphs and global Sellke-type construction.)

Final outcome of major outbreak

- To show that equation (4),

$$1 - z = f_S(\mathrm{e}^{-\lambda_G z \mu_I}),$$

has a *unique* solution in $(0, 1]$ when $R_* > 1$, consider a Galton-Watson process with offspring random variable, R say, having a Poisson distribution with random mean $\lambda_G \mu_I S$.

- Then

$$\mathbb{E}[R] = \lambda_G \mu_I \mathbb{E}[S] = R_*$$

and

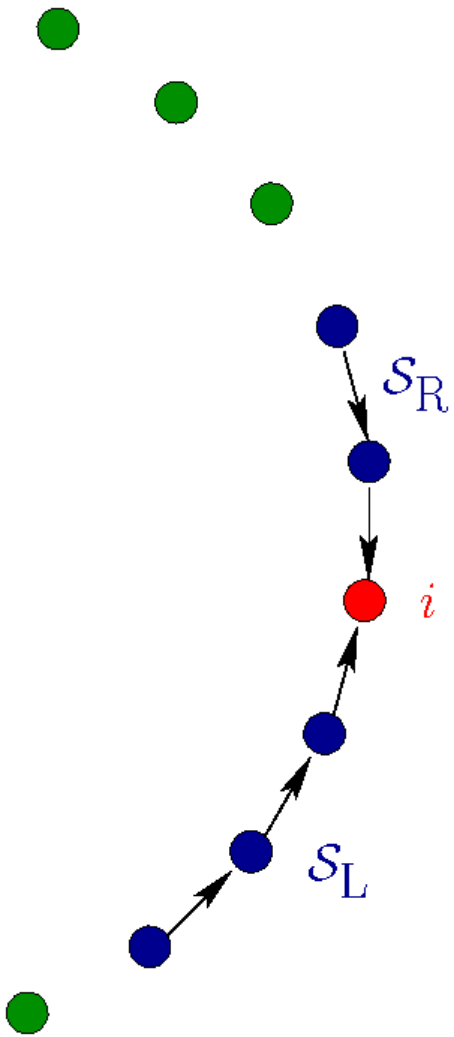
$$f(s) = \mathbb{E}\left[s^R\right] = \mathbb{E}\left[\mathbb{E}\left[s^R \mid S\right]\right] = \mathbb{E}\left[\mathrm{e}^{-\lambda_G \mu_I S(1-s)}\right] = f_S\left(\mathrm{e}^{-\lambda_G \mu_I (1-s)}\right).$$

- Suppose that $R_* > 1$. Then $f(s) = s$ has a unique solution in $[0, 1)$. Hence, setting $z = 1 - s$, shows that (4) has a unique solution in $(0, 1]$.

General two-level mixing model

- Analyse early spread of epidemic via local infectious clumps.
- Analyse final size properties of a major outbreak by local susceptibility sets.

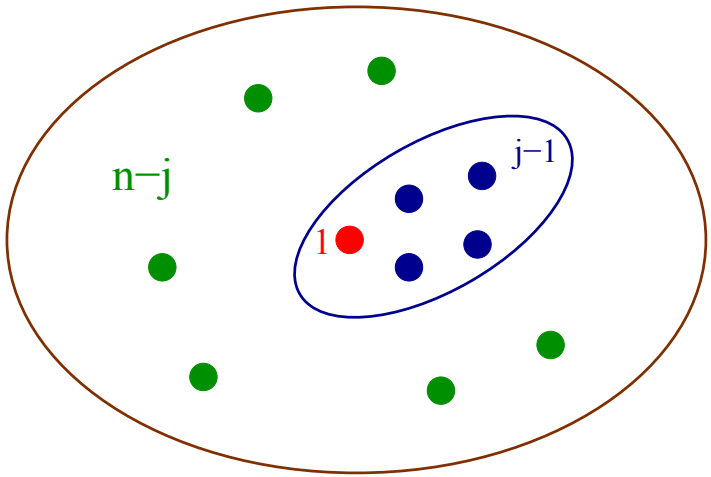
Great circle model



- $\mathcal{S}_i = \{i\} \cup \mathcal{S}_L \cup \mathcal{S}_R$
- $p_L = \mathbb{P}(i \text{ infects } i+1 \text{ locally}) = 1 - \mathbb{E}[\mathrm{e}^{-\lambda_L I}]$
- $\mathbb{P}(S_L = k) = \mathbb{P}(S_R = k) = (1-p_L)p_L^k$
  $(k = 0, 1, \dots)$
- S_L and S_L are independent, so
  $\mathbb{P}(S = k) = (1-p_L)^2 p_L^{k-1} \quad (k = 1, 2, \dots)$
- $\mathbb{E}[S] = 2p_L^{-1} - 1$
- f_S follows easily

Households model

Consider household of n individuals, labelled $1, 2, \ldots, n$, and let $\mathcal{S}$ be the local susceptibility set of individual 1.



Let $P_j^{(n)} = \mathbb{P}(S = j) \qquad (j = 1, 2, \ldots, n)$

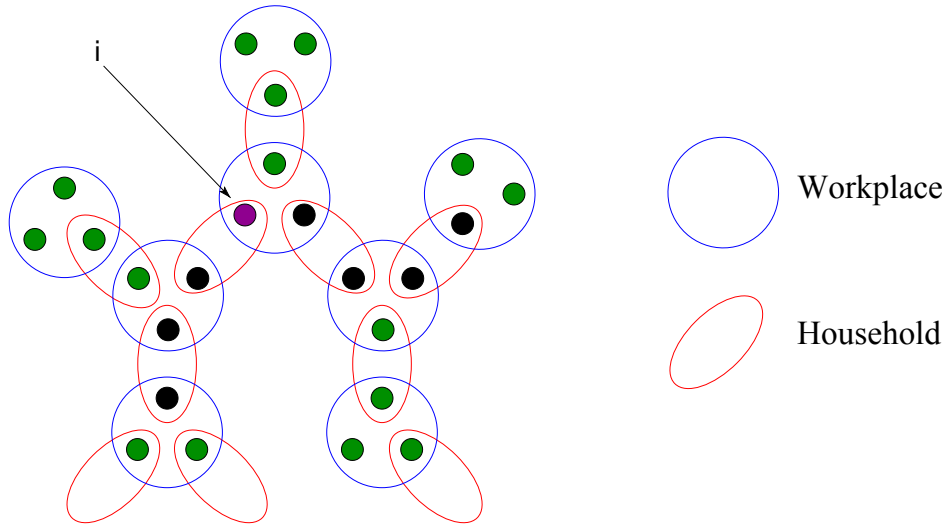
Let $q_k = \mathbb{E}[\mathrm{e}^{-k\lambda_L I}]$ be the probability that a given set of k susceptibles avoids local infection from a given infective

$$P_j^{(n)} = \binom{n-1}{j-1} P_j^{(j)} q_j^{n-j} \qquad (j = 1, 2, \ldots, n)$$

- $$\sum_{j=1}^{k} P_j^{(k)} = 1 \implies \sum_{j=1}^{k} \binom{k-1}{j-1} P_j^{(j)} q_j^{k-j} = 1$$
$$\implies \frac{\sum_{j=1}^{k} \binom{n-k}{n-j} P_j^{(n)}}{q_j^{n-k}} = \binom{n-1}{k-1} \qquad (k = 1, 2, \ldots, n)$$

- Triangular system of linear equations for $\mathbb{P}(S = j) \qquad (j = 1, 2, \ldots, n)$

Overlapping groups model



- Construct local susceptibility set $\mathcal{S}$ of typical individual i via a two-type branching process in which individuals beget only the opposite type and the offspring of a type α (β) individual are the individuals in its workplace (household) susceptibility set.
- If μ_α (μ_β) is the mean size of a household (workplace) susceptiblity set, then

$$\mathbb{E}[S] = \begin{cases} \frac{\mu_\alpha \mu_\beta}{\mu_\alpha + \mu_\beta - \mu_\alpha \mu_\beta} & \text{if } (\mu_\alpha - 1)(\mu_\beta - 1) < 1, \\ \infty & \text{otherwise.} \end{cases}$$

Network: Configuration model

- Population $\mathcal{N} = \{1, 2, \cdots, N\}$.
- $D =$ degree of typical individual,

$$p_k = \mathbb{P}(D = k) \quad (k = 0, 1, \dots) \quad \text{specified} \qquad \mu_D = \mathbb{E}[D].$$

- $D_1, D_2, \dots, D_N$ IID copies of D.
- Attach D_i half-edges to individual i $(i = 1, 2, \dots, N)$.
- Pair up the S_N half-edges uniformly at random to form the network.
- There may be imperfections; but these are sparse if $\sigma_D^2 = \mathrm{var}(D) < \infty$.

(Bollobás (2001))

Networks with casual contacts

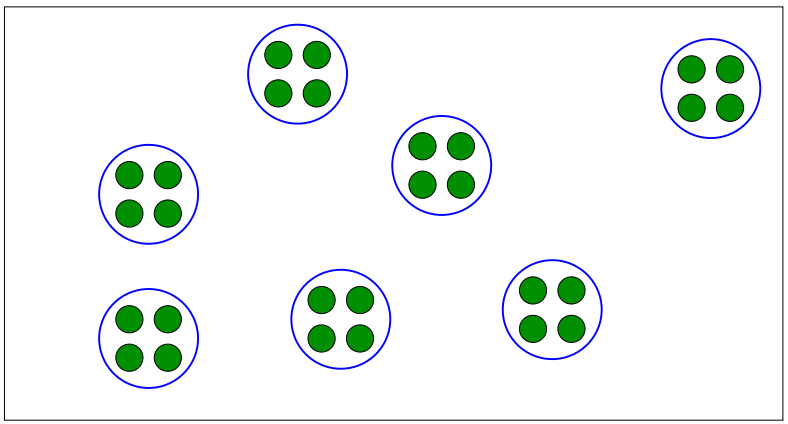
- Let $\tilde{D}$ = degree of typical neighbour of typical individual in the network and $\mu_{\tilde{D}} = \mathbb{E}[\tilde{D}]$. Then

$$\mathbb{P}(\tilde{D} = k) = kp_k/\mu_D \quad (k = 1, 2, \dots) \quad \text{and} \quad \mu_{\tilde{D}} = \frac{\sigma_D^2 + \mu_D^2}{\mu_D}.$$

- Size of typical local susceptibility set $S^{(N)} \xrightarrow{a.s.} S$ as $N \to \infty$, where S is the total size of a Galton-Watson process having offspring law $\mathrm{Bin}(D, p_L)$ for the initial individual and $\mathrm{Bin}(\tilde{D} - 1, p_L)$ for all subsequent individuals.
- It follows that

$$\mathbb{E}[S] = \begin{cases} 1 + \frac{\mu_D p_L}{1-(\mu_{\tilde{D}}-1)p_L} & \text{if } (\mu_{\tilde{D}} - 1)p_L < 1, \\ \infty & \text{otherwise.} \end{cases}$$

'Deterministic' households model



m households of size n, labelled $1, 2, \ldots, m$.

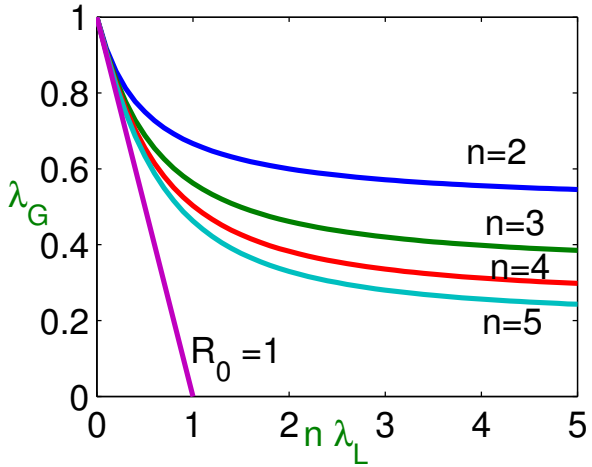
Let $x_i(t)$ and $y_i(t)$ be the number of susceptibles and infectives in household i at time t.

- $$\frac{dx_i}{dt} = -(\lambda_L y_i + N^{-1}\lambda_G \sum_{j=1}^{m} y_j)x_i,$$
$$\frac{dy_i}{dt} = (\lambda_L y_i + N^{-1}\lambda_G \sum_{j=1}^{m} y_j)x_i - \gamma y_i \quad (i = 1, 2, \ldots, m).$$

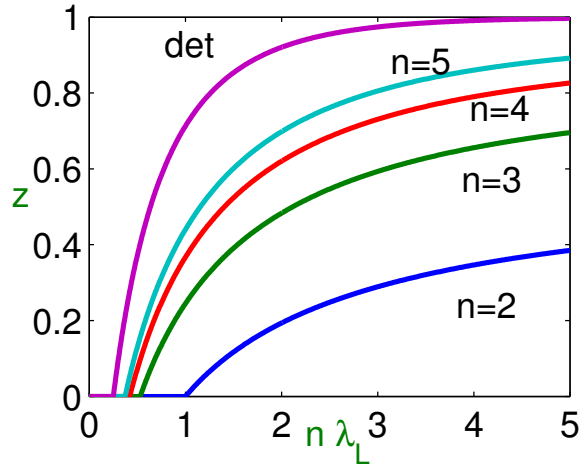
- Basic Reproduction number $R_0 = (\lambda_G + n\lambda_L)/\gamma$.
- Proportion of susceptibles ultimately infected z^*_{det} given by largest root of $1 - z = \exp(-R_0 z)$ in $[0, 1]$

Households and great circle models

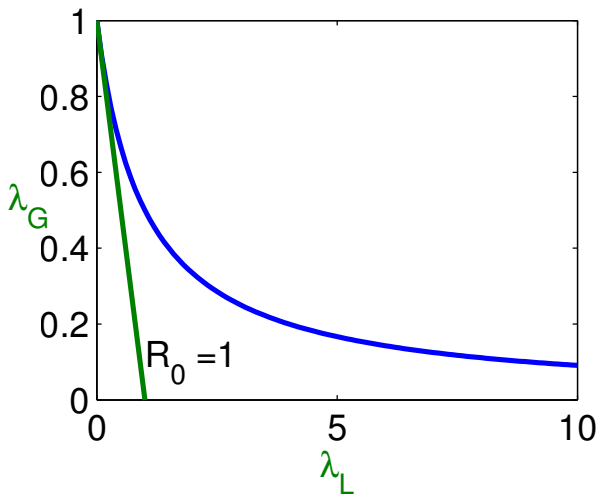
Households, R_*=1.



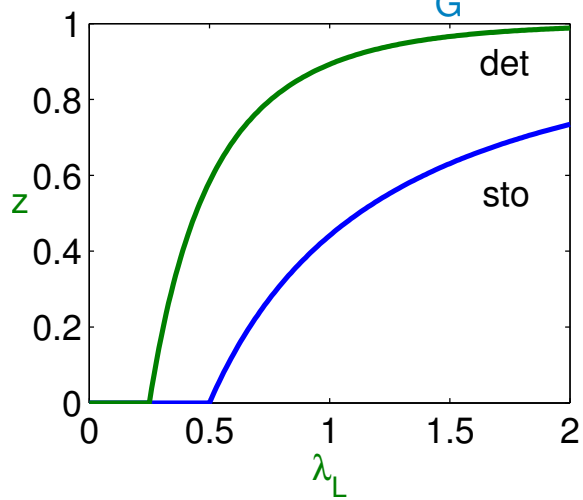
Households, z, λ_G=0.75.



Great circle, R_*=1.

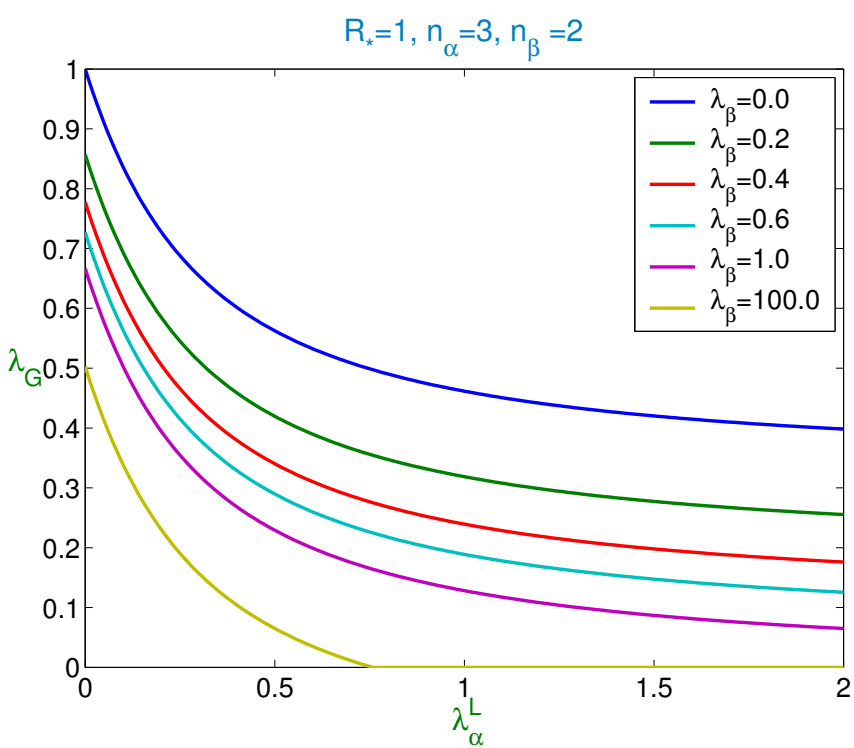


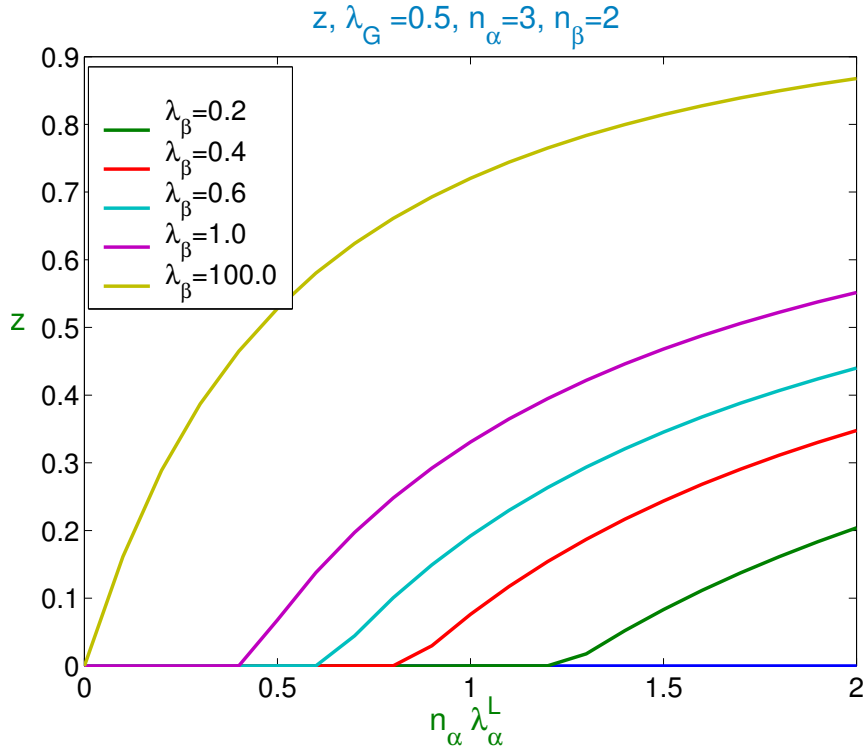
Great circle, z, λ_G=0.5



Critical values of (λ_L, λ_G) so that $R_* = 1$ and final outcome z^* when $I \sim \mathrm{Exp}(1)$.

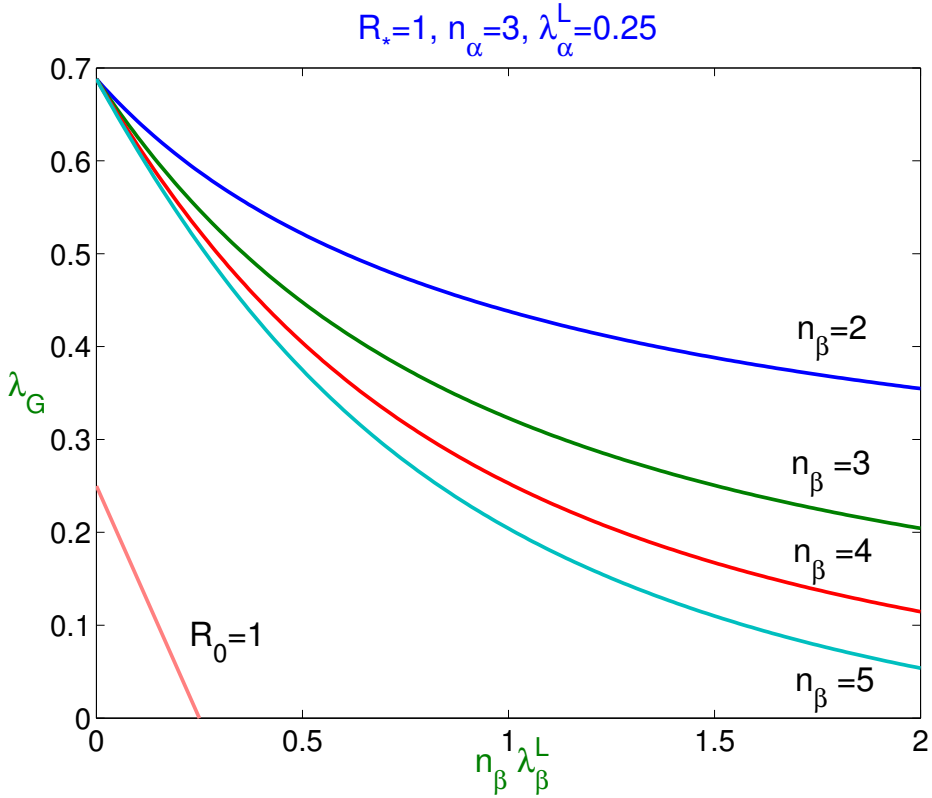
Overlapping groups model, varying λ_β^L

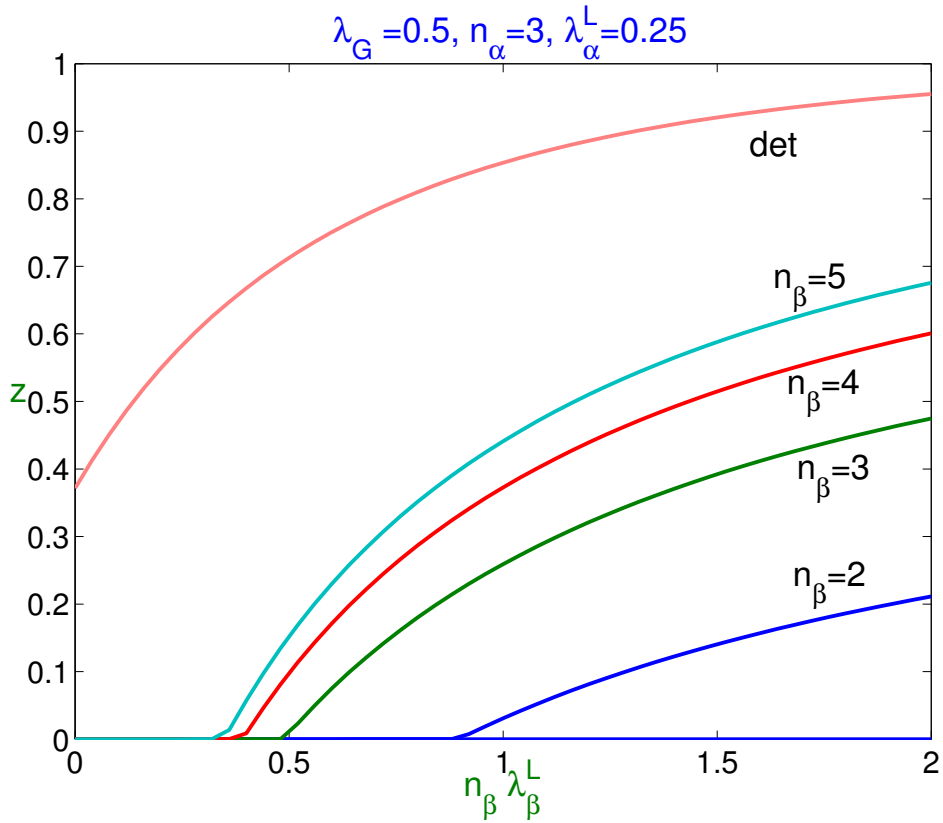




Critical values of $(\lambda_\alpha^L, \lambda_G)$ so that $R_* = 1$ and final outcome z^* when $I \sim \mathrm{Exp}(1)$

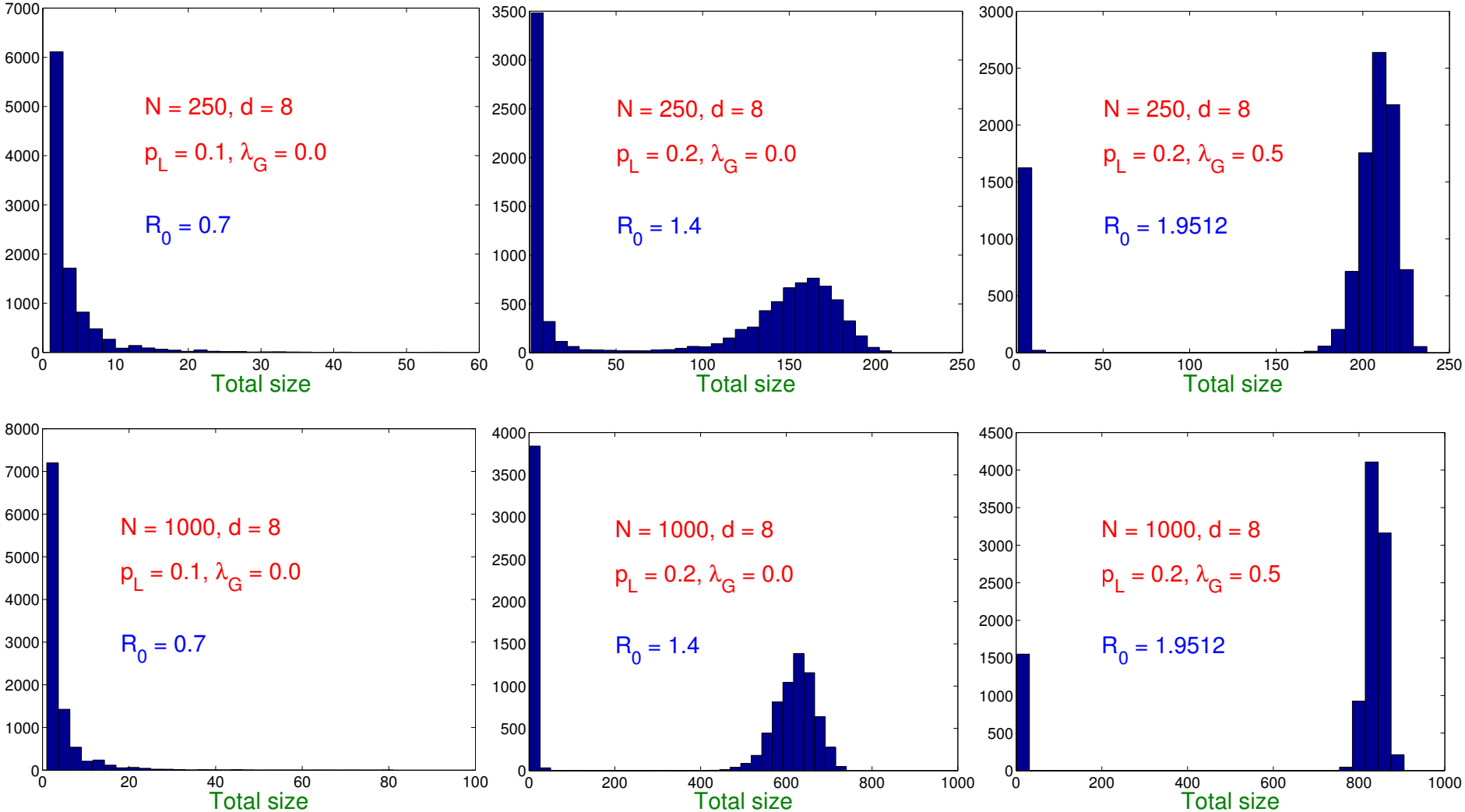
Overlapping groups model, varying n_β





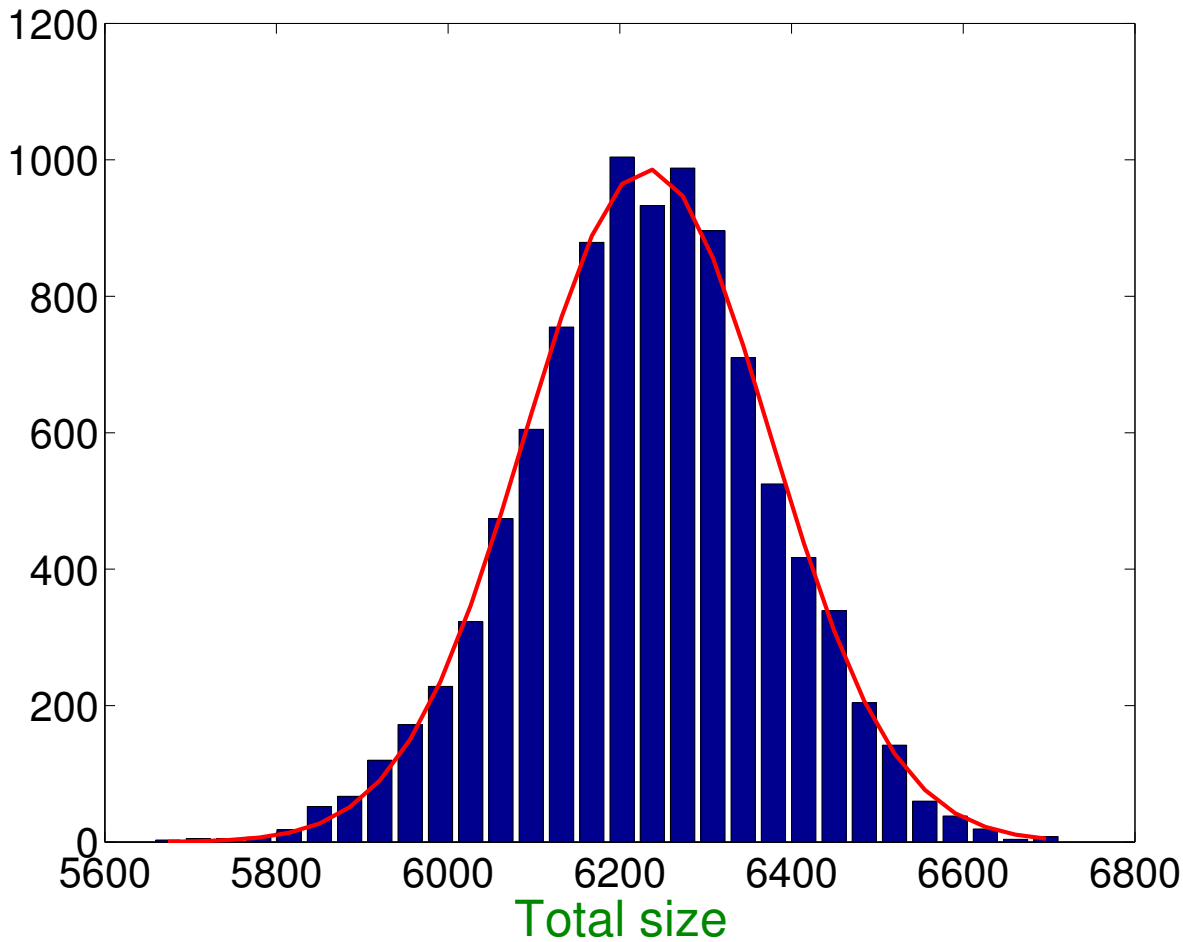
Critical values of $(\lambda_\beta^L, \lambda_G)$ so that $R_* = 1$ and final outcome z^* when $I \sim \mathrm{Exp}(1)$

Networks with casual contacts



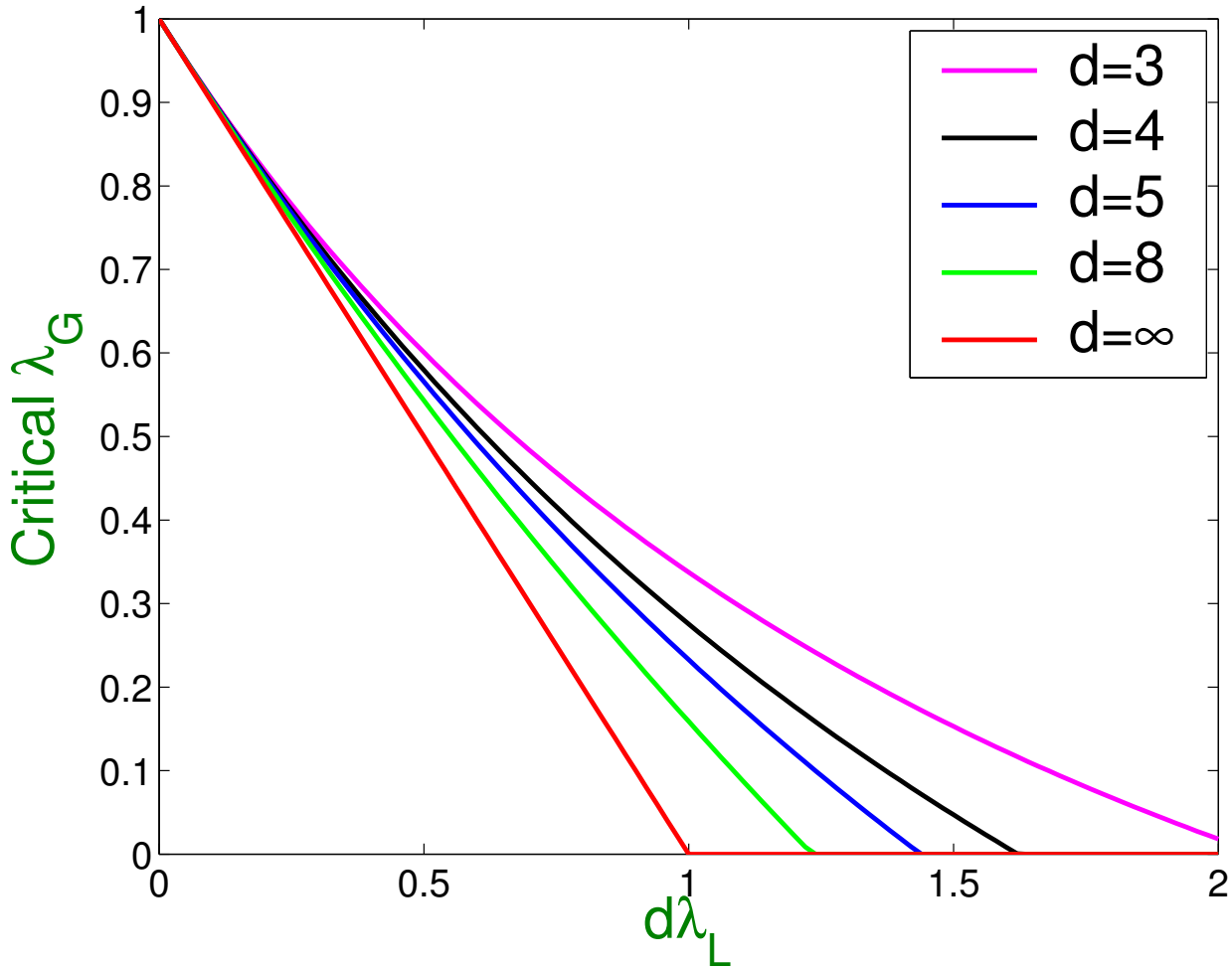
Histograms of size of 10,000 simulated epidemics per parameter combination, for a constant-degree network with $D \equiv d, I \equiv 1$ and other parameters as shown.

Illustration of CLT



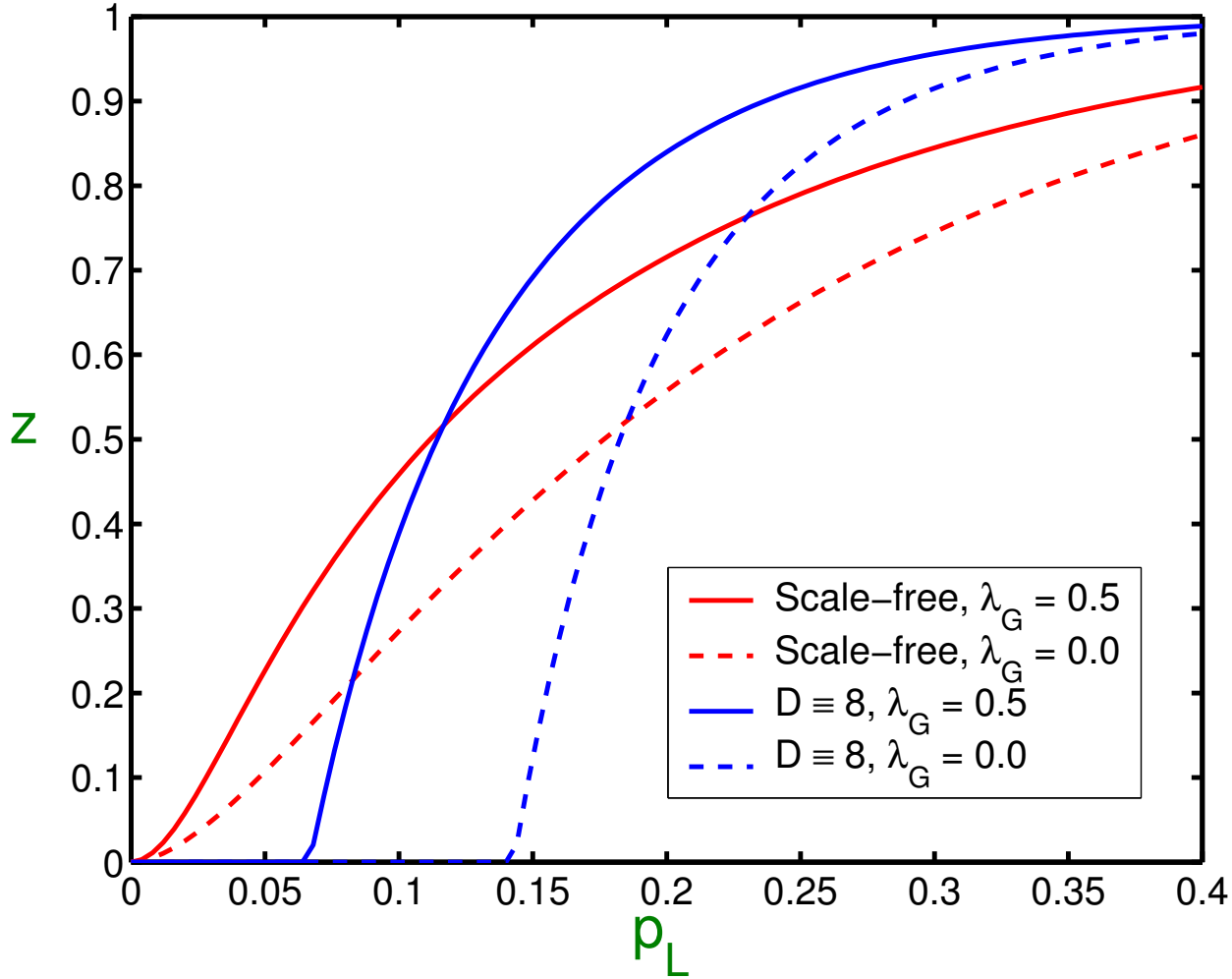
Histogram of size of 10,000 simulated major outbreaks in a population of size $N = 10,000$ when $D \equiv 8, \lambda_G = 0$ and $p_L = 0.2$ ($I \equiv 1$ and $\lambda_L = -\log 0.8$), with asymptotic normal approximation superimposed.

Networks with casual contacts



Critical values of (λ_L, λ_G) so that $R_* = 1$ when $I \equiv 1$. (Expected number of potentially infectious contacts made by an infective is $\lambda_G + d\lambda_L$.)

Networks with casual contacts

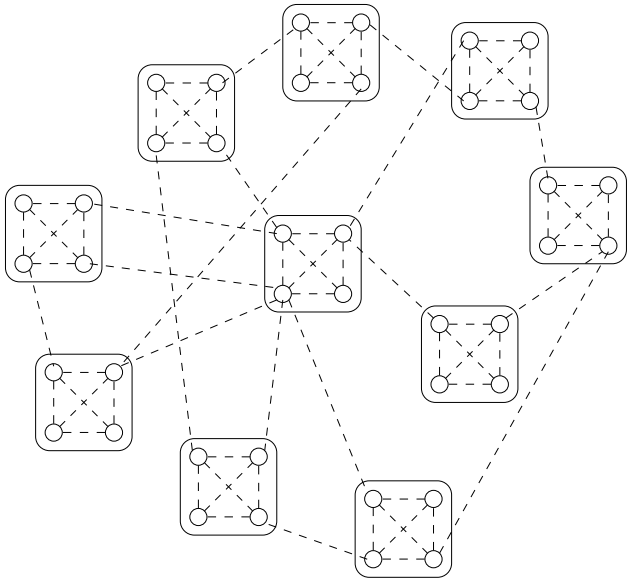


Asymptotic proportion of population infected by major outbreak, z, for constant-degree and scale-free ($\mathbb{P}(D = k) \propto k^{-2.466956}$ $(k = 3, 4, \dots)$) networks with $\mu_D = 8$ when $I \equiv 1$.

Section 4: Households and network model

- Standard households model has household structure and (global) homogeneous mixing.
- Use a network instead of homogeneous mixing for the global mixing.
- The model for the network is the configuration model.

Household and network SIR epidemic model



m_n households of size n $\quad (n = 1, 2, \dots)$.

Total of $m = \sum_{n=1}^{\infty} m_n$ households

and $N = \sum_{n=1}^{\infty} n m_n < \infty$ individuals.

Network/global degrees $\sim D$, arbitrary.

- SIR (susceptible $\rightarrow$ infective $\rightarrow$ removed) progression.
- Infectious period $\sim I$, an arbitrary but specified distribution.
- Infection rates (individual to individual)

  (i) local (within-household) λ_L,
  (ii) network (between-household) λ_G.

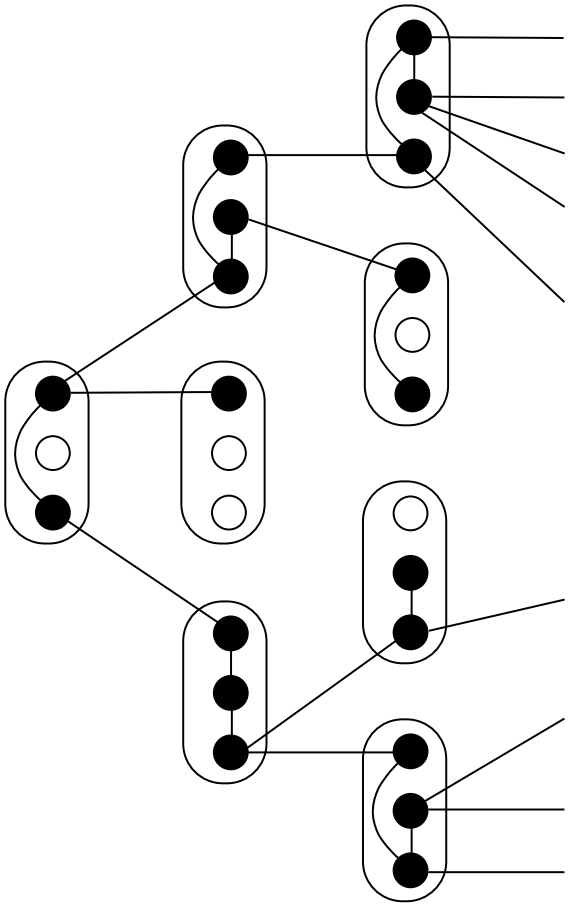
- (Latent period.)

(Ball, Sirl & Trapman (2009, 2010))

Analysis of household/network model

- Basic ideas are the same as for the standard households model: approximate early stages by a branching process of infected households.
- Analysis is more complex, because the within household severity doesn't give enough information to determine the number of global contacts.
- This is because individuals are heterogeneous in their connectivity.

Approximation of early stages



- Branching process of infected households (generation basis).
- Offspring of a household are the households its members infect globally.
- An individual contacted through the network has degree distributed as $\tilde{D}$.
- In the initial household all individuals have degree distributed as D; in subsequent infected households the primary case has degree $\tilde{D}$.

- BP characterised by the distributions of (i) $\tilde{C}$ and (ii) C; the number of network infectious contacts emanating from a household with a single primary case (i) infected through the network / (ii) chosen UAR.

Threshold parameter

- A major outbreak is possible if $R_* = \mathbb{E}[\tilde{C}] > 1$.
- Letting $T = \sum_{n=2}^{\infty} \tilde{\alpha}_n \, \mathbb{E}[Z^{(n)}]$ be the expected number of secondary cases in a household (i.e. the expected final size of $E_{n-1,1}(\lambda_L, I)$, averaged over n),

$$R_* = \Big(\mathbb{E}[\tilde{D} - 1] + \mathbb{E}[T]\mathbb{E}[D]\Big)\, p_G.$$

- The first term is the expected number of network neighbours of those infected in the within-household epidemic
  and $p_G = 1 - \mathbb{E}[\mathrm{e}^{-\lambda_G I}] = 1 - \phi(\lambda_G)$ is the probability of each of those neighbours being infected.
- Evaluate numerically, since $\mathbb{E}[T]$ is complicated.

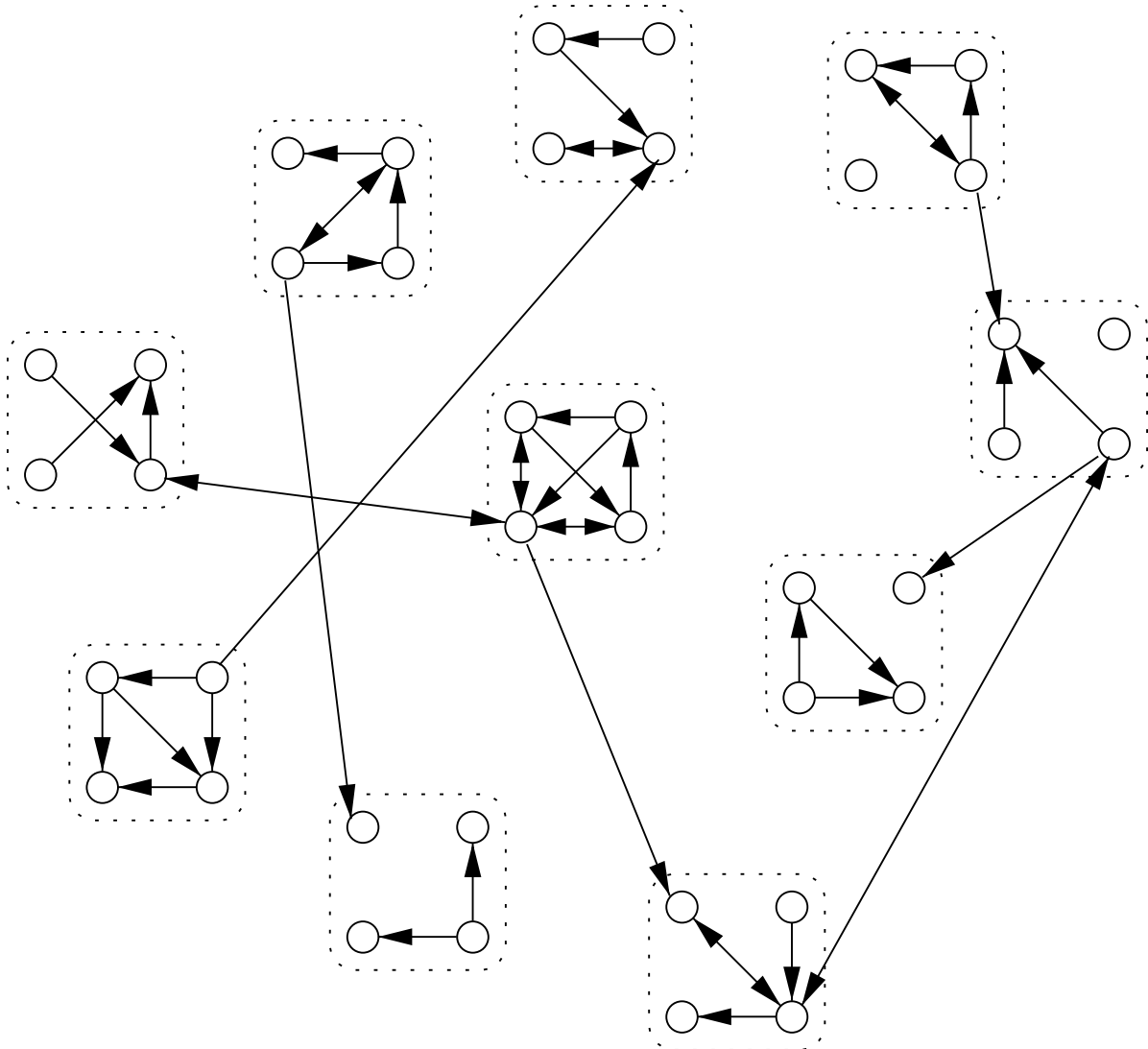
Major outbreak probability

- Early stages of the proliferation of infected households approximated by a GW processes with (i) 1 ancestor, (ii) offspring random variable C in first generation, (iii) offspring random variable $\tilde{C}$ in subsequent generations.
- Therefore $p_{\mathrm{maj}} \approx 1 - f_C(\sigma)$, where σ is the smallest solution of $f_{\tilde{C}}(s) = s$ in $s \in [0, 1]$.
- Calculating these PGFs is difficult due to dependencies between the number of network contacts made by different individuals in the same household.
- "Final state random variables" of Ball & O'Neill (1999) can be used to overcome this.

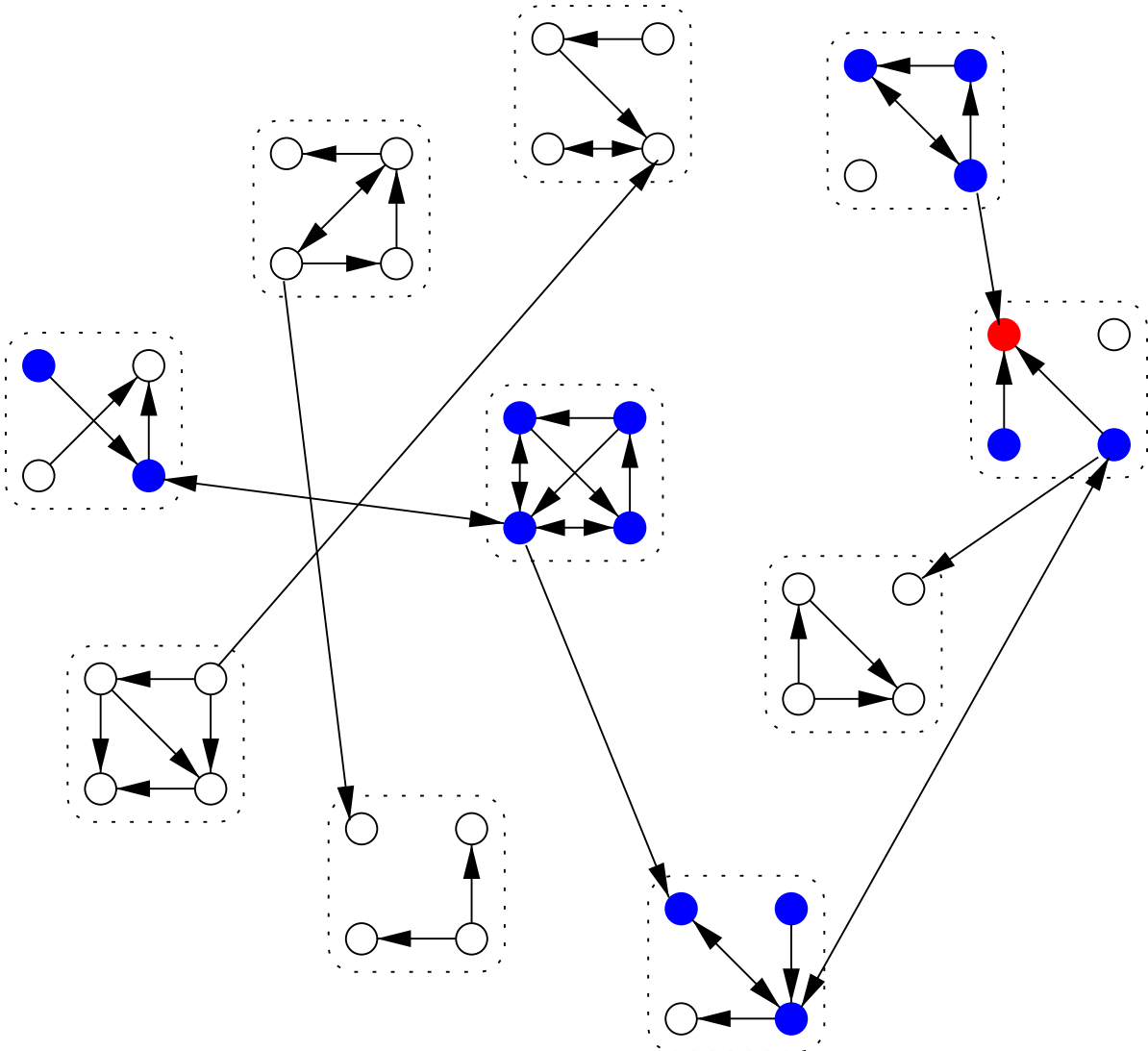
Susceptibility sets

- For each individual in the population, take samples from the infective period distribution and the relevant Poisson processes.
- This gives a list of which individuals each individual would have infectious contact with, were it to become infected.
- Construct a (random) digraph with an arc from i to j iff j is in i's list.
- We then say that $j \in \mathcal{S}_i$ (i's susceptibility set) if there is a path from j to i in this digraph.
- Individual i becomes infected if a member of $\mathcal{S}_i$ becomes infected.

Example



Example



Final size of a major outbreak

- An exhangability argument tells us that the probability that a given individual is infected is equal to the expected proportion of individuals that are ultimately infected.
- We find that, as $m \to \infty$, in the event of a major outbreak, an initially susceptible individual is ultimately infected iff its susceptibility set is infinite.
- We construct the susceptibility set of an individual by 'generations' in much the same manner as our analysis of the early stages of an epidemic.
- This leads to a branching process approximation for the size of the susceptibility set of a typical initial susceptible.

Final size of a major outbreak

- The expected relative final size z of a major outbreak is the probability that this branching process avoids extinction.
- In this BP
  - Individuals are households that have members in the susceptibility set (of an individual chosen UAR).
  - There is 1 ancestor and the initial and subsequent generations have different offspring distributions.
- We have $z \approx 1 - f_B(\xi)$, where ξ is the smallest solution of $f_B(s) = s$ in $[0, 1]$. Here

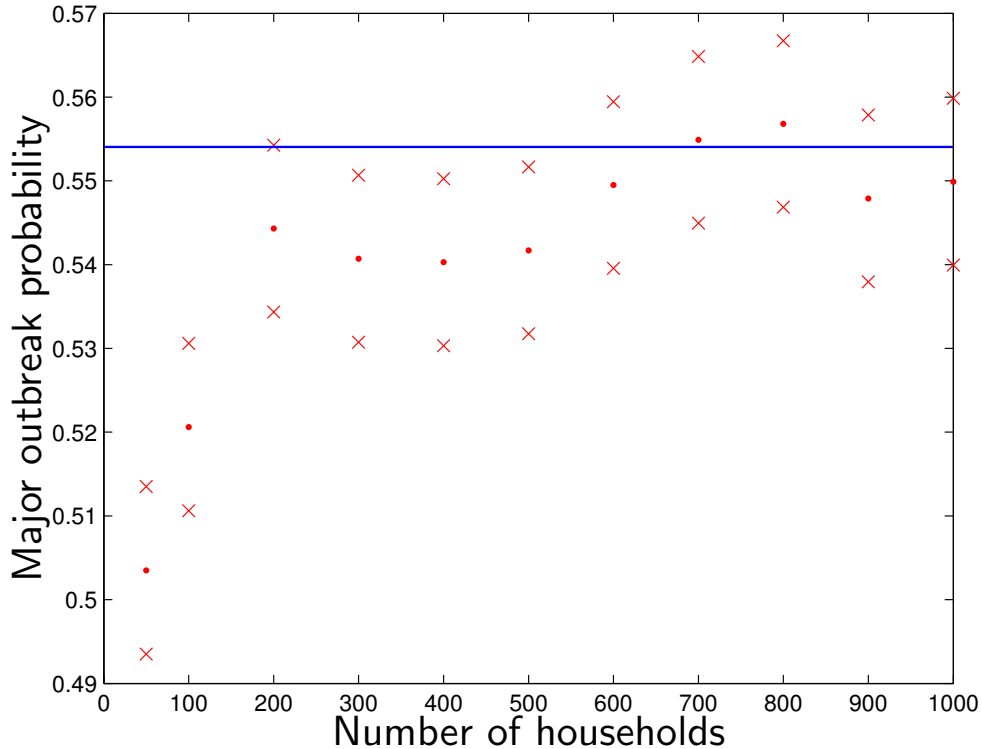
$$f_{\tilde{B}}(s) = f_{\tilde{D}-1}(1 - p_G + p_G s)\, f_M(f_D(1 - p_G + p_G s)),$$
$$f_B(s) = f_D(1 - p_G + p_G s)\, f_M(f_D(1 - p_G + p_G s)),$$

and M is the size of a typical individual's local susceptibility set.

Within-household final size

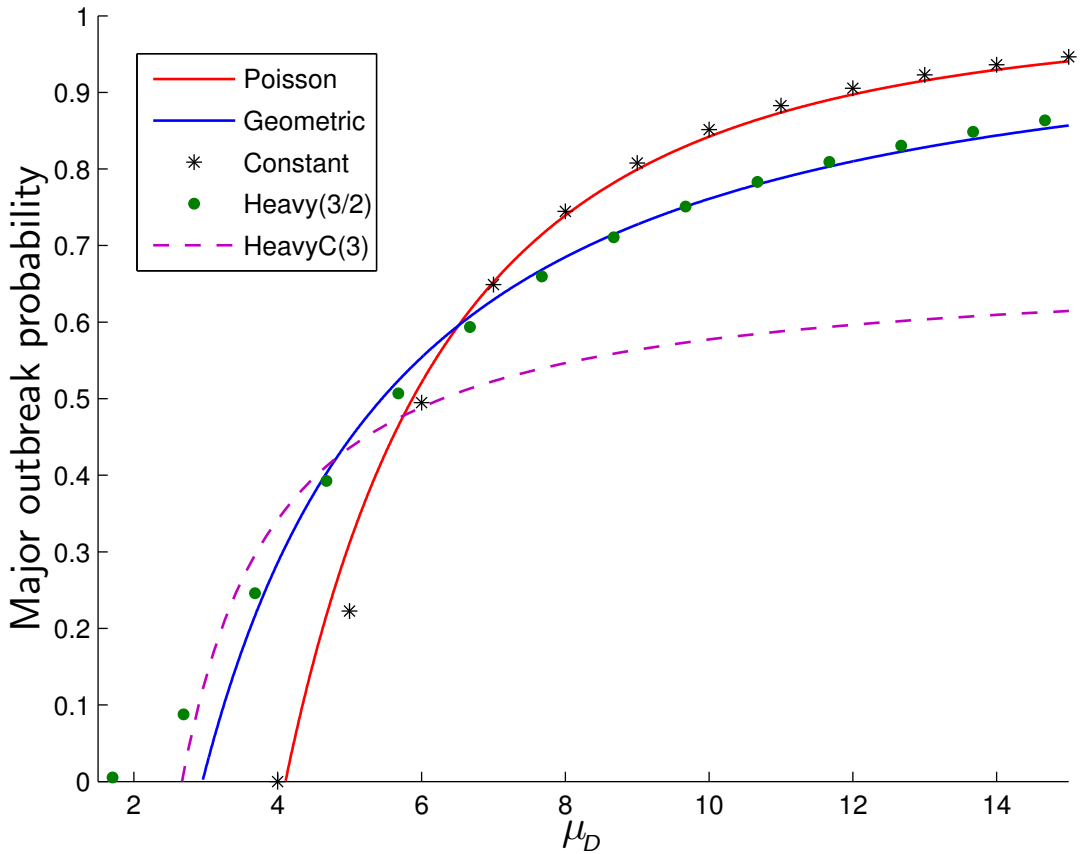
- We can also compute the distribution of the number of ultimately infected individuals within a given household in the event of a major outbreak.
- We find the probability that a set A of individuals within a household avoid infection (using the same branching process approximation of susceptibility set size).
- Some combinatorics then yields a formula for the mass function of the within-household final size, in terms of branching process extinction probabilities.

Numerical results



Simulation-based estimates of major outbreak probability against number of households, together with asymptotic value, for the model with $H \sim \mathrm{U}(\{1,2,3\})$, $D \sim \mathrm{Geom}(3/4)$, $I \equiv 1$, $\lambda_L = 2$, $\lambda_G = 1/4$. Each estimate is based on 10,000 simulations and the plot shows the sample proportions $\pm$ 2SE.

Numerical results



Major outbreak probability dependence on D.
Other parameters are $H \equiv 3$, $I \equiv 1$, $\lambda_L = 1$, $\lambda_G = 1/10$.

Vaccination

- In advance of any outbreak
- Vaccine allocation models
  - Individuals chosen UAR
  - Household based
  - Network based
- Vaccine action model

Network based vaccine allocation

- 'Best' allocation vaccinates individuals of highest degree
- 'Worst' allocation vaccinates individuals of lowest degree
- More realistically try to target individuals of higher degree
  - Sample individuals UAR from the population
  - Sampled individuals name some of their neighbours
  - These named individuals are vaccinated
- Neighbours of individuals are likely to have higher degree than typical individuals, so this achieves that aim.

Cohen *et al.* (2003); Britton *et al.* (2007); Ball & Sirl (2013)

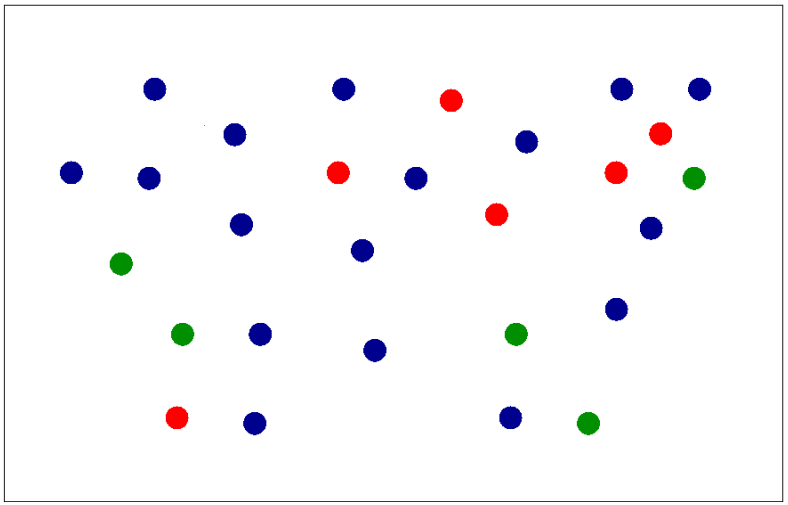
Section 5: Multitype SIR model

- The previous sections have focused on introducing small groups/networks of individuals (local contacts) to the standard homogeneously mixing model.
- The models are analysed in the large population limit, but with these groups remaining 'small'.
- Multitype models allow for the situation where these sub-groups of the population also become large in the large population limit.
- For example, some infectious diseases have the property that there are some groups of the population that have different susceptibility to infection and/or infectivity if infected.

Multitype SIR model

- Suppose that the population can be split into k *types*.
- All individuals of the same type are homogeneous, with respect to infectious period distribution, susceptibility, infectivity and mixing rates.
- Types might reflect one (or more) of
  - age groups,
  - vaccination status,
  - prior immunity,
  - geographic location,

Multitype SIR epidemic model



For each $i = 1, 2, \ldots, k$, there are

n_i initial susceptibles of type i and

m_i initial infectives of type i.

Total of $n = \sum_{i=1}^{k} n_i$ susceptibles

and $m = \sum_{i=1}^{k} m_i$ infectives.

- SIR (susceptible $\rightarrow$ infective $\rightarrow$ removed) progression
- Type i individuals have infectious period $\sim I_i$, an arbitrary but specified distribution.
- A type-i infective infects each type-j susceptible at (individual to individual) rate λ_{ij}/N.
- (Latent period, movement between groups)

(Ball (1986), Ball & Clancy (1993))

Exact results

- Denote this process by $E_{\boldsymbol{n},\boldsymbol{m}}(\Lambda, \boldsymbol{I})$, where $\boldsymbol{n} = (n_i)_{i=1}^k$, $\boldsymbol{m} = (m_i)_{i=1}^k$, $\boldsymbol{I} = (I_i)_{i=1}^k$, $\Lambda = (\lambda_{ij})_{i,j=1}^k$.
- Let Z_i and T_i be the final size and severity amongst type i individuals; and write $\boldsymbol{Z} = (Z_i)_{i=1}^k$, $\boldsymbol{T} = (T_i)_{i=1}^k$.
- There are formulae (cf. single-type case) for
  - final size probabilities $\mathbb{P}(\boldsymbol{Z} = \boldsymbol{z}) \quad \boldsymbol{0} \leq \boldsymbol{z} \leq \boldsymbol{n}$,
  - expected final size $\mathbb{E}[\boldsymbol{Z}]$,
  - joint PGF/MGF of $\boldsymbol{n} - \boldsymbol{Z}$ and $\boldsymbol{T}$, $\mathbb{E}[\prod_{i=1}^k s_i^{n_i - Z_i} \exp(-\theta_i T_i)]$.

Large population limits

- These limits are as $n \to \infty$, with k fixed.
- Let $\pi_i = \lim_{n\to\infty} n_i/n$ be the asymptotic proportion of individuals of type i and $\mu_i = \lim_{n\to\infty} m_i/n_i$ be the asymptotic ratio of initial infective to susceptible type i individuals.
- (Assume the matrix $(\mathbb{E}[I_i]\lambda_{ij}\pi_j)_{i,j=1}^k$ is irreducible.)
- There are two cases, depending on whether $\sum_{i=1}^k \mu_i$ is zero or positive.

Case $\sum_{i=1}^{k} \mu_i = 0$

- Let $Z_i' = Z_i + m_i$.
- Then (Z_i') converges in distribution (as $n \to \infty$) to the distribution of the total progeny of a multitype branching process with
  - $\boldsymbol{m}$ ancestors,
  - lifetime distributions $\boldsymbol{I}$,
  - birth rates $(\lambda_{ij}\pi_j)$.
- From this follow
  - basic reproduction number,
  - major outbreak probability.

Case $\sum_{i=1}^{k} \mu_i > 0$

- Let $\hat{Z}_i = \dfrac{Z_i'}{n_i} = \dfrac{Z_i + m_i}{n_i}$.
- Then $(\hat{Z}_i)$ converges in probability to (z_i), which solves the balance equations

$$1 + \mu_j - z_j = \exp\left(-\sum_{i=1}^{k} \pi_i z_i \mathbb{E}[I_i]\lambda_{ij}\right) \qquad (j = 1, \ldots, k),$$

  uniquely in $[0, 1]^k$.
- The vector with entries $\sqrt{n_i}(\hat{Z}_i - z_i)$ satisfies a CLT.

Section 6: Extensions and variations

There are many directions in which these models and analyses can be extended. Here we briefly address

- Imperfect vaccine action models.
- Models with more 'levels' to represent/capture more features.
- Different branching process approximations (and reproduction numbers) in structured models.
- Inference.

And we don't address time evolution, demography, non S(E)IR progression, multiple severities, contact tracing, control measures imposed during an outbreak, ...

(a) Imperfect vaccine action

- Framework of Becker & Starczak (1998).
- Random vaccine response (A, B), independent for each vaccinated individual, describing relative susceptibility and relative infectivity.
- If (A, B) takes finitely many values then the models we have seen can be extended to allow for this using multi-type methods.
- (E.g. In a households model, analyse within-household spread by conditioning on the number of vaccinated individuals with each possible response to the vaccine.)

(b) Models with more levels of mixing

- Homogeneous mixing.
- Network structure/s.
- Household structure.
- Overlapping groups.
- (And multitype variations)

- Simulation-based models.

(c) Different BP approximations

- In structured population epidemic models there is more than one BP approximation (even after assuming that we just look at final size properties).
- This leads to a variety of reproduction numbers.
- Example: In households models we can set up branching processes which approximate the proliferation of
  - infected households,
  - infected individuals.
- We have focused on R_* as it is (generally) easiest to work with.
- Other reproduction numbers are 'nicer' but harder to calculate: R_0 and R_r.

(Pellis, Ball & Trapman (2012); Ball, Pellis & Trapman (2015))

Reproduction numbers: household/network model

- R_*, a household-to-household reproduction number.
- R_1: maximum eigenvalue of $M = \begin{pmatrix} (\mu_{\tilde{D}} - 1)p_G & \mu_T \\ \mu_D p_G & 0 \end{pmatrix}$.
- Here μ_T is the mean number of secondary cases in a household.
- Entries of M give the mean number of primary/secondary individuals infected by a primary/secondary infective; assigning all within-household infections to the primary case.
- M is a mean matrix (next-generation matrix) which attempts to reflect the proliferation of *individuals*.
- R_1 is a threshold parameter, but does not have a neat interpretation.

(d) Inference

- Models with multiple levels of mixing have complex features and thus likelihood functions which can be difficult to deal with analytically.
- Further difficulties arise due to the nature of the available data: missing data, final size data, partial data.
- Progress can sometimes be made by assuming independence where dependence is weak (pseudolikelihood methods).
- MCMC and data augmentation methods.

References I

Addy, C. L., Longini, I. M. and Haber, M. (1991) A generalized stochastic model for the analysis of infectious disease final size data. *Biometrics* 47:961–974.

Andersson, H. (1999) Epidemic models and social networks. *Math. Sci.* 24:128–147.

Andersson, H. and Britton, T. (2000) *Stochastic Epidemic Models and Their Statistical Analysis*. (Lecture Notes in Statistics 151). Springer, New York.

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

Ball, F. G. and Clancy, D. (1993) The final size and severity of a generalised stochastic multitype epidemic model. *Adv. Appl. Prob.* 25:721–736.

Ball, F. G. and Donnelly, P. J. (1995) Strong approximations for epidemic models. *Stoch. Proc. Applic.* 55:1–21.

References II

Ball, F. G. and Lyne, O. D. (2002) Optimal vaccination policies for stochastic epidemics among a population of households. *Math. Biosci.* 177&178:333–354.

Ball, F. G. and Lyne, O. D. (2006) Optimal vaccination schemes for epidemics among a population of households, with application to variola minor in Brazil. *Statistical Methods in Medical Research* 15:481-497.

Ball, F. G. and Lyne, O. D. (2014) Statistical inference for epidemics among a population of households. *J. R. Statist. Soc. Ser. B*, under revision.

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

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

Ball, F. G. and Neal, P. (2003) The great circle epidemic model. *Stoch. Proc. Applic.* 107:233–268.

Ball, F. G. and Neal, P. (2008) Network epidemic models with two levels of mixing. *Math. Biosci.* 212:69–87.

References III

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

Ball, F. G., Pellis, L. and Trapman P. (2015) Reproduction numbers for epidemic models with households and other social structures II: comparisons and implications for vaccination. Submitted. (arXiv:1410.4469).

Ball, F. G. and Sirl, D. J. (2013) Acquaintance vaccination in an epidemic on a random graph with specified degree distribution. *J. Appl. Prob.* 50:1147–1168. (Correction (2015), *ibid.* 52:908.)

Ball, F .G., Sirl, D. J. and Trapman, P. (2010) Analysis of a stochastic SIR epidemic on a random network incorporating household structure. *Math. Biosci.* 224:53–73. (Erratum (2010), *ibid.* 225:81.)

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

Ball, F. G., Sirl, D. J. and Trapman, P. (2014) Epidemics on random intersection graphs. *Ann. Appl. Prob.* 24:1081–1128.

References IV

Bartoszyński, R. (1972) On a certain model of an epidemic. *Applic. Math.* 13:139–151.

Becker, N. G. and Dietz,K. (1995) The effect of household distribution on transmission and control of highly infectious diseases.*Math. Biosci.* 127:207–219.

Becker, N. G. and Starczak, D. (1997) Optimal vaccination strategies for a community of households. *Math. Biosci.* 139:117–132.

Becker, N. G. and Starczak, D. (1998) The effect of random vaccine response on the vaccination coverage required to prevent epidemics.*Math. Biosci.* 154:117–135.

Bollobás, B. (2001) *Random graphs*. Cambridge University Press.

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

References V

Haccou, P., Jagers, P. and Vatutin, V. A. (2005) *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge University Press, Cambridge.

Kiss, I. Z., Green, D. M. and Kao, R. R. (2006) The effect of contact heterogeniety and multiple routes of transmission on final epidemic size. *Math. Biosci.* 203:124–136.

Ludwig, D. (1975) Final size distributions for epidemics. *Math. Biosci.* 23:33–46.

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

Pellis, L., Ball, F. G. and Trapman, P. (2012) Reproduction numbers for epidemic models with households and other social structures I: definition and calculation of R_0. *Math. Biosci.* 235:85–97.

Picard, P. and Lefèvre, C. (1990) A unified analysis of the final size and severity distribution in collective Reed–Frost epidemic processes. *Adv. Appl. Prob.* 22:269–294.

References VI

Scalia-Tomba, G. (1985) Asymptotic final-size distribution for some chain-binomial processes. *Adv. Appl. Prob.* 17:477–495.